The Use of the Ordered Orthogonalized Multivariate Linear Regression in a Structure−Activity Study of Coumarin and Flavonoid Derivatives as Inhibitors of Aldose Reductase

Dragan Amić, Dušanka Davidović-Amić, and Drago Bešlo

Faculty of Agriculture, The Josip Juraj Strossmayer University, P.O. Box 117, HR-31001 Osijek, The Republic of Croatia

Bono Lučić and Nenad Trinajstić*

The Rugjer Bošković Institute, P.O. Box 1016, HR-10001 Zagreb, The Republic of Croatia

Received November 13, 1996[⊗]

The relationship between molecular descriptors and the inhibitory activity of aldose reductase (AR) for a series of coumarin and flavonoid derivatives has been investigated using a novel multivariate linear regression based on the ordered orthogonalized descriptor set. First, starting from the set of 31 descriptors we produced absolutely the best nonorthogonalized QSAR models with I descriptors ($I = 1-7$). These models are always better than the models that the most authors achieve by the use of the stepwise inclusion-exclusion procedure. In the next step we realized all possible orthogonalization orderings of a given set of N descriptors (there are $N!$ of these). The key result is that some orthogonalization orderings lead to QSAR models with I ordered orthogonalized descriptors that have higher values of both the correlation coefficient R and cross-validated correlation coefficient R_{cv} than the corresponding models with the same number of nonorthogonalized descriptors. In order to achieve the highest possible reliability in predicting the inhibition we produced several models that were obtained applying the ordered orthogonalization procedure on one set with five ($N = 5$) and on two sets with seven ($N = 7$) descriptors. Then the inhibitory activity for 34 coumarins and 30 flavonoids was predicted, and several compounds were detected with a very high inhibitory activity.

INTRODUCTION

Coumarins and flavonoids are naturally occurring substances possessing various pharmacological properties and are able, among other things, to inhibit the enzyme aldose reductase (AR) to a greater or lesser degree.[1−3]

AR catalyzes the NADPH-dependent reduction of a wide variety of carbonyl containing compounds to their corresponding alcohols. Increased AR activity leads to high levels of sorbitol and galactitol in the cells of many tissues. Accumulation of sugar alcohols has been shown to cause osmotic cataracts in the lens. AR also play a key role in diabetic complications of nerve, kidney, and retina.[4,5]

Since the inhibitors of AR are potentially clinically useful, considerable interest has been developed in seeking potent inhibitors. The aim of this report is to establish a possible correlation between the molecular properties and the inhibitory activity shown by coumarin derivatives (which are biogenically and chemically related to flavonoids) and flavonoid derivatives. Quantitative structure−activity relationships (QSARs) can be used to predict inhibitory activity preceding experimental investigations. Consequently, QSAR models can be exploited for development of an AR inhibitor of sufficient potency, so as to be useful in the treatment of diabetic cataracts.

In order to obtain the reliable predictions, one must derive as reliable as possible a QSAR model, that is, one must use in the modeling process the best possible approach. Most often in the QSAR modeling is used the multivariate linear regression method (MLR). Studying numerous examples of the application of this method we noticed that the models obtained by the stepwise inclusion−exclusion procedure are not the best possible MLR models. This observation stimulated us to reinvestigate systematically the MLR method in order to detect conditions which lead to the maximum possible accuracy and reliability of the method.

We succeeded in designing an efficient computer program by which we were able to obtain the best possible MLR models with I descriptors ($I = 1, ..., N$; where N is the number of molecular descriptors used in the modeling).[6−8] To increase the computing speed, we first, before computing the MLR models, orthogonalized descriptors applying the Gram−Schmidt orthogonalization procedure for vectors.[9,10] Several authors, including Lukovits (who first used the orthogonalized quantum chemical indices in QSAR studies),[11] Randić,[9] and Amić *et al.*[12] who used the MLR with orthogonalized descriptors, found that these regression models were more stable but not statistically better than the models with nonorthogonalized descriptors, since the values of the correlation coefficient R, the standard errors (S_f, S_d), and the F-test remained the same for the both models. However, the key improvement in the development of the orthogonal MLR models was achieved by the analysis of the dependence of contributions of individual orthogonalized descriptors on the dependence of the orthogonalization order. This is incorporated in our ordered orthogonalized MLR algorithm.[6−8]

Comparisons with other methods show that the best possible nonorthogonalized MLR models, and especially the best ordered orthogonalized MLR models, are superior to

* Author for correspondence. Fax: 385-1-425-497; E-mail: trina@olimp.irb.hr.

