Chemometrics and Intelligent Laboratory Systems 109 (2011) 94–100

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Short Communication

Comparison of SMILES and molecular graphs as the representation of the molecular structure for QSAR analysis for mutagenic potential of polyaromatic amines

A.A. Toropov [a,*], A.P. Toropova [a], S.E. Martyanov [b], E. Benfenati [a], G.Gini [c], D. Leszczynska [d], J. Leszczynski [e]

[a] Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy
[b] Teleca OOO, 603093, 23, Rodionova st, Nizhny Novgorod, Russia
[c] Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy
[d] Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA
[e] Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA

ARTICLE INFO

Article history:
Received 6 June 2011
Received in revised form 13 July 2011
Accepted 28 July 2011
Available online 5 August 2011

Keywords:
QSAR
Optimal descriptor
Monte Carlo method
Mutagenicity

ABSTRACT

Optimal descriptors calculated with simplified molecular input line entry system (SMILES), hydrogen-suppressed molecular graph (HSG), hydrogen-filled molecular graph (HFG), and graph of atomic orbitals (GAO) have been studied as a basis to build up models for mutagenicity of polyaromatic amines. The optimal descriptors are calculated with correlation weights of the molecular fragments. In the case of the molecular graph, chemical elements (C, N, O, etc.) or their electronic structure (1s2, 2p3, 3d10, etc.) together with their Morgan vertex degrees are the basis for calculation of the descriptor. In the case of SMILES, chemical elements (C, O, N, etc.) together with presence of cycles (1, 2, 3, etc.), cis-, trans- isomerism ('\' and '/') and other are the basis for calculation of the descriptor. In both these cases, descriptors are a mathematical function of the correlation weights of the above-mentioned molecular features. The correlation weights are calculated by the Monte Carlo optimization (the target function is the correlation coefficient between experimental and predicted endpoint values). SMILES-based optimal descriptors have shown the preferable predictive ability. The CORAL software (http://www.insilico.eu/coral/) was used to build up models of the mutagenic potential as the function of the molecular structure. Analysis of three probes of the Monte Carlo optimization with six random splits has shown there are three kinds of the molecular features encoded by SMILES attributes: promoters of increase/decrease of mutagenic potential and ones without defined role.

1. Introduction

Mutagenicity is a toxicity endpoint associated with the chronic exposure to chemicals. There is a similarity of the mutagenicity mechanism with carcinogenicity mechanisms of action. The mutagenicity can be used for detection of substances, potentially hazardous to human health [1].

Quantitative structure–property/activity relationships (QSPR/QSAR) can be useful in praxis of the risk assessment of large sets of various substances [1–10]. Mutagenicity is an important factor in assessing the hazardous effects of chemicals on both human and environmental health. Therefore, it is not surprising that various authors have attempted to predict mutagenicity of chemicals from their structure.

According to Organisation for Economic Co-operation and Development (OECD) principles [11] QSAR model should be associated with the following information:

1) a defined endpoint
2) an unambiguous algorithm
3) a defined domain of applicability
4) appropriate measures of goodness-of-fit, robustness and predictivity
5) a mechanistic interpretation, if possible.

The main reason for the criticisms [12–14] of QSPR/QSAR is poor predictions for external compounds. Probably, the reasons for the criticisms can be reduced if researchers will be concentrated on the statistical quality and reproducibility of QSAR model for external compounds. Apparently, there are good splits into the training and test set (with a very good model that is obtained by a suggested approach) and there are other splits where the suggested approach can give a modest or even poor model.

The aim of this study is the comparison of QSAR obtained with the optimal SMILES-based descriptors and the graph-based descriptors in

* Corresponding author.
E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2011.07.008

the modelling of the mutagenic potential of aromatic amines. Apparently, both the graph and SMILES are the representation of the molecular structure (Fig. 1). Thus, the comparison of the above-mentioned representations is interesting at least from heuristic point of view.

2. Method

2.1. Data

Data on mutagenic potentials of the set of 95 aromatic and heteroaromatic amines were taken from Ref. [15]. The mutagenic activity in *Salmonella typhimurium* TA98 + S9 microsomal reparation is expressed as the natural logarithm of R, where R is the number of revertants per nanomole.

2.2. Descriptors

Optimal graph-based and SMILES-based descriptors have been examined in this study. The graph-based descriptors were calculated as the following

$$DCW(T) = \sum CW(GA_k) \quad (1)$$

where GA_k is an attribute of k-th vertex in graph, i.e. vertex type and vertex degree, $CW(GA_k)$ is the correlation weights of the GA_k. Three kinds of the graph were studied: (1) hydrogen-suppressed graph (HSG, Fig. 1), (2) hydrogen-filled graph (HFG, Fig. 2), and (3) graph of atomic orbitals (GAO, Fig. 3). Vertices in the HSG and HFG are representations of chemical elements and edges are representations of the chemical bonds in the molecule. Vertices in the GAO are representations of atomic orbitals. The algorithm of the translation of the HFG into GAO is described in Ref [16]. From HSG one can obtain the GAO without $1s^2$ vertices (which are the representation of hydrogen; from HFG one can obtain the GAO which involves $1s^2$ vertices.

