Calculation of Molecular Features with Apparent Impact on Both Activity of Mutagens and Activity of Anticancer Agents

Andrey A. Toropov[1,*], Alla P. Toropova[1], Emilio Benfenati[1], Giuseppina Gini[2], Danuta Leszczynska[3] and Jerzy Leszczynski[4]

[1]*Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy;* [2]*Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy;* [3]*Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA;* [4]*Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA*

Abstract: The analysis of influence of molecular features which can be extracted from the simplified molecular input line entry system (SMILES) and involved in the process of the building up of a series of QSAR models (with different splits into training and test sets) by means of the CORAL software for mutagenicity and anticancer activity has been performed. The presence of nitrogen (sp^3) is favorable for decrease of the both endpoints; the presence of only one cycle is also promotor for decrease of the both endpoints; however the presence of two or three cycles is favorable for increase of mutagenicity and decrease of anticancer activity. These findings provide useful criteria for further experimental and computational studies in the search for new anticancer agents.

Keywords: Anticancer activity, Mutagenicity, QSAR, SMILES, Validation, CORAL software.

1. INTRODUCTION

There is a need for efficient computational approaches that provide characteristics of various molecular systems. Among the applied methods the techniques which use initial information obtained from experiments and then link them to structural characteristics are gaining noteworthy recognition. Simplified molecular input line entry system (SMILES) is a representation of the molecular structure [1-4]. This representation can be used for calculation of molecular descriptors for the building up of quantitative structure - property / activity relationships (QSPR/ QSAR) [5-18].

There is a complex correlation between mutagenicity and carcinogenicity [19-32] as well as between mutagenicity and anticancer activity [33,34]. Rigorous research activities are necessary to establish details of such correlations. QSAR methods are capable to accomplish such tasks.

By means of the CORAL software [35] one can calculate so-called correlation weights for different molecular attributes extracted from SMILES. The correlation weights are calculated by the Monte Carlo method. These calculations provide coefficients applied for calculation of the molecular descriptor that is correlated with an endpoint used for the training set. There is a probability that this descriptor is also linked to the endpoint for external test set.

If the process of the Monte Carlo optimization is repeated several times one can obtain three kinds of molecular attributes: 1. attributes with solely positive values of the correlation weights; 2. attributes with solely negative values of the correlation weights; and 3. attributes with both positive and negative values of the correlation weights. In the case 1 one can classify the attribute as a promoter of increase for the endpoint. In the case 2 one can classify the attribute as a promoter of decrease for the endpoint. In the case 3 the role of attribute is undefined.

There are a number of task which can be solved *via* QSPR/QSAR analysis [36-41]. The first task is the building up of QSPR/QSAR models which can be the reliable predictors for various endpoints [42-48]. The CORAL software gives a possibility to compare the correlation weights of molecular attributes for two endpoints related to their prevalence for two sets of compounds (for the first, and the second endpoints, respectively), and according to their correlation weights evaluate them as components of these QSPR/QSAR models.

Using the same method (i.e. applying the same SMILES attributes) one can build up models for anticancer activity and mutagenicity. Establishing a series of such models for different splits (into the training and test sets) one can extract molecular attributes divided into three groups: (1) positive for both anticancer activity and for mutagenicity; (2) negative for both the above-mentioned endpoints; (3) positive for anticancer activity and negative for mutagenicity or vice versa - negative for anticancer activity and positive for mutagenicity. Apparently, this analysis can be useful if (and only if) the models for the both endpoints are characterized by the satisfactory statistical quality. If the prevalence of molecular feature is significant for these two sets one can compare impact of the molecular feature upon the first endpoint and second endpoint [49]. Data on the impact of different molecular features upon the both anticancer activity and mutagenicity can be useful for the search of anti-cancer agents.

The present study was aimed to solve two tasks: (1) To answer the question whether molecular attributes with stable impact for two above-mentioned endpoints do exist? (2) If the answer is yes, to define the list of those molecular attributes.

2. METHOD

Data

The endpoint considered as the anticancer activity of a series of 7- and 3-substituted 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridines, which are novel antitumor quinolone agents is represented by pIC_{50} [i.e. $\log(1/IC_{50})$], where IC_{50} symbolizes the concentration of the agent necessary to reduce cell viability by 50% against Murine P388 Leukemia (*in vitro* cytotoxic activity). Numerical data related to this endpoint were taken from Ref. [50]. Data on mutagenic potentials of the set of 95 aromatic and heteroaromatic amines were taken from Ref. [51]. The mutagenic

*Address correspondence to this author at the Istituto di Ricerche Farmacologiche Mario Negri, 20156, *Via* La Masa 19, Milano, Italy;
Tel: + 39-02-39014595; Fax: + 39-02-39014735;
E-mail: andrey.toropov@marionegri.it

1871-5206/12 $58.00+.00

activity in Salmonella typhimurium TA98+S9 microsomal reparation is expressed as the natural logarithm of R (lnR), where R is the number of revertants per nanomole. SMILES notations for all considered compounds were generated with ACD/ChemSketch software [4]. For both endpoints three splits were examined. These splits are random, but the distribution of compounds for sub-training, calibration, and test sets is done by the manner which gives maximally identical ranges of endpoints in the above-mentioned sets (*Supplementary materials* Table **S1** and **S2**).

Descriptors

The CORAL model represents one-variable model of an endpoint Y, calculated as

$$Y = C_0 + C_1 * DCW(\text{ Threshold, } N_{epoch}) \quad (1)$$

where **DCW(Threshold,N_{epoch})** is the optimal SMILES-based descriptor; C_0 and C_1 are regression coefficients.

The **DCW(Threshold, N_{epoch})** is calculated as

$$DCW(\text{Threshold, } N_{epochs}) = \sum CW(S_k) + \sum CW(SS_k) + CW(BOND) + CW(ATOMPAIR) \quad (2)$$

where S_k, SS_k, ATOMPAIR, and BOND are SMILES attributes (i.e. molecular features) described in the literature [9, 49]. $CW(S_k)$, $CW(SS_k)$, CW(BOND), and CW(ATOMPAIR) are correlation weights of the attributes. The correlation weights are coefficients which are used in Eq.2. They must give maximum of correlation coefficient between experimental and calculated with Eq. 1 values of an endpoint Y for the training set. The threshold defines a coefficient for classification of attributes into two classes: rare and not rare. Correlation weights for rare attributes are fixed equal to zero (blocked). The correlation weights are calculated with the Monte Carlo technique. The number of epochs N_{epoch} of the optimization as well as the threshold have considerable influence on the statistical quality of models [9,49] and their predictability [52]. Fig. (**1**) illustrates the scheme for definition of the preferable threshold and the preferable number of epochs of the Monte Carlo optimization which give a model characterized by the maximal predictive potential.