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0095-2338(96)00158-8 CCC: $14.00

the corresponding models (derived by the use of the identical set of descriptors) obtained by the use of principal component analysis (PCA)[13] and partial least-squares (PLS),[14] neural network (NN), and genetic algorithms (GA).[15] In recent papers[7,8] we showed that our novel MLR algorithm lead to QSAR models that were better than the models that were obtained with the application of the nonorthogonalized[16,17] and orthogonalized MLR procedures[17] carried out in the usual way.

In this report we show by using the cross-validation correlation coefficient R_{cv} that the ordered orthogonalized MLR models are much more stable in their predictions than the models with nonorthogonalized descriptors.

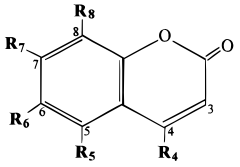
MOLECULAR DESCRIPTORS

The main purpose of QSAR is to relate a given biological activity to the structural features of a particular molecule. Models of different quality for predicting the inhibition of AR by coumarin and flavonoid derivatives can be achieved by considering different QSAR methodologies and different molecular descriptors. Our data base of AR inhibitors consisted of set of 17 compounds: 14 coumarin derivatives and three flavonoid derivatives, whose structures are schematized in Figure 1. The considered set of coumarin and flavonoid aglycons is found to inhibit AR to varying degrees. The experimental values were taken from the paper by Okada *et al.*[18]

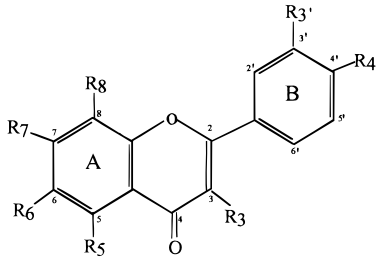
Different molecular descriptors were generated for each studied compound. The descriptors encode the aspects of the structures and electronic properties that are related to the inhibitory activity of molecules. The descriptor pool used contains topological and electronic descriptors. The topological descriptors that were employed include the valence connectivity indexes[19] ${}^{0}\chi^{v}-{}^{5}\chi^{v}$, Wiener number W and path numbers[20] p_0-p_8, Harary indexes[21] H_1-H_6, and molecular topological index[22] MTI. The electronic descriptors include hardness indexes[23] (absolute hardness η, relative hardness η_r), energy of the highest occupied molecular orbital E_{HOMO}, and energy of the lowest unoccupied molecular orbital E_{LUMO}, topological resonance energy indices[24] (topological resonance energy TRE and topological resonance energy per π electron TRE(PE)), dipole moment DM,[25] and the sum of the π charges of the atoms located in the coumarin or chromone moiety $\sum_q$. In addition, various indicator variables I_i were used. Indicator variables express the contribution of certain substituents on the magnitude of biological activity. Typically, such variables flag specific chemical structural features by assigning them a value of 1 for molecules having the feature and 0 otherwise.

METHODS

Nonorthogonalized MLR Method. An algorithm for finding the best possible QSAR models with I nonorthogonalized descriptors ($I = 1, 2, ..., N$)[6-8] is used in the present work. In this algorithm we first apply the orthogonalization procedure on each selected I-tuple of nonorthogonalized descriptors, and, after that, we calculate the correlation coefficient for each selection. In the orthogonal descriptor basis the square of the total correlation coefficient between activity A and its computed value A' is simply the sum of squares of the correlation coefficients between activity A and each individual orthogonal descriptor. For each $I = 1, ..., N$, that I-tuple which produced the highest R value was chosen.



Coumarin derivative	Substitution site					Inhibition[a] %
	R_4	R_5	R_6	R_7	R_8	
1	H	H	OCH_3	OCH_3	H	19
2	H	H	OCH_3	OH	H	60
3	H	H	H	H	H	11
4	H	H	OH	OCH_3	H	19
5	H	OCH_3	OH	OCH_3	H	29
6	H	H	OCH_3	OH	OCH_3	18
7	H	H	H	OH	H	54
8	H	H	OH	OH	H	87
9	OCH_3	H	H	H	H	43
10	H	H	H	OCH_3	H	15
11	H	OCH_3	H	OCH_3	H	42
12	H	H	OCH_3	OAc	H	11
13	H	H	H	OAc	H	0
14	H	H	OAc	OAc	H	14



Flavonoid derivative	Substitution site						Inhibition[a] %
	R_3	R_5	R_6	R_7	$R_{3'}$	$R_{4'}$	
15	H	OH	OCH_3	OCH_3	H	OH	91
16	OH	OH	H	OCH_3	H	OH	21
17	OH	OH	H	OH	OH	OH	88

[a]Experimental values of aldose reductase inhibition percentage were taken from ref. 18.

Figure 1. Structural formulae of coumarin and flavonoid derivatives.

