Comparative Study of Classification Algorithms Using Molecular Descriptors in Toxicological DataBases

Max Pereira[1], Vítor Santos Costa[3], Rui Camacho[1], Nuno A. Fonseca[2,3], Carlos Simões[4], and Rui Brito[4]

[1] LIAAD-INESC Porto LA & FEUP, Universidade do Porto,
Rua Dr Roberto Frias s/n, 4200-465 Porto, Portugal
[2] Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto
Rua do Campo Alegre 823, 4150-180 Porto, Portugal
[3] CRACS-INESC Porto LA, Universidade do Porto,
Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal
[4] Chemistry Department, Faculty of Science and Technology
and Center for Neuroscience and Cell Biology
University of Coimbra, Portugal

Abstract. The rational development of new drugs is a complex and expensive process, comprising several steps. Typically, it starts by screening databases of small organic molecules for chemical structures with potential of binding to a target receptor and prioritizing the most promising ones. Only a few of these will be selected for biological evaluation and further refinement through chemical synthesis. Despite the accumulated knowledge by pharmaceutical companies that continually improve the process of finding new drugs, a myriad of factors affect the activity of putative candidate molecules *in vivo* and the propensity for causing adverse and toxic effects is recognized as the major hurdle behind the current "target-rich, lead-poor" scenario. In this study we evaluate the use of several Machine Learning algorithms to find useful rules to the elucidation and prediction of toxicity using 1D and 2D molecular descriptors. The results indicate that: i) Machine Learning algorithms can effectively use 1D molecular descriptors to construct accurate and simple models; ii) extending the set of descriptors to include 2D descriptors improve the accuracy of the models.

1 Introduction

The amount of information concerning chemicals that is available in databases has been increasing at a considerable pace in the last years, changing the whole process of discovery and development of new drugs. These databases have been used as a starting point for screening candidate molecules, and enable the pharmaceutical industry to produce over 100,000 new compounds per year [1]. The promising compounds are further analysed in the development process, where, among other investigations, their potential toxicity is assessed. This is a complex

and costly process that often requires years before compounds can be tested in human subjects [2–4]. Additionally, about 90% of the initially considered drugs fail to reach the market due to toxicological properties [5]. This fact highlights the importance of determining as early as possible toxicological features.

Toxicity tests determine whether or not a candidate molecule is likely to produce toxic effects in humans, usually involve the use of animal models at a pre-clinical stage. As the number of of biological targets identified as relevant increases (resulting from the humane genome project), and hence the demand for drug screening campaigns there is a growing need of efficient *in silico* methods for the prediction of toxicity of organic compounds. The problem is to identify clear relationships between a molecule's chemical structure and its toxicological activity. These relationships can be used to build predictive models to apply to new compounds [6]. Ultimately, this task can be regarded as a method to predict Quantitative Structure-Activity Relationships (QSARs) [7] which considers a set of structural features associated with a toxicity endpoint. These models offer an inexpensive and fast way of estimating molecules toxicological properties [8, 9].

The problem of estimating the toxicity of drugs has been addressed, mainly, from three methods: i) regression from physical-chemical properties; ii) expert systems and; iii) machine learning [10, 11]. Some toxicity prediction programs are commercially available including TOPKAT (toxicity-prediction by computer-assisted technology), DEREK (deductive estimation of risk from existing knowledge), CSGenoTox, MetaDrug and HazardExpert [12]. These programs have a common characteristic, they are classified as "global" models [8] since they were developed using a non-congeneric set of chemicals. Actually it is not mandatory that the chemicals in these data sets are congeneric, but they should share structural features. Besides the commercially available programs, other studies have been published using machine learning approaches [3, 13, 10, 6, 11].

