

## Dataset Paper

# Rodent Carcinogenicity Dataset

**Natalja Fjodorova and Marjana Novič**

*Laboratory of Chemometrics, National Institute of Chemistry, Hajdrihova 19, 1001 Ljubljana, Slovenia*

Correspondence should be addressed to Natalja Fjodorova; [natalja.fjodorova@ki.si](mailto:natalja.fjodorova@ki.si)

Received 1 June 2012; Accepted 27 June 2012

Academic Editors: E. Frei and K. van Golen

Copyright © 2013 N. Fjodorova and M. Novič. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rodent carcinogenicity dataset was compiled from the Carcinogenic Potency Database (CPDBAS) and was applied for the classification of quantitative structure-activity relationship (QSAR) models for the prediction of carcinogenicity based on the counter-propagation artificial neural network (CP ANN) algorithm. The models were developed within EU-funded project CAESAR for regulatory use. The dataset contains the following information: *common* information about chemicals (ID, chemical name, and their CASRN), *molecular structure* information (SDF files and SMILES), and *carcinogenic (toxicological) properties* information: carcinogenic potency (TD50\_Rat\_mg; carcinogen/noncarcinogen) and structural alert (SA) for carcinogenicity based on mechanistic data. *Molecular structure* information was used to get *chemometrics* information to calculate molecular descriptors (254 MDL and 784 Dragon descriptors), which were further used in predictive QSAR modeling. The dataset presented in the paper can be used in future research in oncology, ecology, or chemicals' risk assessment.

## 1. Introduction

Rodent carcinogenicity datasets were used to build models to predict carcinogenicity within EC-funded project CAESAR (Project no. 022674 (SSPI)) [1]. CAESAR project was aimed to develop quantitative structure-activity relationship (QSAR) models for the REACH (Registration, Evaluation, Authorization, and restriction of CHemicals) legislation for five endpoints: bioconcentration factor, skin sensitization, carcinogenicity, mutagenicity, and developmental toxicity. REACH regulation requires the evaluation of the risks resulting from the use of chemicals produced in industry and testing of their toxicity. Carcinogenicity is among the toxicological endpoints that pose the highest public concern. The standard bioassays in rodents used to assess the carcinogenic potency of chemicals are time consuming and costly and require the sacrifice of large number of animals. Cancer bioassays should be reduced according to REACH regulation [2], while the Seventh Amendment to the EU cosmetics directive will ban the bioassay for cosmetic ingredients from 2013 [3].

The aim of CAESAR project was to reduce the use of animals as well as the cost associated with toxicity tests.

The models predicting carcinogenicity meet the requirements for QSAR models used for regulatory use. Great attention was paid to the quality of data used to build the models; the models were then validated. They are transparent and reproducible and are checked against the OECD principles.

The models at the CAESAR's website have been implemented in java and are freely accessible for public use [1].

Models for prediction of carcinogenicity using rodent carcinogenicity database were described [4–8]. State of the art and perspectives of predictive models for carcinogenicity are discussed in the paper by Benfenati et al. [9].

## 2. Methodology

The chemicals involved in the study belong to different chemical classes, so-called noncongeneric substances. The aim was to cover chemical space as much as possible. The list of 805 chemicals (see Dataset Item 1 (Table)) was extracted from rodent carcinogenicity study findings for 1481 chemicals taken from Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network, which was built from the Lois Gold Carcinogenic Database (CPDBAS)

TABLE 1: A list of 33 SAs for carcinogenicity extracted from Toxtree with number of chemicals in carcinogenicity dataset.

Structural Alert (SA)	Number of Chemicals
SA2: alkyl (C < 5) or benzyl ester of sulphonic or phosphonic acid	4
SA3: N-methylol derivatives	2
SA4: monohaloalkene	5
SA5: S or N mustard	7
SA6: propiolactones or propiosultones	0
SA7: epoxides and aziridines	22
SA8: aliphatic halogens	47
SA9: alkyl nitrite	1
SA10: a, b unsaturated carbonyls	0
SA11: simple aldehyde	4
SA12: quinones	22
SA13: hydrazine	32
SA14: aliphatic azo and azoxy	7
SA15: isocyanate and isothiocyanate groups	4
SA16: alkyl carbamate and thiocarbamate	5
SA17: thiocarbonyl	18
SA18: polycyclic aromatic hydrocarbons	12
SA19: heterocyclic polycyclic aromatic hydrocarbons	5
SA20: (poly)halogenated cycloalkanes	9
SA21: alkyl and aryl N-nitroso groups	107
SA22: azide and triazene groups	3
SA23: aliphatic N-nitro group	1
SA24: a, b unsaturated aliphatic alkoxy group	1
SA25: aromatic nitroso group	4
SA26: aromatic ring N-oxide	1
SA27: nitroaromatic	75
SA28: primary aromatic amine, hydroxyl amine, and its derived esters	52
SA28bis: aromatic mono- and dialkylamine	7
SA28ter: aromatic N-acyl amine	17
SA29: aromatic diazo	4
SA30: coumarins and furocoumarins	8
SA31a: halogenated benzene	16
SA31b: halogenated PAH	7
SA31c: halogenated dibenzodioxins	2