Ordered Orthogonalized MLR Method. The orthogonalization of I selected descriptors is carried out in all possible orthogonalization orderings (there are $I!$ of them). In this case the correlation coefficients between activity A and each single orthogonalized descriptor change, because with the change of the order of orthogonalization the individual orthogonalized descriptors are also changed. However, the total correlation coefficient between A and its computed value A' remains the same for each order with the same I-tuple. Among all possible orderings, we select

those which maximize the contribution of 1, 2, ..., $I - 1$ (significant) descriptors, and which at the same time minimize the contributions of $I - 1$, $I - 2$, ..., *1* (insignificant) descriptors. Insignificant descriptors can be removed from the ordered orthogonalized MLR model, since their removal leads to the (much) more stable model with the smaller number of descriptors.[6−8] It is important to note that, as a result of the QSAR modeling, the ordered orthogonalized MLR method gives not only the linear regression coefficient for each individual significant orthogonalized descriptor but also the orthogonalization ordering.

Statistical Studies. Statistical characteristics of the models in the training procedure were calculated using correlation coefficient R, standard errors of estimate (S_f, S_d), and the results of the F-test. S_d was calculated using eq 1

$$S_d = \sqrt{\frac{\sum_{i=1}^{N}(a_i - a_i')^2}{(N - I - 1)}} \qquad (1)$$

where a_i and a_i' denote the experimental and calculated activity; N and I denote the total number of cases (molecules) and the number of descriptors in the model, respectively. S_f was calculated using eq 1 when I was fixed to 1. The prediction abilities of both the nonorthogonalized and ordered orthogonalized MLR methods were calculated using the leave-one-out cross-validation technique and marked by the cross-validation correlation coefficient R_{cv} and standard error of estimate S_{cv}. S_{cv} is different from S_f because we put $(a_i')_{cv}$ instead of a_i'. We also considered it useful to list the standard errors of all coefficients in the model, because from these data it is possible to see the level of the significance and stability of the individual descriptors.

Ordered Orthogonalized MLR Prediction Procedure. In order to predict activity (inhibition) of compounds not used in the training of the algorithm, one has first to calculate molecular descriptors needed to obtain the model in the orthogonalized basis. The numerical values of these descriptors are then added to the pool of descriptors which were used to produce the initial model. The next is the orthogonalization of all descriptors in that ordering that was selected as the best in the training process when the search for the best model was carried out. In this way the values of the added descriptors are transformed according the rule established in the training process. Then the transformed values of descriptors are inserted into the ordered orthogonalized MLR model obtained in the training process, and the activities of studied compounds are finally predicted.

RESULTS AND DISCUSSION

The first step in the procedure is the descriptor selection for the QSAR model. From the initial descriptor pools is selected the final set of seven favorable nonorthogonal descriptors for studied compounds. Selected nonorthogonal descriptors are subjected to a QSAR analysis to yield a linear regression equation that best describes the inhibitory activity of the compounds in the data base.

The best fit equations (models) were selected on the basis of the highest correlation coefficients and the lowest standard deviations. Starting from 31 descriptors we produced the best possible models with I = 1, ..., *7* descriptors (for example: for $I = 7$, there are 31!/24! × 7! possible models, and among them we selected the best one). The best models are listed in Table 1.

Many factors can affect the potency of coumarin and flavonoid derivatives as AR inhibitors, e.g., structural features, electronic properties of molecules, substitution pattern of molecules, etc.

Some possible relationships of structure to the inhibitory activities of coumarin and flavonoid derivatives are reported.[26−29] However, no unambiguous conclusions about the effect of substitution pattern on the inhibitory potency of coumarin and flavonoid derivatives can be found in the literature. Previous studies indicate that the structural requirements of an AR inhibitor are more general than specific. The main feature of the structure of coumarin and flavonoid derivatives influencing activities seems to be the hydroxylation pattern. Minor modifications in the hydroxylation pattern and the degree of methylation can affect the inhibitory activity in a rather unpredictable way. Although a clear relationship between the structure and inhibitory potency of the coumarin and flavonoid derivatives was not established, some structural features necessary for inhibition may be summarized as follows:

- The effective inhibitors of AR are the hydroxylated derivatives.
- The position and number of the hydroxyl groups are important. The presence of hydroxyl groups at positions C-5 and C-7 in the A ring is necessary for inhibition. Hydroxylation at both the C-3′ and C-4′ positions in the B ring help augment the inhibitory activity.
- Marked activity is associated with the presence of the hydroxyl or methoxyl groups in positions C-6 and C-7.
- The degree of methylation of hydroxyl groups affects the inhibitory potency.
- The planarity of the benzopyrone system seems to be a requirement for inhibition.