There are three approaches of the Monte Carlo optimization aimed to build up a QSPR/QSAR model. The first type represents the "classic" scheme [5-15], i.e. searching for maximum of correlation

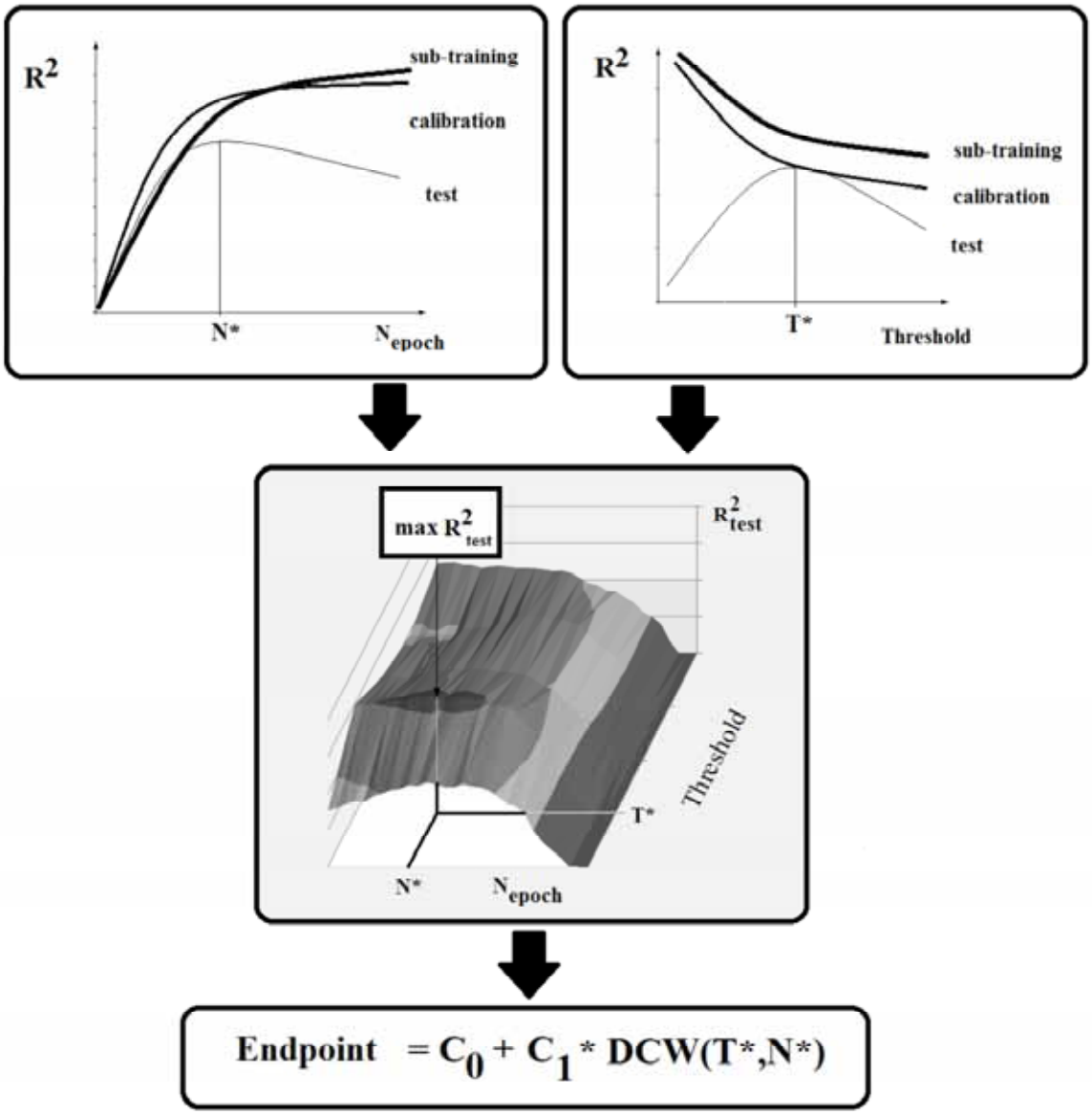


Fig. (1). The general scheme of the building up a model with the CORAL software.

coefficient for the training set hoping that the descriptor will be well correlated for test set. The second type is the distribution of compounds of the training set into sub-training set and calibration set. The role of the calibration set is to form a preliminary test set (the checking up of the identity of the correlation coefficients for the sub-training and calibration sets). This approach is called the balance of correlations [53-60]. The third type is the balance of correlations with ideal slopes, i.e. with checking up the identity of slopes and intercepts for model calculated with the sub-training and calibration sets [16,61]. Models used in this study were built up with CORAL software [35] by means of balance of correlations with ideal slopes [16,61].

3. RESULTS AND DISCUSSION

In order to establish a usefulness of the applied approaches the statistical quality of the model has to be evaluated. The statistical quality of the CORAL models for anticancer activity (pIC_{50}) and mutagencity (lnR) is the following:

Anticancer Activity

Split 1

pIC_{50} = -0.2206(±0.0109) + 0.2248(±0.0023) * DCW(6,69) (3)

n=50, r^2=0.7778, q^2=0.7604, s=0.469, F=168 (sub-training set);

n=25, r^2=0.8684, R^2_{pred}=0.8482, s=0.481, F=152 (calibration set);

n=25, r^2=0.8581, R^2_{pred}=0.8342, s=0.425, F=139, R_m^2=0.7829 (test set)

Split 2

pIC_{50} = -0.0203(±0.0124) + 0.1149(±0.0013) * DCW(5,28) (4)

n=50, r^2=0.7136, q^2=0.6931, s=0.555, F=120 (sub-training set);

n=25, r^2=0.7256, R^2_{pred}=0.6891, s=0.586, F=61 (calibration set);

n=25, r^2=0.7307, R^2_{pred}=0.6842, s=0.517, F=62, R_m^2=0.7137 (test set)

Split 3

pIC_{50} = -0.1734(±0.0095) + 0.1914(±0.0012) * DCW(3,71) (5)

n=50, r^2=0.7774, q^2=0.7626, s=0.445, F=168 (sub-training set);

n=25, r^2=0.9103, R^2_{pred}=0.9003, s=0.355, F=233 (calibration set);

n=25, r^2=0.7054, R^2_{pred}=0.6559, s=0.714, F=55, R_m^2=0.6993 (test set)