In this paper we compare the performance of classification algorithms in predicting the toxicity of compounds and present the results for data sets composed of constitutional molecular descriptors exclusively (*1D descriptors*) and data sets composed of both constitutional and topological molecular descriptors (*2D descriptors*). Although similar studies have been reported [3, 4, 10, 14], they did not assess the relevancy of molecular descriptors in terms of toxicity prediction. We have applied the classification algorithms to predict the toxicity of compounds for the estrogen receptor (DSSTox NCTRER DataBase), mutagenicity using the carcinogenic potency database (DSSTox CPDBAS DataBase), the Fathead Minow Acute Toxicity (DSSTox EPAFHM DataBase) and the Disinfection By-products Carcinogenicity Estimates (DSSTox DBPCAN DataBase).

In this study we aim at proving/disproving the following hypotheses:

HC0: 1D descriptors contain sufficient information for Machine Learning algorithms to construct accurate and robust predictive models that can predict *whether* a given molecule is toxic.

HC1: Extending the set of 1D descriptors with 2D descriptors improves the accuracy of the models produced by Machine Learning algorithms to predict the degree of toxic activity of molecules.

HR0: 1D descriptors contain sufficient information for Machine Learning algorithms to construct accurate and simple predictive models of the *degree* of toxic activity of molecules.

HR1: Extending the 1D set of descriptors of molecules with 2D descriptors improves the models constructed by Machine Learning algorithms to predict the *degree* of toxic activity of molecules.

The remaining portion of the paper has the following structure. Section 2.1 describes the data sets used, and Section 2.2 gives an overview on the Machine Learning algorithms used in this study. Section 3 details the experiments undertaken and discussed the results obtained. We compare our work with previously published work in Section 4. Conclusions are presented in the last section of the paper.

2   Material and Methods

2.1   Data Sets

We used four data sets available from the Distributed Structure-Searchable Toxicity (DSSTox) Public DataBase Network [15] from the U.S.Environmental Protection Agency[5]. The DSSTox database project is targeted to toxicology studies and uses a standard chemical structure annotation. The data sets are characterised as follows.

CPDB: The Carcinogenic Potency DataBase (CPDB) contains detailed results and analyses of 6540 chronic, long term carcinogenesis bioassays [16], which currently contains 1547 compounds. For the purpose of this study the carcinogenicity endpoint was evaluated concerning hamster, mouse and rat species. The experimental results on the remaining species (cynomolgus, dog, rhesus) are insufficient and were discarded. When the same compound was tested in more than one specie, the experimental results for all species were stored in a single entry. Thus, for this study, the database was pre-processed in order to expand a single entry to produce an entry for each specie. This pre-processing resulted in a total of 2272 entries.

EPAFHM: The Fathead Minnow Acute Toxicity database was generated by the U.S. EPA Mid-Continental Ecology Division (MED) for the purpose of developing an expert system to predict acute toxicity from chemical structure based on mode of action considerations [14]. The database contains 614 organic chemicals, for the prediction of acute toxicity endpoint of environmental and industrial chemicals.

NCTRER: Researchers within FDA's National Center for Toxicological Research (NCTR) generated a database of experimental ER (estrogen receptor) binding results with the purpose of developing improved QSAR models to predict ER binding affinities. The NCTRER database provides activity classifications for a total of 224 chemical compounds, with a diverse set of natural, synthetic and environmental estrogens [17].

---

[5] http://www.epa.gov/ncct/dsstox/index.html, accessed Dec 2008

DBPCAN: Water disinfection by-products database contains predicted estimates of carcinogenic potential for 178 chemicals.The goal is to provide informed estimates of carcinogenic potential to be used as one factor in ranking and prioritising future monitoring, testing, and research needs in the drinking water area [18].

The structures of chemicals in DSSTox are stored as SDF [6] files as well as SMILES[7] strings.

In addition to the original database entries, a number of 50 molecular descriptors was calculated with the GenerateMD software[8]. Molecular descriptors belong to a set of pre-defined categories [12, 19]. In our data sets we used the constitutional-based descriptors, also known as *1D descriptors*, and the topological-based descriptors, also called *2D descriptors*.

Furthermore, we generated molecular *fingerprints* of type FP4[9] using Open Babel software [20]. These fingerprints were then converted into binary attributes meaning the presence or absence of a particular chemical substructure in the compound, such as, fused rings, alkene, lactone, enolether, in a total of 300 chemical substructures.