A reasonable explanation why the pattern of hydroxylation is so important for the inhibitory activity could be stated as follows. The inhibitory activities of coumarin and flavonoid derivatives result from their ability to bind to the enzyme. Hydroxyl groups may be involved in the formation of hydrogen bridges between the coumarin or flavonoid derivative and the amino acid located near or at the active site on the enzyme molecule. Degree of methylation probably cause steric hindrance in the essential structural feature of inhibitor molecule. In addition to the steric hindrance brought about by methylations of hydroxyl groups it is suggestive that the resultant elimination of the capacity for proton bonding of the hydroxyl groups exerts an even greater influence on the capacity of the substances to bind with the enzyme.

Our recently introduced approach, based on the consideration of all possible orthogonalization orderings and dominant descriptor analysis, improves the predictivity of QSAR models.[6] The procedure is fully documented in our recent reports.[7,8] The approach results in obtaining QSAR models with orthogonalized indices with better statistical parameters than the models based on the nonorthogonal basis. The best ordered orthogonalized models for studied molecules with their statistical characteristics are listed in Table 2.

To make the prediction of the inhibitory activity of the favorable coumarin and flavonoid derivatives we tried to test

Table 1. The Best Possible Multivariate Linear Regression Models (with I Descriptors) for Inhibition of Aldose Reductase with 14 Coumarin Derivatives and Three Flavonoid Derivatives, Obtained Using the Nonorthogonalized Descriptors[a]

$I = 1$
$R = 0.9313$, $S_f = 11.11$, $S_d = 11.11$, $F = 98.06$
$R_{cv} = 0.9015$, $S_{cv} = 13.24$
%INH = $(15.923 \pm 3.4) + (61.098 \pm 6.2)$ d9

$I = 2$
$R = 0.9555$, $S_f = 9.00$, $S_d = 9.32$, $F = 73.35$
$R_{cv} = 0.9295$, $S_{cv} = 11.29$
%INH = $(-731.487 \pm 276.2) + (61.545 \pm 22.7)$d1 + (54.756 ± 5.7)d9

$I = 3$
$R = 0.9640$, $S_f = 8.11$, $S_d = 8.71$, $F = 56.94$
$R_{cv} = 0.9354$, $S_{cv} = 10.83$
%INH = $(-890.097 \pm 274.0) + (-12.367 \pm 7.1)$d11 + (75.459 ± 22.7)d1 + (54.949 ± 5.3)d9

$I = 4$
$R = 0.970$, $S_f = 7.42$, $S_d = 8.29$, $F = 47.73$
$R_{cv} = 0.9365$, $S_{cv} = 10.77$
%INH = $(-5.518 \pm 8.8) + (0.027 \pm 0.01)$d3 + (-4.428 ± 1.8)d8 + (18.457 ± 7.1)d10 + (58.133 ± 4.8) d9

$I = 5$
$R = 0.9750$, $S_f = 6.78$, $S_d = 7.92$, $F = 42.33$
$R_{cv} = 0.9453$, $S_{cv} = 10.01$
%INH = $(-4.324 \pm 8.4) + (0.080 \pm 0.04)$d3 + (-2.615 ± 1.8)d6 + (-8.752 ± 3.4)d8 + (22.957 ± 7.5)d10 + (55.948 ± 4.8)d9

$I = 6$
$R = 0.9771$, $S_f = 6.49$, $S_d = 7.95$, $F = 35.14$
$R_{cv} = 0.9368$, $S_{cv} = 10.72$
%INH = $(-6.860 \pm 9.9) + (0.343 \pm 0.16)$d2 + (-3.394 ± 1.98)d6 + (-7.289 ± 3.2)d8 + (27.089 ± 8.3)d10 + (2.728 ± 2.2)d4 + (49.047 ± 8.0)d9

$I = 7$
$R = 0.9881$, $S_f = 4.69$, $S_d = 6.05$, $F = 53.10$
$R_{cv} = 0.9654$, $S_{cv} = 8.10$
%INH = $(-526.435 \pm 141.4) + (0.108 \pm 0.04)$d3 + (52.812 ± 14.2)d5 + (-17.055 ± 4.3)d6 + (-7.398 ± 2.0)d7 + (-15.888 ± 3.4)d8 + (-49.72788 ± 10.6)d11 + (57.01196 ± 4.2)d9

descriptor	explanation
d1	Σ_q − sum of the π charges in the coumarin or chromone moiety
d2	W − Wiener number[20]
d3	MTI − molecular topological index[22]
d4	DM − dipole moment[25]
d5	p_1 − number of the paths of length 1
d6	p_4 − number of the paths of length 4
d7	p_7 − number of the paths of length 7
d8	p_8 − number of the paths of length 8
d9	for coumarin derivatives value of 1 for 7-OH (without 8-R) or 4-OH compounds; value of 0.25 for 5,7-diOCH_3 or 4-OCH_3 compounds and elsewhere 0. For flavonoid derivatives value of 1 for 7,3′,4′-diOH or 6,7-diOH (diOCH_3) compounds; and elsewhere 0
d10	value of 1 for derivatives with substituent in position C-4; and elsewhere 0
d11	value of 1 for derivatives with substituent in position C-7; and elsewhere 0