([http://www.epa.gov/ncct/dsstox/sdf\\_cpdbas.html](http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html)) [10]. We have used version 3b of CPDBAS (CPDBAS\_v3b\_1481\_10Apr2006), while in the present moment, it is updated to version 5d, revised on 20 November 2008, and now contains 1547 chemicals.

It should be stressed that in order to obtain data suitable for QSAR modeling the initial dataset (1481 chemicals) has been cleaned of all incorrect structures, ambiguous or mixed structures, polymers, inorganic compounds, metallo-organic compounds, salts, and complexes and compounds without well-defined structure. The obtained data and structures of chemicals were cross-checked by at least two partners using the following online databases: ChemFinder [11], ChemID-plus [12], and PubChem Compound [13]. We selected chemicals with available information about carcinogenic potency in rats. Thus, the final dataset of 805 chemicals, with their ID

number, chemical name, CASRN, experimental  $TD_{50}$  values for rat, and corresponding binary carcinogenicity classes (P: positive; NP: not positive), are available in Dataset Item 1 (Table). For each substance, it is indicated whether it belongs to training or test set.

Rat data only was suggested to be used because a dataset based on data for single species is more consistent and has less variation than a dataset based on two or more species.

Additionally, in our latest study we complemented the dataset with the following alerts collected from Toxtree program: GA, genotoxic alert; nGA, non-genotoxic alert; and NA, no carcinogenic alert. Structural alerts (SAs) for carcinogenicity indicating possible mechanism of carcinogenicity were also collected for each chemical and are presented in Dataset Item 2 (Table), and the list of SAs for carcinogenicity is presented in Table 1 (for more detailed

explanation of terms listed, see [http://www.epa.gov/ncct/dsstox/sdf\\_cpdbas.html](http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html)). The Toxtree expert system with the 33 SAs for carcinogenicity was reported in the Benigni/Bossa rulebase for mutagenicity and carcinogenicity [14]. In a broad sense, the set of chemicals characterized by the same SA could compose a family of compounds with the same mechanism of action (see the recent review written by Benigni and Bossa [15]). From our point of view, SA for carcinogenicity is valuable information in mechanistic interpretation of models [8].

To prepare data for modeling, the dataset of 805 chemicals was subdivided into training (644 chemicals) and test (161 chemicals) sets using the subsorting of chemicals according to a hierarchical system of compound classes in relation to functional group (within classes, the compounds were sorted according to halogen substitution, aromaticity, bond orders, ring contents, and number of atoms), and the following procedure aimed to distinguish between connectivity aspects. This sorting of compounds was implemented with the software system ChemProp [16, 17].

External validation of models was performed using external validation set of 738 chemicals different from those in our dataset of 805 compounds described earlier [4]. ChemFinder Ultra 10.0 software was used [18].

Nowadays, thousands of chemical descriptors such as constitutional, quantum chemical, topological, geometrical, charge related, semiempirical, thermodynamic, and others can be calculated for a given chemical structure [19, 20]. In the present study, the following sets of descriptors for 805 compounds were generated for modeling: 254 MDL descriptors computed using MDL QSAR version 2.2. [21] and 835 Dragon descriptors calculated by DRAGON professional 5.4 software [22].

The obtained descriptors include physicochemical, electrotopological E-state, connectivity, and other descriptors. It should be noticed that E-state indices are a combination of electronic, topological, and valence state information [23–25].