Mutagenicity

Split 1

lnR = -4.8389 (± 0.058) + 0.1142 (± 0.0013) * DCW(3,11) (6)

n=42, r^2=0.7506, q^2=0.7297, s=1.10, F=120 (sub-training set);

n=25, r^2=0.7828, R^2_{pred}=0.7293, s=0.811, F=83 (calibration set);

n=28, r^2=0.8361, R^2_{pred}=0.8048, s=0.782, F=133, R_m^2=0.7076 (test set)

Split 2

lnR = -2.5951(±0.0395) + 0.1506(±0.0022) * DCW(5,25) (7)

n=42, r^2=0.7441, q^2=0.7177, s=0.945, F=116 (sub-training set);

n=25, r^2=0.7936, R^2_{pred}=0.7642, s=0.884, F=88 (calibration set);

n=28, r^2=0.8052, R^2_{pred}=0.7621, s=0.925, F=107, R_m^2=0.7359 (test set)

Split 3

lnR = -0.0928(±0.0217) + 0.2604(±0.0033) * DCW(3,58) (8)

n=43, r^2=0.7791, q^2=0.7578, s=0.890, F=145 (sub-training set);

n=25, r^2=0.8970, R^2_{pred}=0.8812, s=0.599, F=200 (calibration set);

n=27, r^2=0.8870, R^2_{pred}=0.8692, s=0.704, F=196, R_m^2=0.8194 (test set)

In Eqs. 3-8, n is the number of compounds in a set; r is correlation coefficient; q^2 is leave-one-out cross-validated correlation coefficient; R^2_{pred} is external predictive correlation coefficient; s is standard error of estimation (root mean square error); R_m^2 is novel validation metric [52] calculated according to Eq. 9

$$R_m^2 = r^2 \times (1 - \sqrt{r^2 - r_0^2}) \quad (9)$$

where r_0^2 is correlation coefficient between observed and predicted values without intercept [52]. Fig. (**2**) contains graphical representations of models calculated with Eq. 3 and Eq. 6.

The number of attributes which are involved in the building up of CORAL model depends upon the assumed threshold. The typical situation is the following. The increase of threshold is accompanied by decrease of correlation coefficient for the sub-training and test sets, but there is maximum of the correlation coefficient for test set. This maximum occurs for a specific threshold value which is denoted as T* (Fig. **1**). The increase of the number of epochs of the Monte Carlo optimization is accompanied by increase of the correlation coefficients between experimental and calculated values of an endpoint for sub-training and calibration set. For the test set there are two phases. Phase 1: the increase of correlation coefficient till a maximum is reached (the number of epochs is equal to N*); and phase 2: decrease of the correlation coefficient (Fig. **1**). The T* and N* are represented for Eqs. 3-8, e.g. in the case of Eq. 3 T*=6 and N*=69.

The balance of correlations [53-60] with ideal slopes [16,61] has been used to build up the models for anticancer activity (Eqs. 3-5) and mutagenicity (Eqs. 6-8). The statistical quality of models for anticancer activity calculated with Eqs 3-5 is approximately identical for three splits. The same results are obtained for models of mutagenicity calculated with Eqs. 6-8. Consequently, the comparison of molecular features which are involved in these models and which have considerable prevalence provides an interesting and useful way of the investigation of the aforementioned endpoints.

Table **1** shows the results of the analysis of influence of molecular attributes which are extracted from SMILES on the anticancer activity and the mutagenicity. The selection of the attributes has been done by the following scheme. Firstly, only apparent promoters of increase or decrease of endpoints were involved in the analysis i.e. attributes which have only positive or only negative values of the correlation weights in three runs of the Monte Carlo optimization (*Supplementary Materials* Table **S3** and S4). Secondly, only attributes with considerable prevalence were extracted from the above-mentioned apparent promoters of increase or decrease of endpoints. The impact of the apparent promoters with considerable prevalence has been studied for nine combinations of three models for anti-cancer activity and three models for mutagenicity (*Supplementary materials* Table **S5**).

There are three SMILES attributes which have clear function for the nine considered combinations of the models. These are ‘c(‘, ‘1’, and ‘N’. The interpretation for ‘c(‘ can be formulated as presence of branching which starts from carbon (sp^2) in an aromatic system. The attribute ‘1’ means presence of a cycle. The attribute ‘N’ means presence of nitrogen (sp^3). The impact of attributes ‘c(‘ and ‘1’ is increase for the both endpoints, whereas presence of ‘N’ should lead to decrease of both endpoints.

There is attribute ‘c2’ which occurs in eight of nine examined combinations of three anticancer models and three mutagenicity models. The attribute can be interpreted as presence of cycle which contains aromatic carbon (sp^2). The presence of this molecular feature should lead to decrease of the both endpoints.

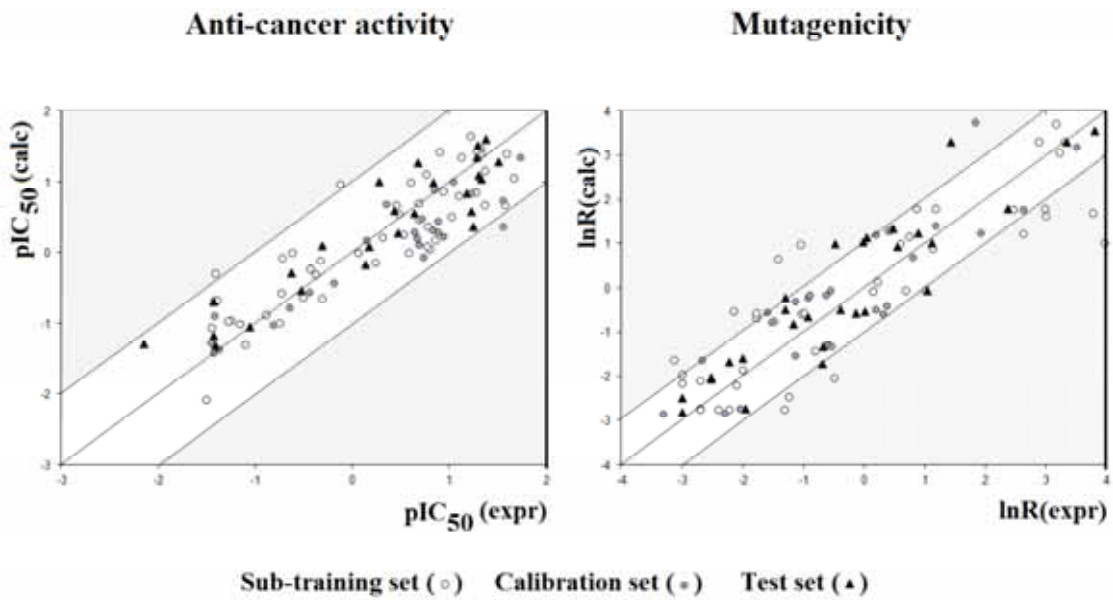


Fig. (2). Graphical representation of models which are calculated with the CORAL software: (i) Eq. 3 for anti-cancer activity pIC_{50}, the concentration of the agent necessary to reduce cell viability by 50% against Murine P388 Leukemia; and (ii) Eq. 6 for mutagenic activity in *Salmonella typhimurium* TA98+S9 lnR, where R is the number of revertants per nanomole.