[a] I = the number of descriptors in model; %INH = calculated percentage of AR inhibition; R = correlation coefficient; S_f and S_d = standard errors of estimate; F = F-test; R_{cv} and S_{cv} = leave-one-out cross-validation correlation coefficient and standard error of estimate (see methods—Statistical Studies).

several models. Hence, we produced the ordered orthogonalized MLR models (with three and four significant descriptors) by orthogonalizing five descriptors (d3, d6, d8, d10, and d9) from the Table 1 which give the best possible MLR model with five nonorthogonalized descriptors (part A in Table 2).

We also tested another set of seven descriptors (d3, d6, d8, d10, d11, d1, and d9) from Table 1 which appears in models with nonorthogonalized descriptors from $I = 1$ to $I = 5$. The models with 2, 3, 4, and 5 significant descriptors are the best according to R_{cv} (see part B in Table 2).

The third set considered for creating the models based on orthogonalized descriptors was made up from seven nonorthogonalized descriptors from Table 1 (d3, d5, d6, d7, d8, d11, and d9) which give the best possible model with seven nonorthogonalized descriptors. We singled out the best models with 3, 4, 5, and 6 significant descriptors (part C in Table 2).

It can be seen from Table 2 that all orthogonalized MLR models are better than the best possible nonothogonalized MLR models (Table 1) according the R, S_f, S_d, and F-test criteria, but also according to the most important statistical parameters: the leave-one-out cross-validation correlation coefficient R_{cv} and standard error of estimate S_{cv}.

How To Interpret the Ordered Orthogonalized MLR Models? We are after the best approximation A' of the activity A in the space which is spanned by I nonorthogonal descriptors. The set of I nonorthogonal descriptors by which we wish to approximate the activity A, which, if they are not linearly dependent, span in general the space of dimension I. By orthogonalization in the $I!$ orderings, we rotate the space spanned by I descriptors toward the fixed vector A (activity) such that it is possible to detect that position at which A is well represented by k orthogonal significant descriptors ($k < I$). Remaining ($I - k$) descriptors are not

Table 2. The Best Possible QSAR Models with *I*-Tuples of Orthogonal Descriptors Obtained by Selecting the Optimum Orthogonalization Orderings of Descriptors[a]

Part A. Orthogonalization of Descriptors: d3, d6, d8, d10, d9

$I = 3$, orthogonalization ordering: ***d9 d10*** d3 d6 ***d8***
$R = 0.9702$, $S_f = 7.39$, $S_d = 7.94$, $F = 69.52$
$R_{cv} = 0.9443$, $S_{cv} = 10.12$
$\%INH = (36.588 \pm 1.9) + (61.098 \pm 4.4)\Omega9 + (-14.886 \pm 4.7)\Omega10 + (-8.75670 \pm 3.4)\Omega8$

$I = 4$, orthogonalization ordering: ***d9 d6*** d3 ***d10 d8***
$R = 0.9736$, $S_f = 6.97$, $S_d = 7.79$, $F = 54.50$
$R_{cv} = 0.9471$, $S_{cv} = 9.87$
$\%INH = (36.588 \pm 1.9) + (61.098 \pm 4.3)\Omega9 + (-0.583 \pm 0.2)\Omega6 + (-13.800 \pm 6.4)\Omega10 + (-8.757 \pm 3.4)\Omega8$

Part B. Orthogonalization of Descriptors: d3, d6, d8, d10, d11, d1, d9

$I = 2$, orthogonalization ordering: ***d9*** d11 ***d1*** d3 d6 d8 d10
$R = 0.9630$, $S_f = 8.22$, $S_d = 8.51$, $F = 89.28$
$R_{cv} = 0.9363$, $S_{cv} = 10.76$
$\%INH = (36.588 \pm 2.1) + (61.098 \pm 4.7)\Omega9 + (75.461 \pm 22.2)\Omega1$

$I = 3$, orthogonalization ordering: ***d9*** d11 ***d1*** d3 d6 d10 ***d8***
$R = 0.9747$, $S_f = 6.8$, $S_d = 7.3$, $F = 82.2$
$R_{cv} = 0.9526$, $S_{cv} = 9.36$
$\%INH = (36.588 \pm 1.8) + (61.098 \pm 4.1)\Omega9 + (75.462 \pm 19.1)\Omega1 + (-7.945 \pm 3.3)\Omega8$