Three topological invariants of the molecular graphs have been involved in the study: vertex degree (EC0), extended connectivity of first order (EC1), and extended connectivity of second order (EC2) [16]. Fig. 4 contains examples of calculations of the EC1 and EC2.

Optimal SMILES-based descriptors were calculated as the following

$$DCW(T) = \alpha\sum CW(S_k) + \beta\sum CW(SS_k) + \gamma\sum CW(SSS_k) \quad (2)$$

where S_k, SS_k, SSS_k are SMILES attributes which contain one-, two-, and three SMILES elements respectively; $CW(S_k)$, $CW(SS_k)$, and $CW(SSS_k)$ are the correlation weights of the SMILES attributes; α, β, and γ are coefficients which can be 1 or 0, one can select model which is based on the attributes of one-element (α = 1, β = 0,γ = 0), or model based on the S_k and SS_k (α = 1, β = 1,γ = 0), etc. T is threshold (1,2,…5). These values are used to classify the molecular attributes into two categories, i.e. rare and not rare (active). The attribute which is defined as rare has the correlation weight equal to zero, consequently, the attribute has no influence for the modelling process.

The numerical data for the correlation weights are calculated with the Monte Carlo method optimization which provides maximal value of correlation coefficient between experimental and calculated lnR for the training set (classic scheme) or maximal value of the balance of correlation with ideal slopes [17,18] or without ideal slopes [19].

The predictive potentials of the graph-based model and the SMILES-based model are mathematical functions the threshold (T) and the number of epochs (Nepoch) of the Monte Carlo optimization (Fig. 5). One can find the most predictive combination of T and N_{epoch} values for a split (Fig. 6). Having data on the calculation for several splits, one can estimate average predictive potential of the model. Apparently, the average values of statistical characteristics of the model will be poorer than ones for a 'good' split, but it will be more reliable information.

3. Results and discussion

Table 1 contains statistical quality of the models obtained with molecular graphs (HSG, HFG, and GAO) using the extended connectivity of zero-, first-, and second orders. The extended connectivity of first order seems preferable version of the optimal descriptor for all types of graphs (in fact best average correlation coefficient value (external test set) for HSG was obtained with EC0 (08441), but for the EC1 of HSG the values is very similar (0.8422). We have selected model based on the EC0 in the HSG as the best graph-based model.

Table 2 contains statistical quality of the models obtained with SMILES. One can see that the best average correlation coefficient (external test set) takes place for version descriptor calculated with Eq. (2) using combination of α = 1, β = 0, γ = 0. These models have good predictability according to criterion R^2_m[20,21] which should be larger than 0.5. The SMILES-based models are the following:

Split 1

$$\ln R = 0.0033(\pm 0.0288) + 0.4321(\pm 0.0057)*DCW(2) \quad (3)$$

$$n = 42, r^2 = 0.7245, q^2 = 0.7007, s = 1.15, F = 105 \text{ (sub−training set)}$$

$$n = 25, r^2 = 0.8257, r^2_{pred} = 0.7877, s = 0.707, F = 109 \text{ (calibration set)}$$

$$n = 28, r^2 = 0.8478, r^2_{pred} = 0.8203, R^2_m = 0.8008, s = 0.700, F = 14 \text{ (test set)}$$