Table 1. The analysis of Influences of Various Molecular Features on the Anti-cancer Activity and the Mutagenicity

		Mutagenicity		
		Split1	**Split2**	**Split3**
Anti-cancer activity	**Split1**	c(↑↑, 1↑↑, N↓↓, c2↑↑, (↓↑, 2↓↑	c(↑↑, 1↑↑, N↓↓, c2↑↑	c(↑↑, 1↑↑, N↓↓, c2↑↑, (↓↑, 2↓↑
	Split2	c(↑↑, 1↑↑, N↓↓, c2↑↑, (↓↑, 2↓↑	c(↑↑, 1↑↑, N↓↓, c2↑↑	c(↑↑, 1↑↑, N↓↓, (↓↑, 2↓↑
	Split3	c(↑↑, 1↑↑, N↓↓, c2↑↑, (↓↑, 2↓↑	c(↑↑, 1↑↑, N↓↓, c2↑↑	c(↑↑, 1↑↑, N↓↓, c2↑↑, (↓↑, 2↓↑

↑ is an indicator of increase; ↓ is an indicator of decrease; each molecular feature is accompanied by two indicators, the first is related to anti-cancer activity, the second is related to mutagenicity.

Finally, there are two SMILES attributes which occur in six of nine examined combinations of models for two above-mentioned endpoints. These are '(' and '2'. The attribute '(' means presence of any branching. It is to be noted 'c(' and '(' are not the same. The attribute '2' means presence of any two cycles. It is also to be noted that '2' and 'c2' are not the same. We deem that 6/9 occurrences hardly can be classified as absolutely random result. Consequently, presence of these two molecular features can be interpreted as quite probable decrease of anticancer activity together with quite probable increase of mutagenicity. The lack of influence of these two attributes for the two endpoints is observed for the three combinations which involve models of mutagenicity obtained for split 2. Therefore, possibly this split is not 'typical' in respect of distribution of these SMILES attributes in the sub-training, the calibration, and the test sets. *Supplementary materials* section contains the technical details of the described analysis.

Table **2** shows possible ways to construct anticancer agents with using model (the split 1, Eq. 3) based on the molecular features with stable positive or negative influence on the pIC50. We have attempted to carry out modifications of five arbitrary molecular structures according to data from Table **1**. In fact Table **2** contains a group of hypotheses, which need confirmation by the experiment, however good quality of the model calculated with Eq. 3 is argument to estimate these predictions as quite reliable. Modifications for #1, #24, #59, and #74 illustrate the influence of presence of fragment "c(". Modification for #89 illustrates influence of presence of fragment "N".

We believe that the results of this study are useful for investigations the links between mutagenicity and carcinogenicity since the list of attributes with clear influence on the endpoints is not empty and the influence is statistically significant. This approach is general and can provide useful tools for analysis of other types of endpoints. It is very probable that using similar or even identical substances in described analysis can be more beneficial. However, the comparison of very different molecular structures of mutagens and anti-cancer agents is attractive from heuristic point of view.

It is to be noted the number of SMILES attributes can be increased [35]. In this case the statistical quality of a model for training set (or sub-training and calibration sets) will be improved, but it is

Table 2. Analysis of Influence of Various Modifications of Structures Upon the pIC_{50} Values (Anticancer Activity)

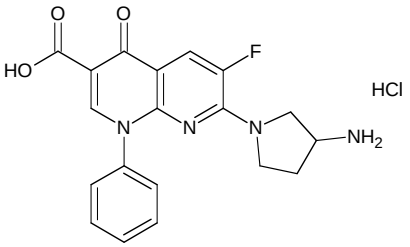
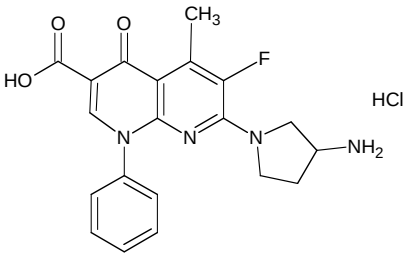
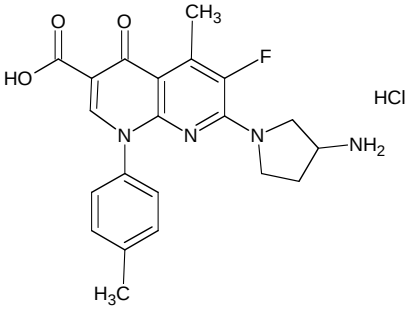
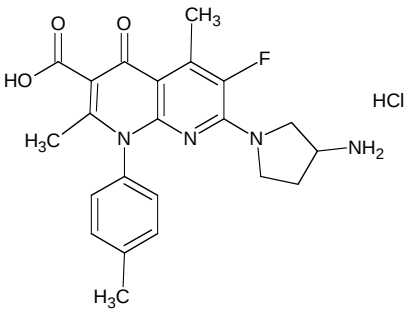
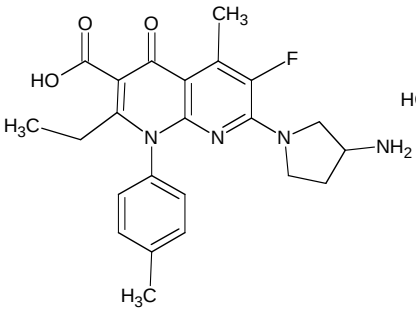
ID	Structure and SMILES	pIC_{50} Experiment	pIC_{50} Calculated with Eq. 3
1	 Cl.O=C(O)C2=CN(c1nc(c(F)cc1C2=O)N3CCC(N)C3)c4ccccc4	-0,8139	-1.013
	 Cl.O=C(O)C2=CN(c1nc(c(F)c(C)c1C2=O)N3CCC(N)C3)c4ccccc4		-0,414
	 Cl.O=C(O)C2=CN(c1nc(c(F)c(C)c1C2=O)N3CCC(N)C3)c4ccc(C)cc4		0,185
	 Cl.NC1CCN(C1)c4nc2c(C(=O)C(=C(C)N2c3ccc(C)cc3)C(=O)O)c(C)c4F		1,203
	 Cl.NC1CCN(C1)c4nc2c(C(=O)C(=C(CC)N2c3ccc(C)cc3)C(=O)O)c(C)c4F		1,502

Table 2. contd….