$I = 4$, orthogonalization ordering: ***d9 d1*** d6 d8 ***d10*** d11 ***d3***
$R = 0.9759$, $S_f = 6.65$, $S_d = 7.43$, $F = 60.12$
$R_{cv} = 0.9529$, $S_{cv} = 9.33$
$\%INH = (36.588 \pm 1.803) + (61.098 \pm 4.1)\Omega9 + (-61.548 \pm 18.1)\Omega1 + (14.650 \pm 6.8)\Omega10 + (0.082 \pm 0.04)\Omega3$

$I = 5$, orthogonalization ordering: ***d9 d3 d10*** d11 ***d1*** d6 ***d8***
$R = 0.9777$, $S_f = 6.40$, $S_d = 7.47$, $F = 47.78$
$R_{cv} = 0.9487$, $S_{cv} = 9.75$
$\%INH = (36.588 \pm 1.8) + (61.100 \pm 4.2)\Omega9 + (-0.0056 \pm 0.002)\Omega3 + (13.748 \pm 6.2)\Omega10 + (60.006 \pm 26.1)\Omega1 + (-7.946 \pm 3.3)\Omega8$

Part C. Orthogonalization of Descriptors: d3, d5, d6, d7, d8, d11, d9

$I = 3$, orthogonalization ordering: ***d9*** d11 ***d6*** d3 d5 d7 ***d8***
$R = 0.9828$, $S_f = 5.63$, $S_d = 6.04$, $F = 123.08$
$R_{cv} = 0.9688$, $S_{cv} = 7.83$
$\%INH = (36.589 \pm 1.5) + (61.098 \pm 3.4)\Omega9 + (0.720 \pm 0.2)\Omega6 + (-15.8881 \pm 3.4)\Omega8$

$I = 4$, orthogonalization ordering: ***d9*** d11 ***d5 d7*** d3 d6 ***d8***
$R = 0.9869$, $S_f = 4.91$, $S_d = 5.50$, $F = 112.55$
$R_{cv} = 0.9738$, $S_{cv} = 7.30$
$\%INH = (36.589 \pm 1.3) + (61.098 \pm 3.1)\Omega9 + (1.552 \pm 0.4)\Omega5 + (1.794 \pm 0.7)\Omega7 + (-15.888 \pm 3.1)\Omega8$

$I = 5$, orthogonalization ordering: ***d9 d7*** d8 ***d5 d3*** d6 ***d11***
$R = 0.9880$, $S_f = 4.70$, $S_d = 5.49$, $F = 90.28$
$R_{cv} = 0.9746$, $S_{cv} = 7.15$
$\%INH = (36.589 \pm 1.3) + (61.0974 \pm 3.1)\Omega9 + (-0.522 \pm 0.2)\Omega7 + (-4.019 \pm 1.4)\Omega5 + (0.071 \pm 0.03)\Omega3 + (-49.727 \pm 9.6)\Omega11$

$I = 6$, orthogonalization ordering: ***d9 d7*** d8 ***d5 d11 d6 d3***
$R = 0.9881$, $S_f = 4.69$, $S_d = 5.74$, $F = 68.82$
$R_{cv} = 0.9708$, $S_{cv} = 7.56$
$\%INH = (36.589 \pm 1.4) + (61.097 \pm 3.2)\Omega9 + (-0.522 \pm 0.2)\Omega7 + (-4.019 \pm 1.4)\Omega5 + (-17.759 \pm 5.4)\Omega11 + (11.286 \pm 3.6)\Omega6 + (0.108 \pm 0.03)\Omega3$

[a] Boldface italic letters denote those descriptors (e.g., dominant descriptors) which, after the orthogonalization in the indicated order, take part in the construction the model. Orthogonal descriptors are denoted by Greek character Ω. For explanation of *I*, %INH, *R*, S_f, S_d, *F*, R_{cv}, and S_{cv} see footnote in Table 1.

significant, that is, they are nearly orthogonal on *A*. In this way we design the QSAR model of considerable predictive power with the limited number of parameters (in the space of smaller dimension than the initial space of dimension *I*).

Predictions for 34 favorable coumarin and 30 flavonoid derivatives with all ordered orthogonalized MLR models from Table 2 are similar, although these are models with different descriptors. This indicates that the models obtained are dependable and stable. These predictions are also congruent with the highly predictive nonorthogonalized models from Table 1 (models with 2, 3, 4, 5, and 7 descriptors). The agreement with the nonorthogonalized models with 1 and 6 descriptors is poorer, because this model is found to be less reliable and unstable as it can be seen from the statistical parameters for the model.

Among the studied molecules we detected four favorable coumarin and four flavonoid derivatives with the highest predicted value of inhibition. These compounds are shown in Figure 2. Their predicted values of inhibition are larger than those of compounds used in the training set (Figure 1).

CONCLUDING REMARKS

The results presented in this report suggest the possibility that a relatively simple QSAR model, based on molecular descriptors of coumarin and flavonoid derivatives, can be used to predict successfully the percent inhibition of AR.