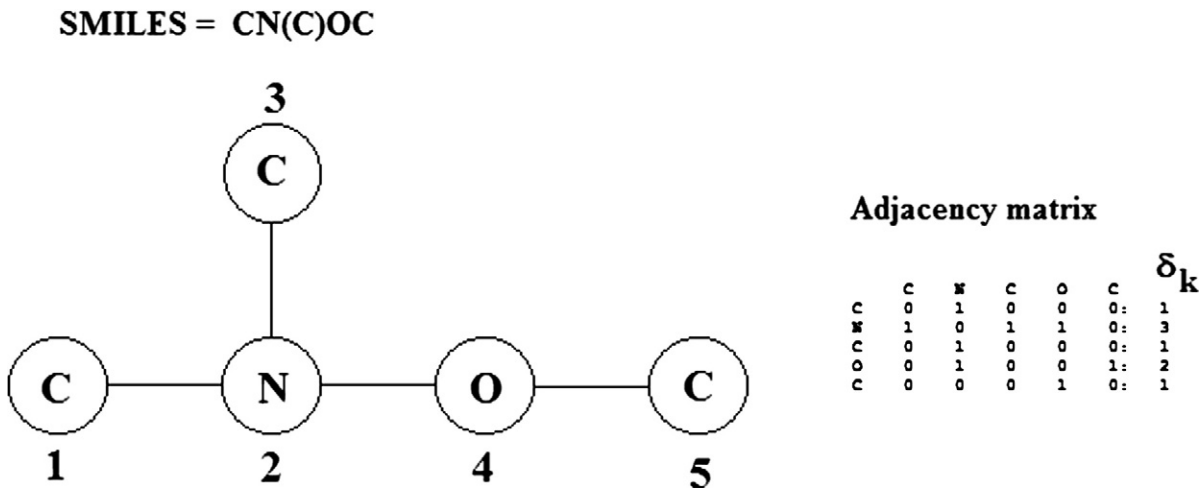
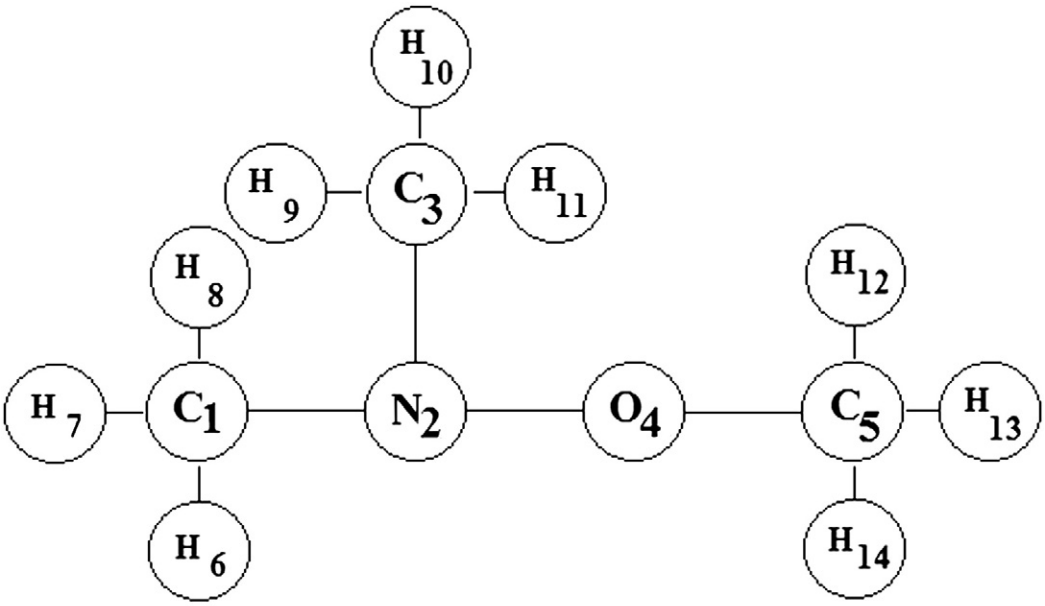


Fig. 1. Example of SMILES and the molecular graph with suppressed hydrogen atoms (HSG).



Adjacency matrix δ_k

	C	N	C	O	C	H	H	H	H	H	H	H	H	H	
C	0	1	0	0	0	1	1	1	0	0	0	0	0	0:	4
N	1	0	1	1	0	0	0	0	0	0	0	0	0	0:	3
C	0	1	0	0	0	0	0	0	1	1	1	0	0	0:	4
O	0	1	0	0	1	0	0	0	0	0	0	0	0	0:	2
C	0	0	0	1	0	0	0	0	0	0	0	1	1	1:	4
H	1	0	0	0	0	0	0	0	0	0	0	0	0	0:	1
H	1	0	0	0	0	0	0	0	0	0	0	0	0	0:	1
H	1	0	0	0	0	0	0	0	0	0	0	0	0	0:	1
H	0	0	1	0	0	0	0	0	0	0	0	0	0	0:	1
H	0	0	1	0	0	0	0	0	0	0	0	0	0	0:	1
H	0	0	1	0	0	0	0	0	0	0	0	0	0	0:	1
H	0	0	0	0	1	0	0	0	0	0	0	0	0	0:	1
H	0	0	0	0	1	0	0	0	0	0	0	0	0	0:	1

Fig. 2. The molecular graph with hydrogen atoms (HFG).