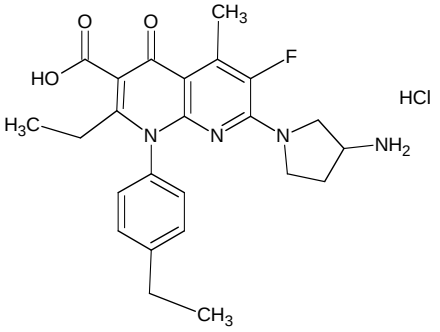
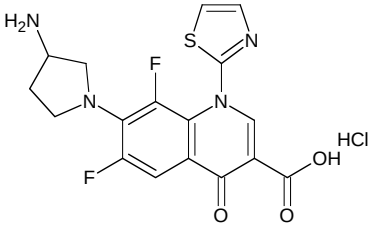
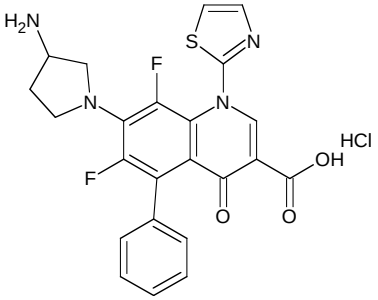
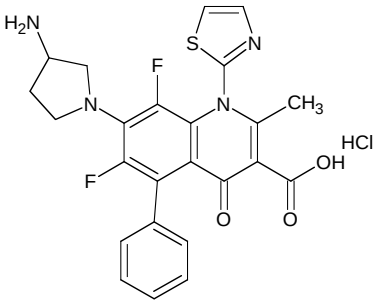
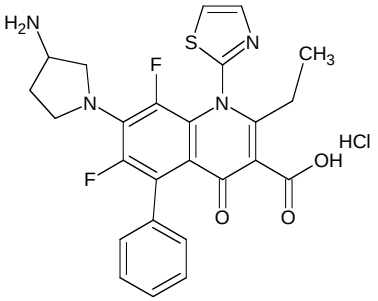
ID	Structure and SMILES	pIC_{50} Experiment	pIC_{50} Calculated with Eq. 3
	 Cl.NC1CCN(C1)c4nc2c(C(=O)C(=C(CC)N2c3ccc(CC)cc3)C(=O)O)c(C)c4F		1,801
24	 Cl.NC1CCN(C1)c3c(F)cc4C(=O)C(=CN(c2nccs2)c4c3F)C(=O)O	-2,1467	-1,281
	 Cl.NC1CCN(C1)c4c(F)c(c2ccccc2)c5C(=O)C(=CN(c3nccs3)c5c4F)C(=O)O		0,674
	 Cl.NC1CCN(C1)c4c(F)c(c2ccccc2)c5C(=O)C(=C(C)N(c3nccs3)c5c4F)C(=O)O		0,845
	 Cl.NC1CCN(C1)c4c(F)c(c2ccccc2)c5C(=O)C(=C(CC)N(c3nccs3)c5c4F)C(=O)O		1,143

Table 2. contd….

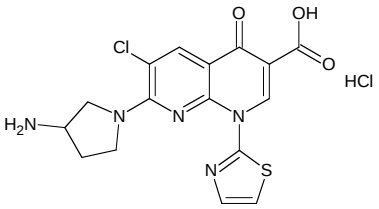
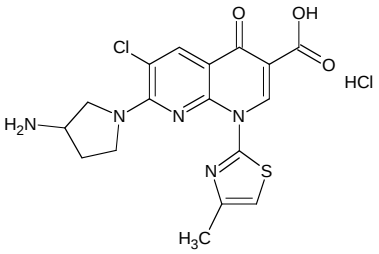
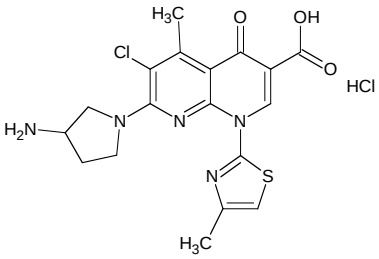
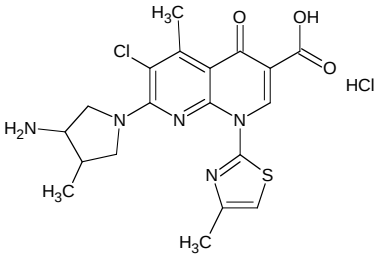
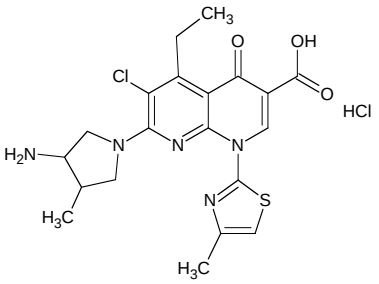
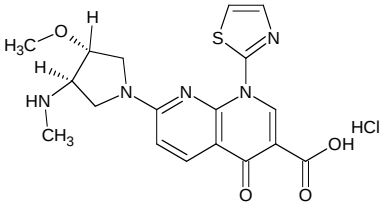
ID	Structure and SMILES	pIC_{50} Experiment	pIC_{50} Calculated with Eq. 3
59	 Cl.NC1CCN(C1)c2nc3N(C=C(C(=O)c3cc2Cl)C(=O)O)c4nccs4	0,2371	-0,139
	 Cl.NC1CCN(C1)c2nc3N(C=C(C(=O)c3cc2Cl)C(=O)O)c4nc(C)cs4		0,460
	 Cl.NC1CCN(C1)c2nc3N(C=C(C(=O)c3c(C)c2Cl)C(=O)O)c4nc(C)cs4		1,059
	 Cl.CC1CN(CC1N)c2nc3N(C=C(C(=O)c3c(C)c2Cl)C(=O)O)c4nc(C)cs4		1,459
	 Cl.CC1CN(CC1N)c2nc3N(C=C(C(=O)c3c(CC)c2Cl)C(=O)O)c4nc(C)cs4		1,758
74	 Cl.CN[C@@H]1CN(C[C@H]1OC)c3ccc4C(=O)C(=CN(c2nccs2)c4n3)C(=O)O	1,7282	1,355

Table 2. contd….