In order that the predictions are dependable it is necessary that in the QSAR modeling is used the most reliable method. Our investigations reveal that the best possible nonorthogonalized MLR models, and especially ordered orthogonalized MLR models, are more dependable than models obtained by the use of other methods. We reached this conclusion owing to the development of the algorithm by which we were able to detect, even in the case of the large number of molecules with the large number of descriptors, the best MLR models with nonorthogonalized descriptors. The application

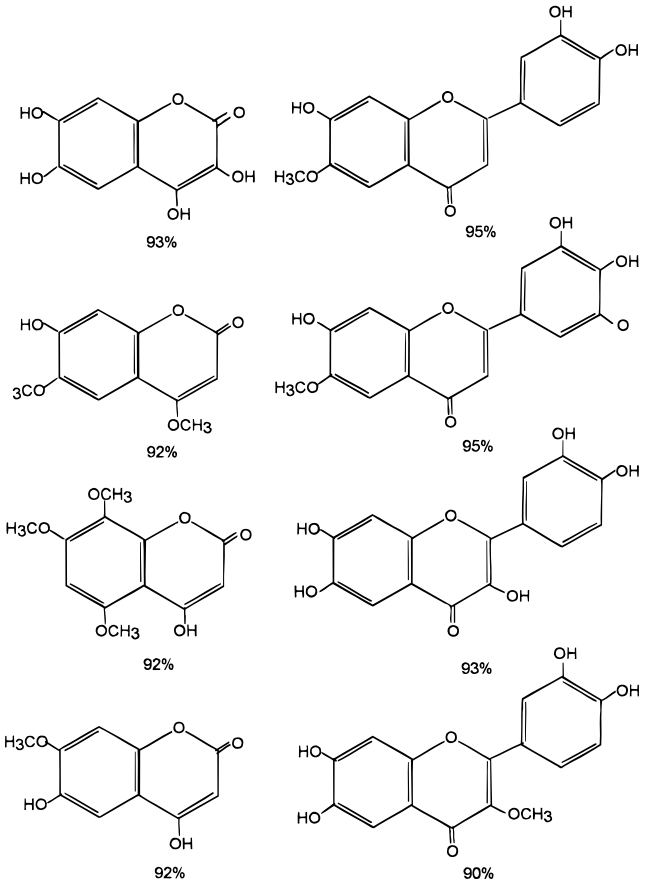


Figure 2. Structural formulas of favorable aldose reductase (AR) inhibitors. Values under each structure are AR inhibitory activity (%INH) calculated as an average of predicted %INH values obtained with all models from Tables 1 and 2, except models with $I = 1$ and $I = 6$ from Table 1.

of the ordered orthogonalized MLR algorithms on the nonorthogonalized MLR models leads to essentially better QSAR models. We observed this in all cases studied so far, some of which are already published and some are being prepared for publication.[6−8,14,15] Based on the above, we conclude that the ordered orthogonalized MLR method is a novel powerful QSAR approach.

Additional advantage of the MLR models is in that these models are interpretable. This is not the case with PCA, PLS, or NN models. The interpretability of the model facilitate the predictability of the model, since it is clearly seen the contribution of each descriptor. This is important because we seek a compound with values of descriptor(s) in the range of active compounds. Thus, the introduced QSAR methodology can be used to design new molecules with the purpose of obtaining more potent activity against AR.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Technology of the Republic of Croatia through Grants 079301 (D.A., D.D.A., D.B.) and 00980605 (B.L., N.T.). We are thankful to the reviewers for their constructive comments.

REFERENCES AND NOTES

(1) For a review, see: Middleton, E. Jr.; Kandaswami, C. The Impact of Plant Flavonoids on Mammalian Biology: Implications for Immunity, Inflammation and Cancer. In *The Flavonoids: Advances in Research Since 1986*; Harborne, J. B., Ed.; Chapman and Hall: London; 1993; pp 619−652.

(2) Mabry, T. J.; Ulubelen, A. Chemistry and Utilization of Phenylpropanoids Including Flavonoids, Coumarins, and Lignans. *J. Agric. Food Chem.* **1980**, *28*, 188−196.

(3) Okada, Y.; Miyauchi, N.; Suzuki, K.; Kobayashi, T.; Tsutsui, C.; Mayuzumi, K.; Okuyama, T. Search for Naturally Occurring Substances for Prevention against the Complications of Diabetes. Inhibitory Effect on Aldose Reductase and Platelet Aggregation. *Nat. Med.* **1994**, *48*, 324−329.

(4) Varma, S. D.; Mizuno, A.; Kinoshita, J. H. Diabetic Cataracts and Flavonoids. *Science* **1977**, *195*, 205−206.

(5) Dvornik, D. A Perspective of Aldose Reductase Inhibitors and Diabetic Complications. *Croat. Chem. Acta* **1996**, *69*, 613−630.