Split 2

$$\ln R = -1.160(\pm 0.0309) + 0.3369(\pm 0.0061)*DCW(5) \quad (4)$$

$n = 42, r^2 = 0.6654, q^2 = 0.6311, s = 1.08, F = 80$ (sub−training set)

$n = 25, r^2 = 0.8257, r^2_{pred} = 0.7921, s = 0.830, F = 109$ (calibration set)

$n = 28, r^2 = 0.8544, r^2_{pred} = 0.8228, R^2_m = 0.7881, s = 0.829,$ $F = 153$(test set)

Split 3

$$\ln R = -2.934(\pm 0.0598) + 0.1925(\pm 0.0036)*DCW(3) \quad (5)$$

$n = 43, r^2 = 0.6376,\ q^2 = 0.6014,\ s = 1.14,\ F = 72$ (sub−training set)

$n = 25,\ r^2 = 0.8181,\ r^2_{pred} = 0.7864,\ s = 0.869,\ F = 103$ (calibration set)

$n = 27, r^2 = 0.8904, r^2_{pred} = 0.8753, R^2_m = 0.8598, s = 0.700,$ $F = 203$(test set)

Split 4

$$\ln R = -4.576(\pm 0.0698) + 0.1756(\pm 0.0031)*DCW(3) \quad (6)$$

$n = 46, r^2 = 0.6540, q^2 = 0.6201, s = 1.17, F = 83$(sub−training set)

$n = 24, r^2 = 0.7320, r^2_{pred} = 0.6877, s = 0.993, F = 60$ (calibration set)

$n = 25, r^2 = 0.8922, r^2_{pred} = 0.8773, R^2_m = 0.8707, s = 0.629,$ $F = 203$(test set)

Split 5

$$\ln R = 0.0008(\pm 0.0233) + 0.3673(\pm 0.0050)*DCW(3) \quad (7)$$

$n = 50, r^2 = 0.6764, q^2 = 0.6506, s = 1.10, F = 100$ (sub−training set)

$n = 21, r^2 = 0.8746, r^2_{pred} = 0.8371, s = 0.717, F = 132$ (calibration set)

$n = 24, r^2 = 0.8943, r^2_{pred} = 0.8792, R^2_m = 0.8215, s = 0.725,$ $F = 186$ (test set)

Split 6

$$\ln R = -0.1911(\pm 0.0227) + 0.3869(\pm 0.0050)*DCW(3) \quad (8)$$

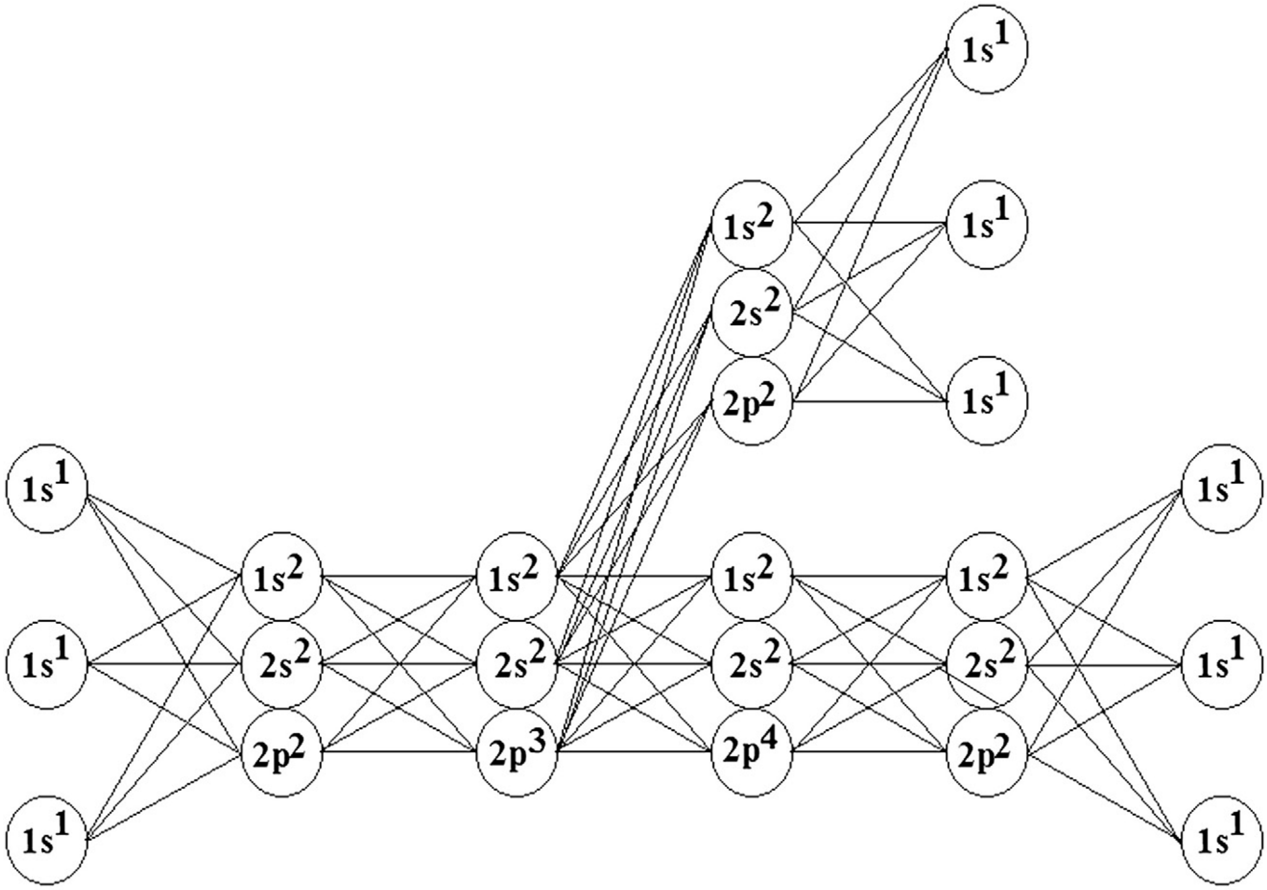
$n = 48, r^2 = 0.7082, q^2 = 0.6844, s = 1.05, F = 112$(sub−training set)