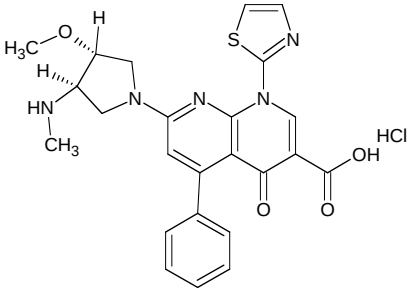
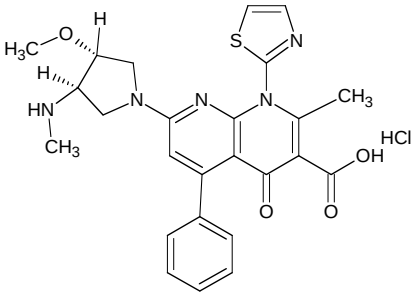
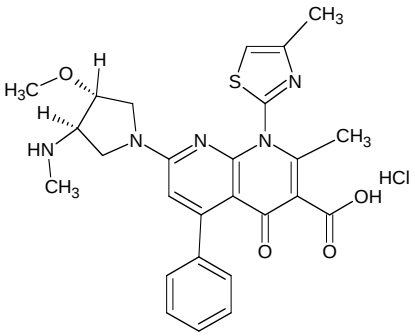
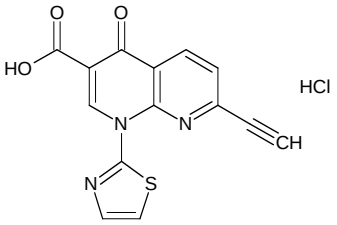
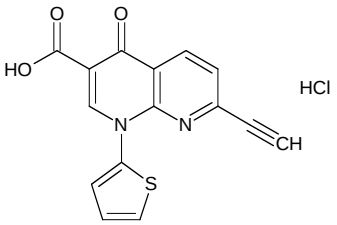
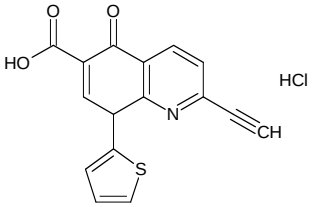
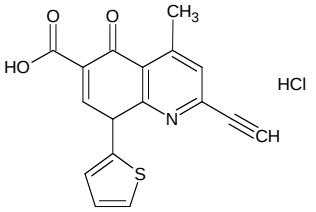
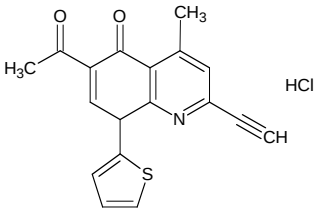
ID	Structure and SMILES	pIC_{50} Experiment	pIC_{50} Calculated with Eq. 3
	 Cl.CN[C@@H]1CN(C[C@H]1OC)c4cc(c2ccccc2)c5C(=O)C(=CN(c3nccs3)c5n4)C(=O)O		1,611
	 Cl.CN[C@@H]1CN(C[C@H]1OC)c4cc(c2ccccc2)c5C(=O)C(=C(C)N(c3nccs3)c5n4)C(=O)O		1,781
	 Cl.CN[C@@H]1CN(C[C@H]1OC)c4cc(c2ccccc2)c5C(=O)C(=C(C)N(c3nc(C)cs3)c5n4)C(=O)O		2,381
89	 Cl.O=C(O)C2=CN(c1nccs1)c3nc(C#C)ccc3C2=O	-0,8886	-0,874
	 Cl.O=C(O)C2=CN(c1cccs1)c3nc(C#C)ccc3C2=O		-0,089

Table 2. contd….

ID	Structure and SMILES	pIC_{50} Experiment	pIC_{50} Calculated with Eq. 3
	 Cl.O=C(O)C2=CC(c1cccs1)c3nc(C#C)ccc3C2=O		0,228
	 Cl.O=C(O)C2=CC(c1cccs1)c3nc(C#C)cc(C)c3C2=O		0,828
	 Cl.CC(=O)C2=CC(c1cccs1)c3nc(C#C)cc(C)c3C2=O		1,563

not clear whether it will be accompanied by the improving of the statistical quality of this model for the external test set.

4. CONCLUSIONS

This study revealed interesting and useful information related to link between anti-cancer activity and mutagenicity of series of compounds. In spite of the considerable differences in molecular architecture of substances used for QSAR modeling of these two properties, there are molecular features (which can be extracted from SMILES) with considerable prevalence in examined data sets with apparent influence on the endpoints.

The presence of branching in an aromatic system (it is encoded in SMILES by 'c(') and presence of a cycle (it is encoded in SMILES by '1') represent promoters of increase for both anti-cancer activity and mutagenicity. The presence of nitrogen (sp^3) is an indicator of decrease for the both endpoints. With high probability (it occurs for eight of nine comparisons of models) one can conclude that the presence of two cycles together with aromatic system (this is indicated in SMILES by 'c2') is an indicator of increase of both endpoints.

Finally, it is quite probably (this occurs in six of nine comparisons), that the presence of a branching (it is encoded in SMILES by '(') as well as presence of two cycles (it is encoded in SMILES by '2') are promoters of decrease of anti-cancer activity and promoters of increase of mutagenicity.

CONFLICT OF INTEREST

Declared none.

ACKNOWLEDGEMENT

We thank ANTARES (the project number LIFE08-ENV/IT/00435), and the National Science Foundation (NSF/CREST HRD-0833178, and EPSCoR Award #:362492-190200-01\NSFEPS-090378) for financial support. Also, the authors express their gratitude to J. Baggott for English edition.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Weininger, D.; SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31-36.

[2] Weininger, D.; Weininger, A.; Weininger, J.L. SMILES. 2. Algorithm for generation of unique SMILES notation *J. Chem. Inf. Compu. Sci.* **1989**, *29*, 97-101.

[3] Weininger, D. Smiles. 3. Depict. Graphical depiction of chemical structures *J. Chem. Inf. Comput. Sci.* **1990**, *30* (3), 237-243.

[4] Advanced Chemistry Development, Inc., Toronto, Canada, ACD/Chemsketch software, A complete software package for drawing chemical structures, http://www.acdlabs.com/products/draw_nom/draw/chemsketch/ (accessed February 7, **2012**)

[5] García, J.; Duchowicz, P.R.; Rozas, M.F.; Caram, J.A., Mirífico, M.V.; Fernández, F.M., Castro, E.A. A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. *J. Mol. Graph. Model.* **2011**, *31*, 10-19.

[6] Mullen, L.M.A.; Duchowicz, P.R.; Castro, E.A. QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents. *Chemometr. Intell. Lab.* **2011**, *107* (2), 269-275.