To develop robust and reliable models, the descriptors' space should be reduced by extracting the most significant variables correlated with carcinogenicity. The *Hybrid Selection Algorithm (HSA)* method was used to select among the different molecular descriptors series the best parameters to classify chemicals by their carcinogenic potency. It combines the Genetic Algorithm (GA) concepts and a stepwise regression [26]. In this way, the descriptors' space was reduced from 254 to 8 MDL descriptors [4]. Thus, we used topological descriptors, including atom-type and group-type, E-State and hydrogen E-state indices, molecular connectivity, Chi indices, and topological polarity to obtain the molecular structure information, which is correlated with carcinogenic potency. Among the 8 MDL descriptors, there are two connectivity indices (dpx9 and nxch6), three constitutional parameters (SdssC\_acnt, SdsN\_acnt, and SHBint2\_acnt), and three electrotopological parameters (SdsCH, Gmin, and SHCsats).

Selection of Dragon descriptors was performed using cross-correlation matrix, multicollinearity, and fisher ratio techniques [27, 28].

Among statistical approaches such as linear multivariate regressions, GMDH, and fuzzy logic, artificial neural networks (ANNs), particularly the CP ANN, appeared to be one of the most suitable approaches to predict the complex endpoint such as carcinogenicity for noncongeneric datasets of chemicals with the most reproducible results. The main advantage of neural network modeling is that the complex, nonlinear relationships can be modeled without any assumptions about the form of the model. Large datasets can be examined. Neural networks are able to cope with noisy data and are fault tolerant. However, the interpretation of the acquired knowledge is often a challenge [29].

The detailed description of CPANN can be found in the literature [30–33]. The models used to predict carcinogenicity using 8 MDL descriptors as well as 12 Dragon descriptors and their characterization have been published [4].

### 3. Dataset Description

The dataset associated with this Dataset Paper consists of 7 items, which are described as follows.

*Dataset Item 1 (Table).* A list of 805 chemicals from CPDBAS used for carcinogenicity modeling with indication of training and test sets, which were extracted from the original dataset of 1481 chemicals downloaded from Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network ([http://www.epa.gov/ncct/dsstox/sdf\\_cpdbas.html](http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html)). The column ID\_v5 presents the codes of the chemicals used in CAESAR project (ID of chemicals in database version 5); ID\_CPDBAS-Original, the ID number taken from Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network version 3b; Chemical Name, the chemical names taken from DSSTox and double checked from PubChem Compound (NCBI) (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pccompound>); CASRN, the registry number of the Chemical Abstract Service taken from DSSTox and double checked from PubChem Compound (NCBI). In the column Carcinogenic Potency Expressed as TD<sub>50</sub>, TD<sub>50</sub> is the dose rate in milligram per kilogram of body weight per day, which, if administered chronically for the standard lifespan of the species, will halve the probability of remaining tumorless throughout that period. The TD<sub>50</sub> value reported is the harmonic mean of the most potent TD<sub>50</sub> values from each positive experiment in the species. All the values were derived from the Carcinogenic Potency Database (<http://potency.berkeley.edu/cpdb.html>). In the column Carcinogenic Potency Expressed as P or NP, "P" means positive or active (carcinogens) and "NP" means not positive or inactive (noncarcinogens). In the column Set, "Training" is for the training set and "Test" is for the test (prediction) set.

Column 1: ID\_v5

Column 2: ID\_CPDBAS\_Original

Column 3: Chemical Name

Column 4: CASRN

Column 5: Carcinogenic Potency Expressed as TD<sub>50</sub>  
(mg kg<sup>-1</sup> d<sup>-1</sup>)