$n = 25, r^2 = 0.8966, r^2_{pred} = 0.8795, s = 0.844, F = 200$ (calibration set)

$n = 22, r^2 = 0.8434, r^2_{pred} = 0.8183, R^2_m = 0.7342, s = 0.808,$ $F = 108$(test set)

In Eqs (3)–(8), the n is the number of compounds in the set; q^2 and R^2_{pred} are leave one out cross validation correlation coefficients; s is the standard error estimation (or root means squared error RMSE); F is Fischer F-ratio.

The range of correlation coefficient values for models of the mutagenic potentials (lnR) from Ref [22] is 0.78–0.834; and the range of standard error is 0.810–0.979. The statistical characteristics of



	1s2	2s2	2p2	1s2	2s2	2p3	1s2	2s2	2p2	1s2	2s2	2p4	1s2	2s2	2p2	1s1	1s1	1s1	1s1	1s1	1s1	1s1	1s1	1s1	δ_k
1s2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0:	6
2s2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0:	6
2p2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0:	6
1s2	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0:	9
2s2	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0:	9
2p3	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0:	9
1s2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0:	6
2s2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0:	6
2p2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0:	6
1s2	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0:	6
2s2	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0:	6
2p4	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0:	6
1s2	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1:	6
2s2	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1:	6
2p2	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1:	6
1s1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0:	3
1s1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0:	3
1s1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0:	3
1s1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0:	3
1s1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0:	3
1s1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0:	3
1s1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0:	3
1s1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0:	3
1s1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0:	3

Fig. 3. Graph of atomic orbitals (GAO).

model calculated with topological state indices suggested by Cash [15] are the following: n = 95, $r^2 = 0.771$, s = 0.979, F = 31 (ten-variable model). Later, Cash et al. have improved the model: n = 95, $r^2 = 0.77$, s = 0.89, F = 48 (six-variable model) [23]. Unfortunately the statistical characteristics of aforementioned models for the external test set are not available in Refs. [15,22,23].

We deem the most important characteristics of models calculated with Eqs. (3)–(8) are statistical characteristics related to the external test set. However, the correlation coefficient between experimental and calculated lnR values and the standard error of estimation for all 95 substances are similar to the above-mentioned values from the literature [15,22,23], the average values for six splits are the following

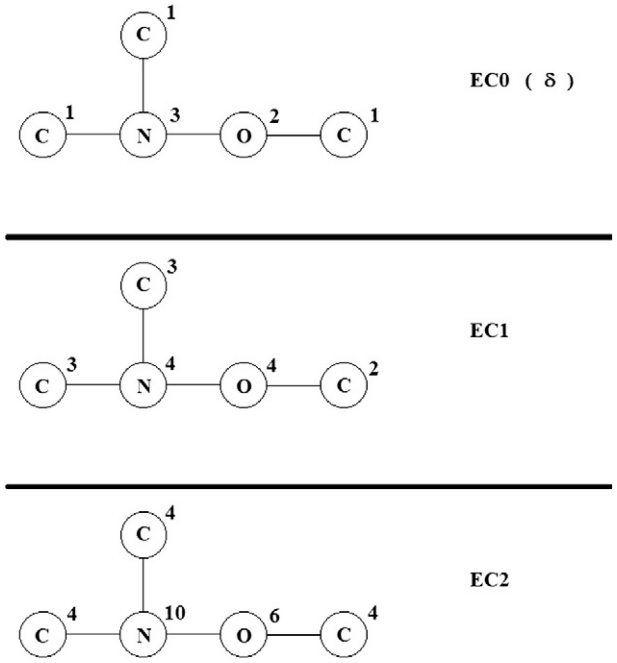


Fig. 4. Example of calculation of the Morgan extended connectivity (EC) of first and second orders. The zero-order EC0 is the vertex degree δ.