All the information was encoded in ARFF format which provides the necessary information to run the classification and regression algorithms. The classification data sets were obtained from the regression ones by establishing a threshold of 50% for the toxicity activity: below the 50 % the drug was not considered toxic and above that value it was considered toxic.

data set	active	inactive
CPDB	1059	1213
EPAFHM	580	34
NCTRER	131	93
DBPCAN	80	98

(a)

Type of Features	Number of Attributes
1D	22
1D+FP	322
2D	564

(b)

Table 1. Characterisation of the data sets. (a) Class distribution. (b) Number of Attributes.

Table 1(a) summarizes the four datasets used in this study. Except for the EPAFHM dataset all the datasets are balanced in terms of the number of active and inactive compounds. In all the experiments, we used the same number of attributes: 22 pure 1D descriptors, 300 molecular fingerprints, and 242 2D descriptors (Table 1(b)).

2.2 Machine Learning methods

In this study we used classification and regression methods. The classification methods used cover most popular methods in Machine Learning and include

---

[6] Structure Data Format
[7] Simplified Molecular Input Line Entry Specification.
[8] http://www.chemaxon.com, accessed Oct 2008
[9] a set of SMARTS queries

several variants of decision trees, Bayesian methods, instance base classification, and a variant of Support Vector Machines (SVMs). We experimented with different decision trees as they are widely considered one of the most interpretable methods, and interpretability may be of interest in this experiment. Regarding regression, we used approaches based in trees and based on SVMs.

All the machine learning methods used in our study are implemented in the Weka [21] software package. The classification methods used are enumerated in Table 2 and the list of regression methods used is enumerated in Table 3. The tree construction methods we have used are BF trees, CART, J48, ADTrees, and Functional Trees. We used Breiman's "random forests" as a popular methods that benefits from some of the understandability of trees and has been shown to achieve quite good accuracies. We also used two bayes methods: Naive Bayes is usually a good reference, and the "Bayes Network" method uses a greedy algorithm to construct the network. The K Star algorithm is a version of the famous $K - NN$ instance based classifier, and, last but not least, we used Weka's SMO SVM implementation.

Method (Weka name)	Type
BF Tree	Decision Trees.
CART	Decision Trees
SVM SMO	Support Vector Machines
Naive Bayes	Bayesian classifier
Bayes Network	Bayesian classifier
K Star	Instance based classifier
AD Tree	Alternate Decision Trees
FT	Functional trees
J48	Decision tree
Random Forest (RF)	Ensemble method

Table 2. Classification methods used in the present work

We have used four regression methods: M5P and M5rules are based on model trees; in contrast, SMO is based on SVM technology, Linear Regression is a well known statistical method.

Method (Weka name)	Type
M5P	Model Trees.
M5rules	Model Trees.
SVM SMO	Support vector machines
Linear Regression	statistical method

Table 3. Regression methods used in the present work

2.3 Experimental Design

The experiments were carried out on a cluster of 9 nodes where each node has two quad-core Xeon 2.4GHz and 32 GB of RAM and runs Linux Ubuntu 8.10. The Weka [21] version used was 3.6.0.

To estimate the predictive quality of the classification models we performed 10 fold cross-validation. The quality of the regression models were estimated measuring the Relative Absolute Errors (RAE):

$$RAE = \frac{\sum_{j=1}^{N} |P_{ij} - T_j|}{\sum_{j=1}^{N} |\overline{T_i} - T_j|}$$

in the formula, given the classifier i and the drug j, P_{ij} is the prediction made by i on j, $\overline{T_i}$ is the average value of actual activity over all drugs, and T_j is the actual activity value of drug j.

To handle missing values we considered three approaches: i) let the ML algorithm deal with the missing values; ii) use Weka pre-processing procedure for missing values and; iii) remove entries with missing values.

All ML algorithms were applied for each combination of missing values approach, and dataset (using 1D descriptors, 1D descriptors and Finger Prints (1D+FP) and all the information available).