Column 6: Carcinogenic Potency Expressed as P or NP

Column 7: Set

*Dataset Item 2 (Table).* A list of the same 805 chemicals from CPDBAS with additional chemical information for studied compounds and detailed information about carcinogenic potency by results on animal tests (rats and mice). Structural alerts (SAs) for carcinogenicity extracted from Toxtree are included in last columns. The structural diversity of CAESAR dataset of 805 chemicals by the presence of specific structural alerts (SAs) extracted from Toxtree program with the number of chemicals in carcinogenicity dataset is presented in Table 1. In Dataset Item 2 (Table), the column ID\_v5 presents the ID number used in the model; ID\_CPDBAS-Original, the ID number in CPDBAS; STRUCTURE\_Formula, the empirical molecular formula; STRUCTURE\_MolecularWeight, the molecular weight or molar mass (atomic mass units); TestSubstance\_ChemicalName, the common or trade name of chemical; and TestSubstance\_CASRN, the Chemical Abstracts Service (CAS) Registry Number of the tested substance. In the column STRUCTURE\_ChemicalName\_IUPAC, IUPAC (International Union of Pure and Applied Chemistry) refers to the standardized nomenclature of organic chemistry. The column STRUCTURE\_SMILES presents the Simplified Molecular Input Line Entry System (SMILES) molecular text code of displayed STRUCTURE. In the columns TD50\_Rat\_mg, TD50\_Rat\_mmol, TD50\_Mouse\_mg, and TD50\_Mouse\_mmol, TD<sub>50</sub> is a standardized quantitative measure of carcinogenic potency (analogous to an LD<sub>50</sub>) and is computed in the CPDB for each species/sex/tissue/tumor type for each experiment (see <http://potency.berkeley.edu/td50harmonicmean.html>). In the columns TargetSites\_Rat\_Male, TargetSites\_Rat\_Female, TargetSites\_Rat\_BothSexes, TargetSites\_Mouse\_Male, TargetSites\_Mouse\_Female, and TargetSites\_Mouse\_BothSexes, target sites (e.g., liver, lung, etc.) are reported for each sex-species group with a positive result in the CPDB (see <http://potency.berkeley.edu/pathology.table.html>). The column NTP\_TechnicalReport presents the National Toxicology Program Technical Report number of study; Website URL, the Internet URL website address for chemical-specific data or content; Alert Type, the structural alert (SA) for carcinogenicity, where GA stands for genotoxic alert, nGA stands for non-genotoxic alert, and NA stands for no alert; Alert 1, the structural alert (SA1) for carcinogenicity, the first SA in molecule; Alert 2, the structural alert (SA2) for carcinogenicity, the second SA in molecule.

Column 1: ID\_v5

Column 2: ID\_CPDBAS\_Original

Column 3: STRUCTURE\_Formula

Column 4: STRUCTURE\_MolecularWeight

Column 5: TestSubstance\_ChemicalName

Column 6: TestSubstance\_CASRN

Column 7: STRUCTURE\_ChemicalName\_IUPAC

Column 8: STRUCTURE\_SMILES

Column 9: TD50\_Rat\_mg (mg kg<sup>-1</sup> d<sup>-1</sup>)

Column 10: TD50\_Rat\_mmol (mmol kg<sup>-1</sup> d<sup>-1</sup>)

Column 11: TargetSites\_Rat\_Male

Column 12: TargetSites\_Rat\_Female

Column 13: TargetSites\_Rat\_BothSexes

Column 14: TD50\_Mouse\_mg (mg kg<sup>-1</sup> d<sup>-1</sup>)

Column 15: TD50\_Mouse\_mmol (mmol kg<sup>-1</sup> d<sup>-1</sup>)

Column 16: TargetSites\_Mouse\_Male

Column 17: TargetSites\_Mouse\_Female

Column 18: TargetSites\_Mouse\_BothSexes

Column 19: NTP\_TechnicalReport

Column 20: Website URL

Column 21: Alert Type

Column 22: Alert 1

Column 23: Alert 2

*Dataset Item 3 (Chemical Structure Data).* Collection of SDF files for 805 chemicals listed in carcinogenicity dataset. To create QSAR models, we calculated chemical descriptors using chemical structures.

*Dataset Item 4 (Table).* Values of 254 MDL descriptors for 805 chemicals.

Column 1: ID\_v5

Column 2: ID\_CPDBAS\_Original

Column 3: (1) SsCH3

:

Column 254: (252) totop

Column 255: (253) Wt

Column 256: (254) nclass

*Dataset Item 5 (Table).* A list of 254 MDL descriptors with their signs and definitions.

Column 1: ID

Column 2: MDL Number

Column 3: Descriptors' Sign

Column 4: Definition

Column 5: Class

*Dataset Item 6 (Table).* Values of 784 Dragon descriptors for 805 chemicals.

Column 1: ID\_v5

Column 2: DRA0001 MW

Column 3: DRA0002 AMW

:

Column 783: DRA0833 MLOGP2

Column 784: DRA0834 ALOGP

Column 785: DRA0835 ALOGP2

*Dataset Item 7 (Table).* A list of 784 Dragon descriptors with their signs and definitions.