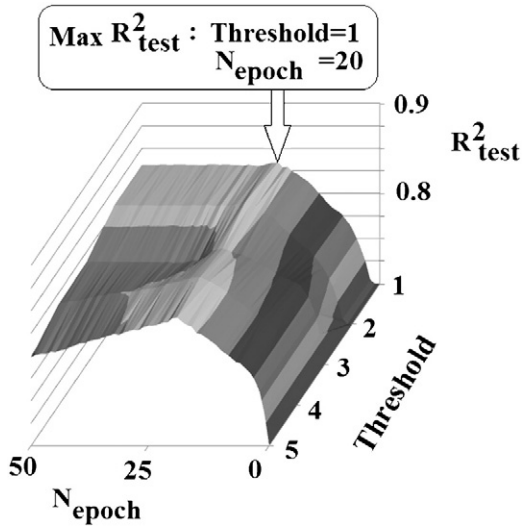


Fig. 6. The preferable model can be established by analysis of the surface $R^2_{(test)} = F(T, N_{epoch})$.

$r^2 = 0.761 \pm 0.013$; $s = 0.942 \pm 0.024$ ($n = 95$). Thus, suggested SMILES-based models have the satisfactory statistical characteristics for six random splits into the sub-training, calibration, and test sets.

Table 3 contains classification of SMILES attributes S_k according to their role during three attempts of the Monte Carlo method optimization: if the correlation weight of S_k $CW(S_k) > 0$ in all three runs of the optimization than the S_k is promoter of lnR increase; if $CW(S_k) < 0$ in all three runs of the optimization than the S_k is promoter of lnR decrease; and if there are both $CW(S_k) < 0$ and $CW(S_k) > 0$ or S_k is blocked then S_k plays an undefined role. One can see (Table 3) that, from probabilistic point of view, '2', '(', and '3' must be classified as the promoters of lnR increase, whereas '1' and 'N' must be classified as the promoters of lnR decrease. Digits in SMILES are indicators of cycles of different kinds. Brackets are indicators of the branching in the molecular skeleton. This information can be useful in searching for substances with low/high mutagenic potential: presence of cycles which are represented by '1' and nitrogen (sp3) is promoter of decrease of the mutagenic potential, vice versa, cycles which are represented by '2' and '3' as well as the high branching of molecular skeleton (i.e. the large number of brackets) is a promoter of increase of the mutagenic potential of a molecule, that is represented by a given SMILES. Unfortunately, the role of other SMILES attributes is less clear, from probabilistic point of view. However, one can see (Table 3) that the role of many SMILES attributes can be changed with change of split into the sub-training, calibration, and test set (owing to change of the distribution of attributes in these sets). This circumstance is important from heuristic point of view.

The advantages of SMILES in comparison with graph is ability to take into account some important molecular features, such as, presence of cycles, cis-/trans- isomerism, the presence of sp^2 and sp^3 atoms, etc. Since there are also some advantages of molecular graphs, the using of hybrid representation of the molecular structure that takes into account both the SMILES attributes and the graph invariants is a possible way to improve the optimal descriptors.

Supplementary materials section contains six random splits which were studied. One can download on the Internet CORAL software [19] and check the suggested models.