3 Results and Discussion

3.1 Classification

Table 4 presents the average accuracy over 10 folds obtained by all classification methods considered, in all data sets and with different sets of features (1D, 1D+FP, and 1D+FP+2D)[10].

The last line in the table, for the ZeroR classifier, gives the baseline performance for the other learning methods. The ZeroR simply predicts the majority class in the training data. The values in bold are the best value for the column (for a data set and set of features). Overall, the classifiers are quite accurate on DBP-CAN, where they achieve over 90% accuracy. All classifiers (except `bayesNet`) perform well in this dataset. EPAFHM is a very skewed dataset: predicting base class would give 95% accuracy. Even so, all classifiers exceed default accuracy, in fact the Functional Trees (FT) classifier achieves almost perfect accuracy. Performance is also quite good for NCTRER, between 77–86%, with best results for `ft` and `rf`. The hardest dataset was CPDBAS, where performance ranges from 55% for naive bayes (a little better than default class) up to 73% with random forests. Notice that in contrast to other datasets, nearest neighbor is quite effective in this dataset.

---

[10] The empty cells for NCTRER and BF tree and Cart are due to the incapacity of those two algorithms to handle missing values and all examples of the NCTRER data set have a missing value.

Algorithm	CPDBAS			DBPCAN		
	1D	1D+FP	1D+FP+2D	1D	1D+FP	1D+FP+2D
J48	66.4	64.0	68.7	89.9	93.3	91.6
BFTree	66.1	63.9	66.2	85.4	92.1	89.3
cart	65.6	66.7	69.1	85.9	91.0	90.6
ibk	62.1	59.6	62.4	91.0	92.7	91.0
SMO	58.4	60.9	65.9	**92.8**	92.7	92.1
Nbayes	57.8	58.0	54.4	87.1	89.3	83.7
bayesNet	57.7	56.9	58.4	81.5	88.2	84.3
kStar	70.8	68.9	72.0	88.2	92.7	89.3
ADTree	61.3	60.8	63.3	90.4	93.8	92.7
ft	66.9	**70.3**	70.2	92.7	**94.4**	**95.5**
rf	**71.8**	**70.3**	**72.8**	92.7	93.8	92.7
ZeroR	53.4	53.4	53.4	55.1	55.1	55.1
Algorithm	EPAFHM			NCTRER		
	1D	1D+FP	1D+FP+2D	1D	1D+FP	1D+FP+2D
J48	98.2	98.2	98.2	83.0	81.7	85.57
BFTree	97.9	97.9	97.5	-	-	-
cart	97.9	97.9	97.3	-	-	-
ibk	96.3	95.4	96.1	82.6	**85.7**	83.0
SMO	96.3	94.6	93.8	81.7	85.3	83.5
Nbayes	94.8	93.3	94.6	79.0	83.5	80.8
bayesNet	94.1	92.7	89.1	80.8	81.3	83.5
kStar	97.2	96.2	96.1	79.0	79.0	76.8
ADTree	98.5	98.2	98.2	83.9	81.7	84.4
ft	**99.0**	**98.9**	**98.7**	**85.7**	85.3	86.7
rf	98.7	97.6	97.2	83.9	85.3	**87.1**
ZeroR	94.5	94.5	94.5	58.5	58.5	58.5

Table 4. Performance of classification algorithms. The values are obtained from the Weka package and are the average (over 10 folds) correctly classified instances obtained by different learning algorithms using 1D, 1D plus finger prints (1D+FP) plus 2D features (1D+FP+2D).

Functional trees (FT) performed quite well on these datasets. FT tend to perform close to the best, and for the columns where FT performance is not the best, it is the second or third best. In all cases, the performance obtained by FT is better than the baseline.

Focusing on FT, the performance values for 1D descriptors in the four data sets clearly supports the hypothesis that 1D descriptors contain sufficient information for Machine Learning algorithms to construct accurate and simple predictive models to determine if a given molecule is toxic or not.