Column 1: Internal Code

Column 2: Symbol

Column 3: Definition

Column 4: Class

#### 4. Concluding Remarks

The CPDB rodent carcinogenic database was used for the development of models for the categorization of carcinogenic potency. Initial preprocessing of data and selection of data with carcinogenic potency for rats give us consistent data suitable for QSAR modeling with carcinogenic potency response closer to human. The MDL and Dragon software programs were applied for calculating the molecular descriptors. The topological structure descriptors provided sound bases for classifying molecular structures.

The CP ANN model for prediction of carcinogenicity demonstrated good prediction statistics on the test set of 161 compounds with sensitivity of 75% and specificity of 61%–69% in addition to accuracy of 69%–73%. A diverse external validation set of 738 compounds confirmed the robustness of our models regarding a large applicability domain, yielding the accuracy 60.0%–61.4%, sensitivity 61.8%–64.0%, and specificity 58.4%–58.9%.

The carcinogenicity models presented in the study [4] can be used as a support in risk assessment, for instance, in setting priorities among chemicals for further testing. The dataset and additional information presented in the paper can be used in the QSAR modeling, in future research in oncology, and in risk assessment of chemicals.

#### Dataset Availability

The dataset associated with this Dataset Paper is dedicated to the public domain using the CC0 waiver and is available at <http://dx.doi.org/10.1155/2013/361615/dataset>. In addition, the data presented in Dataset Item 1 (Table) is taken from supplement information Table 1SI of [4] accessible at <http://journal.chemistrycentral.com/content/4/S1/S3/suppl/S1>.

#### Conflict of Interests

The authors declare that they have no conflict of interests.

#### Acknowledgments

The financial support of the European Union through CAESAR project (SSPI-022674) as well as of the Slovenian Ministry of Higher Education, Science and Technology (Grant P1-017) is gratefully acknowledged.

#### References

- [1] CAESAR Project, <http://www.caesar-project.eu/>.
- [2] EC, “Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Directive 78/786/EEC on the approximation of the laws of the member states relating to cosmetic products,” *Official Journal of the European Union L*, vol. 66, pp. 26–35, 2003.
- [3] EC, “Corrigendum to regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006,” *Official Journal of the European Union L*, vol. 136, no. 50, pp. 3–280, 2007.
- [4] N. Fjodorova, M. Vračko, M. Novič, A. Roncaglioni, and E. Benfenati, “New public QSAR model for carcinogenicity,” *Chemistry Central Journal*, vol. 4, supplement 1, article S3, 2010, <http://journal.chemistrycentral.com/content/4/S1/S3/>.
- [5] N. Fjodorova, M. Vračko, M. Tušar et al., “Quantitative and qualitative models for carcinogenicity prediction for non-congeneric chemicals using CP ANN method for regulatory uses,” *Molecular Diversity*, vol. 14, no. 3, pp. 581–594, 2010.
- [6] N. Fjodorova, M. Vračko, A. Jezierska, and M. Novič, “Counter propagation artificial neural network categorical models for prediction of carcinogenicity for non-congeneric chemicals,” *SAR and QSAR in Environmental Research*, vol. 21, no. 1-2, pp. 57–75, 2010.
- [7] N. Fjodorova, M. Novič, A. Roncaglioni, and E. Benfenati, “Evaluating the applicability domain in the case of classification predictive models for carcinogenicity based on the counter propagation artificial neural network,” *Journal of Computer-Aided Molecular Design*, vol. 25, no. 12, pp. 1147–1158, 2011, <http://link.springer.com/article/10.1007%2Fs10822-011-9499-9>.
- [8] N. Fjodorova and M. Novič, “Some findings relevant to the mechanistic interpretation in the case of predictive models for carcinogenicity based on the counter propagation artificial neural network,” *Journal of Computer-Aided Molecular Design*, vol. 25, no. 12, pp. 1159–1169, 2011, <http://link.springer.com/article/10.1007%2Fs10822-011-9500-7>.
- [9] E. Benfenati, R. Benigni, D. M. Demarini et al., “Predictive models for carcinogenicity and mutagenicity: frameworks, state-of-the-art, and perspectives,” *Journal of Environmental Science and Health. Part C*, vol. 27, no. 2, pp. 57–90, 2009.
- [10] CPDBAS: Carcinogenic Potency Database Summary Tables—All Species, 2008, [http://www.epa.gov/ncct/dsstox/sdf\\_cpdbas.html](http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html).
- [11] ChemFinder, <http://www.cambridgesoft.com/databases/login/?serviceid=128>.
- [12] ChemIDPlus, <http://chem.sis.nlm.nih.gov/chemidplus/>.
- [13] PubChem, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pccompound>.
- [14] R. Benigni, C. Bossa, N. Jeliakova, T. I. Netzeva, and A. P. Worth, The Benigni/Bossa rulebase for mutagenicity and carcinogenicity- a module of Toxtree. EUR 23241 EN: 1-70, 2008.
- [15] R. Benigni and C. Bossa, “Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology,” *Chemical Reviews*, vol. 111, no. 4, pp. 2507–2536, 2011.
- [16] G. Schüürmann, R. Kühne, F. Kleint, R. Ebert-U, C. Rothenbacher, and P. Herth, “A software system for automatic chemical property estimation from molecular structure,” in