[7] Mercader, A.G.; Duchowicz, P.R.; Fernández, F.M.; Castro, E.A. Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories. *J. Chem. Inf. Model.* **2010**, *50* (9), 1542-1548.

[8] Toropov, A.A.; Toropova, A.P.; Martyanov, S.E.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. Comparison of SMILES and molecular graphs as the representation of the molecular structure for QSAR analysis for mutagenic potential of polyaromatic amines. *Chemometr. Intell. Lab.* **2011**, *109* (1), 94-100.

[9] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. CORAL: Quantitative structure-activity relationship models for estimating toxicity of organic compounds in rats. *J. Comput. Chem.* **2011**, *32* (12), 2727-2733.

[10] Benfenati, E.; Toropov, A.A.; Toropova, A.P.; Manganaro, A.; Gonella Diaza, R. CORAL software: QSAR for anticancer agents. *Chem. Biol. Drug Des.* **2011**, *77* (6), 471-476.

[11] Toropov, A.A.; Toropova, A.P.; Lombardo, A.; Roncaglioni, A.; Benfenati, E.; Gini, G. CORAL: Building up the model for bioconcentration factor and defining it's applicability domain. *Eur. J. Med. Chem.* **2011**, *46* (4), 1400-1403.

[12] Toropova, A.P.; Toropov, A.A.; Diaza, R.G.; Benfenati, E.; Gini, G. Analysis of the co-evolutions of correlations as a tool for QSAR-modeling of carcinogenicity: An unexpected good prediction based on a model that seems untrustworthy. *Cent. Eur. J. Chem.* **2011**, *9* (1), 165-174.

[13] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Gini, G. Co-evolutions of correlations for QSAR of toxicity of organometallic and inorganic substances: An unexpected good prediction based on a model that seems untrustworthy *Chemometr. Intell. Lab.* **2011**, *105* (2), 215-219.

[14] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. CORAL: QSPR models for solubility of [C60] and [C70] fullerene derivatives. *Mol. Divers.* **2011**, *15* (1), 249-256.

[15] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Gini, G. QSAR modelling toxicity toward rats of inorganic substances by means of CORAL *Cent. Eur. J. Chem.* **2011**, *9* (1), 75-85.

[16] Toropov, A.A.; Toropova, A.P.; Benfenati, E. SMILES-based optimal descriptors: QSAR modeling of carcinogenicity by balance of correlations with ideal slopes. *Eur. J. Med. Chem.* **2011**, *45* (9), 3581-3587.

[17] Toropova, A.P.; Toropov, A.A.; Lombardo, A.; Roncaglioni, A.; Benfenati, E.; Gini, G. A new bioconcentration factor model based on SMILES and indices of presence of atoms. *Eur. J. Med. Chem.* **2010**, *45* (9), 4399-4402.

[18] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. QSAR modeling of measured binding affinity for fullerene-based HIV-1 PR inhibitors by CORAL. *J. Math. Chem.* **2010**, *48* (4), 959-987.

[19] Benigni, R.; Bossa, C. Alternative strategies for carcinogenicity assessment: An efficient and simplified approach based on *in vitro* mutagenicity and cell transformation assays. *Mutagenesis*, **2011**, *26* (3), 455-460.

[20] Benigni, R.; Bossa, C. Mechanisms of chemical carcinogenicity and mutagenicity: A review with implications for predictive toxicology. *Chem. Rev.* **2011**, *111* (4), 2507-2536.

[21] Benigni, R.; Bossa, C.; Tcheremenskaia, O.; Giuliani, A. Alternatives to the carcinogenicity bioassay: In silico methods, and the *in vitro* and *in vivo* mutagenicity assays. *Expert Opin. Drug Metabol. Toxicol.* **2010**, *6* (7), 809-819.

[22] Combes, R.; Grindon, C.; Cronin, M.T.D.; Roberts, D.W.; Garrod, J.F. Integrated decision-tree testing strategies for mutagenicity and carcinogenicity with respect to the requirements of the EU REACH legislation. *ATLA Altern. Lab. Anim*. **2008**, *36* (SUPPL. 1), 43-63.

[23] Combes, R.; Grindon, C.; Cronin, M.T.D.; Roberts, D.W.; Garrod, J. Proposed integrated decision-tree testing strategies for mutagenicity and carcinogenicity in relation to the EU REACH legislation. *ATLA Altern. Lab. Anim*. **2007**, *35* (2), 267-287.

[24] Helma, C. Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. *Mol. Divers.* **2006**, *10* (2), 147-158.

[25] Benigni, R. Computational prediction of drug toxicity: The case of mutagenicity and carcinogenicity *Drug Discov. Today: Technol.* **2004**, *1* (4),457-463.

[26] Benigni, R.; Passerini, L.; Rodomonte, A. Structure-Activity Relationships for the Mutagenicity and Carcinogenicity of Simple and α-β Unsaturated Aldehydes. *Environ. Mol. Mutagen.* **2003,** *42* (3), 136-143.

[27] Patlewicz, G.; Rodford, R.; Walker, J.D. Quantitative structure-activity relationships for predicting mutagenicity and carcinogenicity. *Environ. Toxicol. Chem.* **2003**, *22* (8), 1885-1893.

[28] Novak, M.; Rajagopal, S. Correlations of nitrenium ion selectivities with quantitative mutagenicity and carcinogenicity of the corresponding amines. *Chem. Res. Toxicol.* **2002**, *15*(12), 1495-1503.

[29] Rosenkranz, H.S.; Karol, M.H. Chemical carcinogenicity: Can it be predicted from knowledge of mutagenicity and allergic contact dermatitis? *Mut. Res. – Fund. Mol. M.* **1999**, *431* (1), 81-91.

[30] Lewis, D.F.V.; Ioannides, C.; Parke, D.V. A combined COMPACT and HazardExpert study of 40 chemicals for which information on mutagenicity and carcinogenicity is known, including the results of human epidemiological studies *Human Exper. Toxicol.* **1998**, *17* (10), 577-586.

[31] Richard, A.M. Structure-based methods for predicting mutagenicity and carcinogenicity: Are we there yet? *Mut. Res. – Fund. Mol. M.* **1998**, *400* (1-2), 493-507.

[32] Klopman, G.; Rosenkranz, H.S. Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/ mutagenicity using MULTI-CASE. *Mut. Res. – Fund. Mol. M.* **1994**, *305* (1), 33-46.

[33] Venitt, S.; Crofton-Sleigh, C.; Agbandje, M., Jenkins, T.C.; Neidle, S. Anthracene-9,10-diones as potential anticancer agents: Bacterial mutation studies of amido-substituted derivatives reveal an unexpected lack of mutagenicity. *J. Med. Chem.* **1998**, *41*(19), 3748-3752.