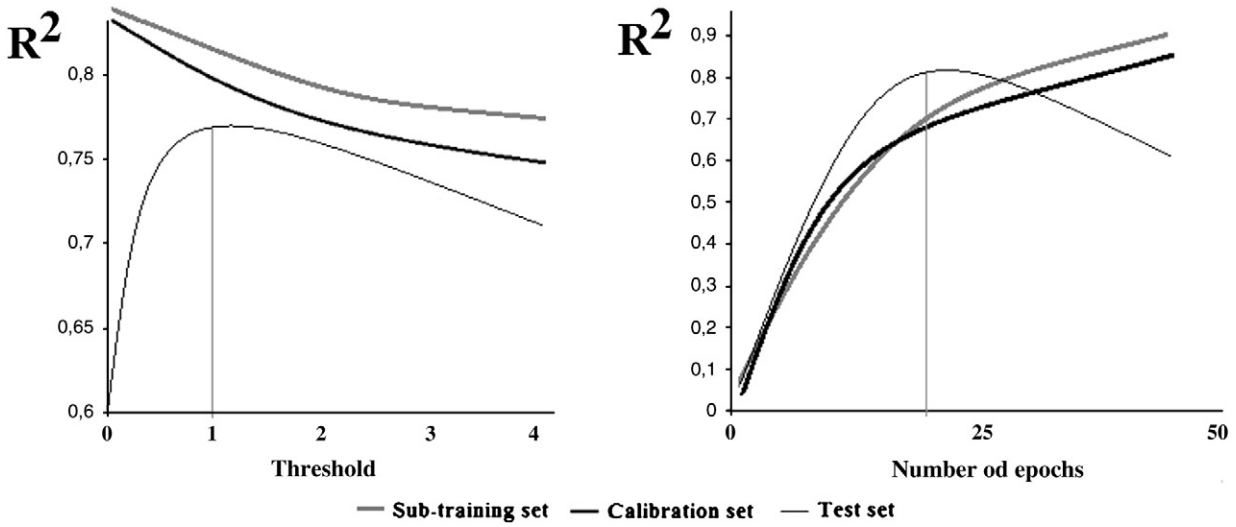


Fig. 5. The correlation coefficient between experimental and calculated lnR values is a mathematical function of threshold and number of epochs of the Monte Carlo method optimization.

Table 1
Statistical characteristics of graph-based model of the mutagenic potentials (lnR).

	EC0				EC1				EC2			
Type of graph	S	T	N	$R^2_{(test)}$	S	T	N	$R^2_{(test)}$	S	T	N	$R^2_{(test)}$
HSG	1	1	20	0.8423	1	2	14	0.8243	1	3	47	0.7685
	2	5	20	0.8046	2	3	28	0.8405	2	2	13	0.8097
	3	2	24	0.8962	3	3	18	0.9351	3	3	11	0.8510
	4	5	46	0.8262	4	4	7	0.8069	4	1	11	0.8557
	5	3	50	0.8708	5	3	20	0.9009	5	2	5	0.6977
	6	4	25	0.8243	6	1	27	0.7453	6	2	13	0.7717
				0.8441 ± 0.031				0.8422 ± 0.062				0.7924 ± 0.054
HFG	1	2	32	0.8450	1	1	34	0.8522	1	1	25	0.7638
	2	2	39	0.7710	2	1	16	0.7932	2	4	30	0.8185
	3	1	50	0.8476	3	1	23	0.8608	3	1	22	0.8565
	4	4	22	0.8839	4	1	33	0.8802	4	2	5	0.6452
	5	3	29	0.8389	5	4	43	0.8322	5	2	50	0.8125
	6	3	31	0.8123	6	3	49	0.8304	6	1	49	0.6004
				0.8331 ± 0.035				**0.8415 ± 0.027**				0.8331 ± 0.035
GAO	1	2	39	0.8421	1	2	22	0.7530	1	5	21	0.7320
	2	1	28	0.7833	2	4	19	0.8074	2	4	35	0.8108
	3	4	50	0.8870	3	4	19	0.9261	3	3	50	0.7620
	4	1	10	0.7818	4	3	16	0.8561	4	1	4	0.6724
	5	5	34	0.8334	5	3	9	0.8214	5	3	17	0.7605
	6	1	49	0.8269	6	1	13	0.8453	6	2	16	0.8543
				0.8257 ± 0.036				**0.8349 ± 0.052**				0.7653 ± 0.057

Bold is indicator of best models.
S = split.
T = threshold.
N = number of epochs of the Monte Carlo optimization.
$R^2_{(test)}$ = correlation coefficient between experimental and calculated lnR for the test set.

Table 2
Statistical characteristics of the best SMILES-based model of the mutagenic potentials (lnR).

α = 1, β = 0, γ = 0				α = 1, β = 1, γ = 0				α = 1, β = 1, γ = 1			
S	T	N	$R^2_{(test)}$	S	T	N	$R^2_{(test)}$	S	T	N	$R^2_{(test)}$
1	2	48	0.8498	1	3	13	0.8387	1	5	50	0.8540
2	5	29	0.8568	2	2	40	0.8225	2	5	23	0.8692
3	3	18	0.8912	3	3	15	0.8918	3	1	12	0.8907
4	3	10	0.8912	4	3	9	0.8664	4	4	8	0.8520
5	3	31	0.9011	5	1	19	0.8660	5	4	14	0.8666
6	3	42	0.8435	6	4	46	0.8742	6	2	18	0.8430
			0.8722 ± 0.023				0.8599 ± 0.023				0.8626 ± 0.015

Bold is indicator of best models.
S = split.
T = threshold.
N = number of epochs of the Monte Carlo optimization.
$R^2_{(test)}$ = correlation coefficient between experimental and calculated lnR for the test set.

4. Conclusions

In the case of the examined 95 aromatic amines, the split into sub-training, calibration, and test sets has considerable influence for the accuracy of prediction. Blocking of the rare molecular features which are detecting with molecular graphs or SMILES can improve the accuracy of the prediction. For each random split examined in this study, there is the most informative number of epochs of the Monte Carlo optimization, which gave most accuracy prediction of the mutagenic potentials. SMILES-based optimal descriptors gave more accurate prediction than optimal descriptors calculated with molecular graphs.