- Quantitative Structure-Activity Relationships in Environmental Sciences*, F. Chen and G. Schüürmann, Eds., pp. 93–114, SETAC Press, Pensacola, Fla, USA, 1997.
- [17] G. Schüürmann, R. Ebert-U, M. Nendza, J. C. Dearden, A. Paschke, and R. Kühne, “Prediction of fate-related compound properties,” in *Risk Assessment of Chemicals. An Introduction*, K. van Leeuwen and T. Vermeire, Eds., pp. 375–426, Springer Science, Dordrecht, The Netherlands, 2007.
- [18] ChemFinder Ultra 10.0: CambridgeSoft Corp., Cambridge, Mass, USA. FDA SAR Carcinogenicity database, Leadscope Inc., Columbus, Ohio, USA, 2009.
- [19] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Willey-VCH, New York, NY, USA, 2000.
- [20] I. V. Tetko, J. Gasteiger, R. Todeschini et al., “Virtual computational chemistry laboratory—design and description,” *Journal of Computer-Aided Molecular Design*, vol. 19, no. 6, pp. 453–463, 2005.
- [21] MDL-QSARv version 2.2., 2002–2004. MDL Information Systems Inc., San Leandro, Calif, USA, 94577, <http://www.drugdiscoveryonline.com/buyersguide.mvc/sponsors>.
- [22] Dragon home page, [http://www.taletе.mi.it/products/dragon\\_description.htm](http://www.taletе.mi.it/products/dragon_description.htm).
- [23] L. B. Kier and L. H. Hall, *Structure Description: The Electrotopological State*, Academic Press, New York, NY, USA, 1999.
- [24] L. B. Kier and L. H. Hall, “Database organization and searching with E-State indices,” *SAR and QSAR in Environmental Research*, vol. 12, no. 1–2, pp. 55–74, 2001.
- [25] L. B. Kier and L. H. Hall, “The electrotopological state: structure modeling for QSAR and database analysis,” in *Topological Indices and Related Descriptors in QSAR and QSPR*, J. Devillers and A. T. Balaban, Eds., pp. 491–562, Gordon and Breach, Reading, UK, 1999.
- [26] F. Ros, M. Pintore, and J. R. Chrétien, “Molecular descriptor selection combining genetic algorithms and fuzzy logic: application to database mining procedures,” *Chemometrics and Intelligent Laboratory Systems*, vol. 63, no. 1, pp. 15–26, 2002.
- [27] T. Hill and P. Lewicki, *STATISTICS Methods and Applications*, StatSoft, Tulsa, Okla, USA, 2007.
- [28] Electronic Version, *Electronic Statistics Textbook*, StatSoft, Tulsa, Okla, USA, 2007, <http://www.statsoft.com/textbook/>.
- [29] J. Taskinen and J. Yliruusi, “Prediction of physicochemical properties based on neural network modelling,” *Advanced Drug Delivery Reviews*, vol. 55, no. 9, pp. 1163–1183, 2003.
- [30] J. Zupan, M. Novič, and I. Ruisánchez, “Kohonen and counter-propagation artificial neural networks in analytical chemistry,” *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 1, pp. 1–23, 1997.
- [31] J. Zupan and J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, Weinheim, Germany, 2nd edition, 1999.
- [32] J. Zupan, M. Novič, and I. Ruisánchez, “Kohonen and counter-propagation artificial neural networks in analytical chemistry,” *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 1, pp. 1–23, 1997.
- [33] P. Mazzatorta, M. Vračko, A. Jezierska, and E. Benfenati, “Modeling toxicity by using supervised Kohonen neural networks,” *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 2, pp. 485–492, 2003.



**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