[34] Juneja, T.R.; Bala, A.; Kumar, P.; Gupta, R.L. Mutagenicity of nitrobenzyl derivatives: Potential bioreductive anticancer agents. *Mutat Res Lett.* **1995**, *348* (3), 137-145.

[35] The CORAL freeware (Correlation and Logic = CORAL), http://www.insilico.eu/CORAL (accessed January 9, **2012**)

[36] Concu, R.; Podda, G.; Ubeira, F.M.; González-Díaz, H. Review of QSAR models for Enzyme classes of drug targets: Theoretical background and applications in parasites, hosts and other organisms. *Curr. Pharm. Des.* **2010**, *16* (24), 2710-2723.

[37] Verma, J.; Khedkar, V.M.; Coutinho, E.C. 3D-QSAR in drug design - A review. *Curr. Top. Med. Chem.* **2010**, *10* (1), 95-115.

[38] Dudek, A.Z.; Arodz, T.; Gálvez, J. Computational methods in developing quantitative structure-activity relationships (QSAR): A review. *Comb. Chem. High T. Scr.* **2006**, *9* (3), 213-228.

[39] Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR applicability domain estimation by projection of the training set in descriptor space: A review. *ATLA Altern. Lab. Anim*. **2005**, *33* (5), 445-459.

[40] Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A.R. QSPR and QSAR models derived using large molecular descriptor spaces. A review of CODESSA applications. Coll. Czechosl. Chem. Commun. **1999**, *64* (1), 1551-1571.

[41] Tuppurainen, K. Frontier orbital energies, hydrophobicity and steric factors as physical QSAR descriptors of molecular mutagenicity. A review with a case study: MX compounds. *Chemosphere*, **1999**, *38* (13), 3015-3030.

[42] Roy, K.; Mitra, I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb. Chem. High T. Scr.* **2011**, *14* (6), 450-474.

[43] Ojha, P.K.; Roy, K. Exploring QSAR, pharmacophore mapping and docking studies and virtual library generation for cycloguanil derivatives as PfDHFR-TS inhibitors. *Med. Chem.* **2011**, *7* (3), pp. 173-199.

[44] Hemmateenejad, B.; Yousefinejad, S.; Mehdipour, A.R. Novel amino acids indices based on quantum topological molecular similarity and their application to QSAR study of peptides. *Amino Acids*, **2011**, *40* (4), pp. 1169-1183.

[45] Hemmateenejad, B.; Mehdipour, A.R.; Miri, R., Shamsipur, M. Comparative qsar studies on toxicity of phenol derivatives using quantum topological molecular similarity indices. *Chem. Biol. Drug Des.* **2010**, *75* (5), 521-531.

[46] Melagraki, G., Afantitis, A. Ligand and structure based virtual screening strategies for hit-finding and optimization of Hepatitis C virus (HCV) inhibitors. *Curr. Med. Chem.* **2011**, *18*(17), 2612-2619.

[47] Afantitis, A.; Melagraki, G.; Koutentis, P.A.; Sarimveis, H.; Kollias, G. Ligand - Based virtual screening procedure for the prediction and the identification of novel β-amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. *Eur. J. Med. Chem.* **2011**, *46* (2), 497-508.

[48] Melagraki, G.; Afantitis, A.; Sarimveis, H.; Igglessi-Markopoulou, O.; Koutentis, P.A., Kollias, G. In silico exploration for identifying structure-activity relationship of MEK inhibition and oral bioavailability for isothiazole derivatives *Chem. Biol. Drug Des.* **2010**, *76*(5), 397-406.

[49] Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. SMILES-based QSAR approaches for carcinogenicity and anticancer activity: Comparison of correlation weights for identical SMILES attributes. *Anti-cancer. Agents. Med. Chem.* **2011**, *11*(10), 974-982.

[50] Atanasova, M.; Ilieva, S.; Galabov, B. QSAR analysis of 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridines with anticancer activity *Eur. J. Med. Chem.* **2007**, *42* (9), 1184-1192.

[51] Cash, G.G. Prediction of the genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices *Mutat. Res.* **2001**, *491* (1-2), 31-37.

[52] Ojha, P.K.; Mitra, I.; Das, R.N.; Kunal Roy, K. Further exploring rm 2 metrics for validation of QSPR models *Chemometr. Intell. Lab.* **2011**, *107*(1), 194–205.

[53] Toropov, A.A.; Toropova, A.P.; Benfenati, E. QSAR modelling of the toxicity to Tetrahymena pyriformis by balance of correlations *Mol. Divers.* **2010**, *14*(4), 821-827.

[54] Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. SMILES-Based Optimal Descriptors: QSAR Analysis of Fullerene-Based HIV-1 PR Inhibitors by Means of Balance of Correlations. *J. Comput. Chem.* **2010**, *31*(2), 381-392.

[55] Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. InChI-based optimal descriptors: QSAR analysis of fullerene[C60]-based HIV-1 PR inhibitors by correlation balance. *Eur. J. Med. Chem.* **2010**, *45*(4), 1387–1394.

[56] Toropov, A. A.; Toropova, A.P.; Raska, I.; Benfenati, E. QSPR modeling of octanol/water partition coefficient of antineoplastic agents by balance of correlations. *Eur. J. Med. Chem.* **2010**, *45* (5), 1639-1647.

[57] Toropov, A.A.; Toropova, A.P.; Benfenati, E. QSAR-modeling of toxicity of organometallic compounds by means of the balance of correlations for InChI-based optimal descriptors. *Mol. Divers.* **2010**, *14* (1), 183-192.

[58] Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Manganaro, A. QSAR modelling of carcinogenicity by balance of correlations. *Mol. Divers.* **2009**, *13*(3), 367-373.

[59] Toropov, A.A.; Toropova, A.P.; Benfenati, E. QSPR modeling bioconcentration factor (BCF) by balance of correlations. *Eur. J. Med. Chem.* **2009**, *44* (6), 2544-2551.

[60] Toropov, A.A.; Rasulev, B.F.; Leszczynski, J. QSAR modeling of acute toxicity by balance of correlations. *Bioorg. Med. Chem.* **2008**, *16*(11), 5999-6008.

[61] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. QSAR modeling of anxiolytic activity taking into account the presence of keto- and enol-tautomers by balance of correlations with ideal slopes. *Cent. Eur. J. Chem.* **2011**, *9*(5), 846-854.

Received: November 21, 2011 Revised: February 14, 2012 Accepted: February 20, 2012