Table 3
Statistical classification of SMILES attributes S_k into three categories: promoters of lnR increase, promoters of lnR decrease, and SMILES attributes which are undefined or blocked. Attributes with apparent role are marked by bold: in other words, 2 and 3 are promoters of lnR increase; 1, N, and C are promoters of lnR decrease; and Br and S are attributes which are not promoters of increase or decrease for lnR.

Split	Promoters of lnR increase	Promoters of lnR decrease	Undefined or blocked
1	**2**, **(**, **3**, +,−,[,4,Cl,Br	**1,N**, c, **C**,=,n,F	O, **S**
2	c, **2**, **(**, **3**, +,−,[	**1,N, C**,O,n,=	Cl, 4,F, **Br**, **S**
3	c, **2**, **(**, **3**, =,4	**1,N**, O, **C**	+,−,[,n, **Br**,Cl,F, **S**
4	c, **2**, **(**, **3**, C, +, −, =, [, n, 4, Cl		1,N,O,F, **Br**, **S**
5	**2**, **(**, **3**, −,4,Cl	**1,N**, c, **C**, =,n	O,+,[F,**S**, **Br**
6	**2**, **(**, **3**, O,+,4,Cl	**1,N**, c, **C**,=,−,n	[,F, **Br, S**

Acknowledgements

The authors express their gratitude to OSIRIS and the NSF CREST Interdisciplinary Nanotoxicity Center NSF-CREST— Grant # HRD-0833178 for financial support. Also, authors express their gratitude to Dr. L. Cappellini and Dr. G. Bianchi for technical assistance and to J. Baggott for the English editing.

References

[1] I. Valkova, M. Vracko, S.C. Basa, Anal. Chim. Acta 509 (2004) 179–186.
[2] P.K. Ojha, K. Roy, Eur. J. Med. Chem. 45 (2010) 4645–4656.
[3] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, G. Kollias, O. Igglessi-Markopoulou, Mol. Divers. 13 (2009) 301–311.
[4] A. Afantitis, G. Melagraki, P.A. Koutentis, H. Sarimveis, G. Kollias, Eur. J. Med. Chem. 46 (2011) 497–508.
[5] A.A. Toropov, A.P. Toropova, I. Gutman, Croat. Chem. Acta 78 (2005) 503–509.
[6] E. Vicente, P.R. Duchowicz, E.A. Castro, A. Monge, J. Mol. Graph. Model. 28 (2009) 28–36.
[7] P.R. Duchowicz, E.A. Castro, Int. J. Mol. Sci. 10 (2009) 2558–2577.

[8] T. Puzyn, A. Mostrag, A. Falandysz, Y. Kholod, J. Leszczynski, J. Hazard. Mater. 170 (2009) 1014–1022.
[9] I. Raska Jr., A. Toropov, Eur. J. Med. Chem. 41 (2006) 1271–1278.
[10] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T.P. Dasari, A. Michalkova, H.-M. Hwang, A. Toropov, D. Leszczynska, J. Leszczynski, Nat. Nanotechnol. 6 (2011) 175–178.
[11] Organisation for Economic Co-operation and Development (OECD) at: http://www.oecd.org/dataoecd/33/37/37849783.pdf.
[12] A.M. Doweyko, J. Comput. Aided Mol. Des. 18 (2004) 587–596.
[13] A.M. Doweyko, J. Comput. Aided Mol. Des. 22 (2008) 81–89.
[14] S.R. Johnson, J. Chem. Inf. Model. 48 (2008) 25–26.
[15] G.G. Cash, Mutat. Res. 491 (2001) 31–37.
[16] A.A. Toropov, A.P. Toropova, J. Mol. Struct. (Theochem) 538 (2001) 287–293.
[17] A.P. Toropova, A.A. Toropov, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Math. Chem. 48 (2010) 959–987.
[18] A.A. Toropov, B.F. Rasulev, J. Leszczynski, Bioorg. Med. Chem. 16 (2008) 5999–6008.
[19] CORALSEA (2011) at http://www.insilico.eu/CORAL.
[20] P.P. Roy, K. Roy, QSAR Comb. Sci. 27 (2008) 302–313.
[21] P.K. Ojha, I. Mitra, R.N. Das, K. Roy, Chemom. Intell. Lab. 107 (2011) 194–205.
[22] S.C. Basak, D. Mills, B.D. Gute, R. Natarajan, Top. Heterocycl. Chem. 3 (2006) 39–80.
[23] G.G. Cash, B. Anderson, K. Mayo, S. Bogaczyk, J. Tunke, Mutat. Res. 585 (2005) 170–183.