(6) 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.

(7) Lučić, B.; Nikolić, S.; Trinajstić, N.; Jurić, A.; Mihalić, Z. A Structure-Property Study of the Solubility of Aliphatic Alcohols in Water. *Croat. Chem. Acta* **1995**, *68*, 417−434.

(8) Lučić, B.; Nikolić, S.; Trinajstić, N.; Juretić, D.; Jurić, A. A Novel QSPR Approach to Physicochemical Properties of the α-Amino Acids. *Croat. Chem. Acta* **1995**, *68*, 435−450.

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

(10) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; McGraw-Hill: New York, 1989.

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

(12) Amić D.; Davidović-Amić, D.; Jurić, A.; Lučić, B.; Trinajstić, N. Structure−Activity Correlation of Flavone Derivatives for Inhibition of cAMP Phosphodiesterase. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1034−1038.

(13) Basak, S. C.; Niemi, G. J.; Veith, G. D. Predicting Properties of Molecules Using Graph Invariants. *J. Math. Chem.* **1991**, *7*, 243−272.

(14) (a) Lučić, B.; Trinajstić, N.; Juretić, D.; Grunwald, G. D.; Basak, S. C. MLR excels PCA. Book of Abstracts of The 10th Dubrovnik International Course & Conference on the Interfaces among Mathematics, Chemistry and Computer Science, Dubrovnik, Croatia, 1995. (b) Lučić, B.; Trinajstić, N. QSAR/QSPR Modelling: MLR Excels PLS Algorithm. Book of Abstracts of The 11th Dubrovnik International Course & Conference on the Interfaces among Mathematics, Chemistry and Computer Science, Dubrovnik, Croatia, 1996.

(15) Lučić, B.; Trinajstić, N. Unpublished results.

(16) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

(17) Pogliani, L. Molecular Connectivity Descriptors of the Physicochemical Properties of α-Amino Acids. *J. Phys. Chem.* **1994**, *98*, 1494−1499.

(18) Okada, Y.; Miyauchi, N.; Suzuki, K.; Kobayashi, T.; Tsutsui, C.; Mayuzumi, K.; Nishibe, S.; Okuyama, T. Search for Naturally Occurring Substances to Prevent the Complications of Diabetes. II. Inhibitory Effect of Coumarin and Flavonoid Derivatives on Bovine Lens Aldose Reductase and Rabbit Platelet Aggregation. *Chem. Pharm. Bull.* **1995**, *43*,1385−1387.

(19) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; John Wiley and Sons, Inc.: New York, 1986.

(20) Wiener, H. Prediction of Isomeric Differences in Paraffin Proprties. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(21) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235−250.

(22) (a) Schultz, H. P. Topological Organic Chemistry. 1. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227−228. (b) Nikolić, S.; Trinajstić, N.; Mihalić, Z. Molecular Topological Index: An Extension to Heterosystems. *J. Math. Chem.* **1993**, *12*, 251−264.

(23) Zhou, Z.; Parr, R. G. New Measures of Aromaticity: Absolute Hardness and Relative Hardness. *J. Am. Chem. Soc.* **1989**, *111*, 7371−7379.

(24) Gutman, I.; Milun, M.; Trinajstić, N. Graph Theory and Molecular Orbitals. 19. Nonparametric Resonance Energies of Arbitrary Conjugated Systems. *J. Am. Chem. Soc.* **1977**, *99*, 1692−1704.

(25) Herman, A.; Blonski, S. Software Package SAA-012-S88. *Spectrochim. Acta* **1988**, *44A* (8).

(26) Varma, S. D.; Kinoshita, J. H. Inhibition of Lens Aldose Reductase by Flavonoids -Their Possible Role in the Prevention of Diabetic Cataracts. *Biochem. Pharmacol.* **1976**, *25*,2505−2513.

(27) Okuda, J.; Miwa, I.; Inagaki, K.; Horie, T.; Nakayama, M. Inhibition of Aldose Reductase by 3′,4′-Dihydroxyflavones. *Chem. Pharm. Bull.* **1984**, *32*, 767−772.

(28) Shimizu, M.; Ito, T.; Terashima, S.; Hayashi, T.; Arisawa, M.; Morita, N.; Kurokawa, S.; Ito, K.; Hashimoto, Y. Inhibition of Lens Aldose Reductase by Flavonoids. *Phytochemistry* **1984**, *23*, 1885−1888.

(29) Iinuma, M.; Tanaka, T.; Mizuno, M.; Katsuzaki, T.; Ogawa, H. Structure−Activity Correlation of Flavonoids for Inhibition of Bovine Lens Aldose Reductase. *Chem. Pharm. Bull.* **1989**, *37*, 1813−1815.

CI960158W