Extending the 1D set of descriptors to include fingerprints did not boost performance. However, extending the set of features to include the 2D set of descriptors improves the performance in most cases. The only exception is observed in the EPAFM data set. However, in this case, it was difficult to improve the performance obtained using 1D descriptors since it was already at 99%. Nevertheless, the results support the second hypothesis that states that extending the 1D set of descriptors of molecules with 2D descriptors improves the accuracy of the models constructed by Machine Learning algorithms to predict the degree of toxic activity of molecules.

(a) (b)
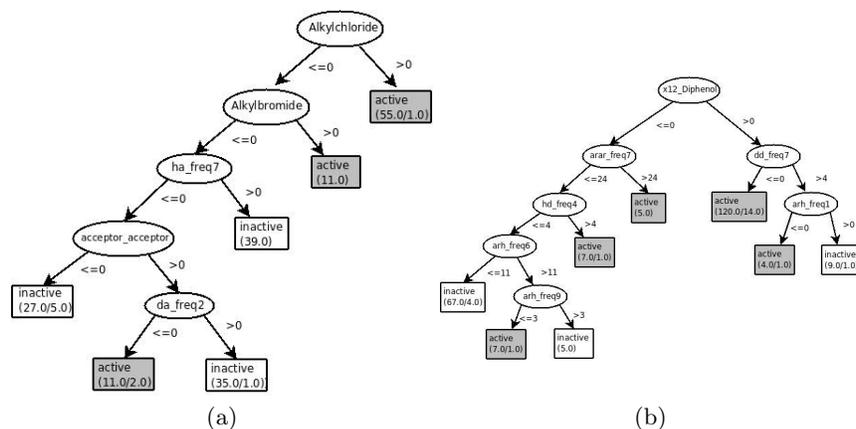
Fig. 1. Decision Tree to predict toxicity for the (a) DBPCAN and (b) NCTRER data sets.

An interesting question is whether the models discovered are supported or contribute to expertise in the area. Figure 3.1 shows two example trees obtained by the J48 algorithm. We chose J48 because it is particularly easy for an experiment to understand, and we chose two applications where J48 performs well.

The tree in Figure 3.1(a) presents a model for DBPCAN. The model is based on halogenated hydrocarbons (e.g., alkylchloride, alkylbromide), which are closely associated to toxicity. That such elements would be present in the model is unsurprising. First, it is well known that organisms have difficulty in metabolising halogens. Indeed the halogens such as fluorine, chlorine, bromine and iodine are elements avoided by medical chemists. Second, halogenated hy-

drocarbons (and aromatic rings such as the ones in diphenol) are hydrophobic. This property allows these molecules to cross biological membranes and deposit in the the fatty tissue, where they eventually can be involved in cancer process (a classical example is DDT).

More curious is that in this model the absence of halogenated hydrocarbons in "active" molecules seems to be compensated by having close-by hydrogen acceptors. ($acceptor\_acceptor > 0$). Donor groups seem to relate with the functionality of drugs (and namely to specificity for certain targets). The existence of a large number of accepting groups with hydrogen bonds (more than 10), that seems to be contemplated in $acceptor\_acceptor > 10$ may be associated with more promiscuous molecules, that is, with molecules connecting to a number of different targets with secondary effects.

Regarding the J48 model for NCTRER, the Diphenol group associated to molecules with a positive sharing coefficient between octanol and water seems to explain molecule activity/toxicity, very much in line with the previous discussion on hydrophobicity.

3.2   Regression

Algorithm	CPDBAS			DBPCAN			EPAFHM			NCTRER		
	1D	1D+FP	1D+FP+2D	1D	1D+FP	1D+FP+2D	1D	1D+FP	1D+FP+2D	1D	1D+FP	1D+FP+2D
m5p	73.6	70.8	67.5	41.5	42.4	45.6	34.6	33.9	36.9	50.2	**46.2**	56.2
m5rules	**72.4**	**68.8**	66.8	**39.3**	41.6	**44.4**	**26.6**	27.8	**24.3**	47.9	51.2	**46.1**
l.regression	95.0	90.9	82.9	52.5	53.4	99.2	110.1	122.1	147.5	61.5	63.1	100.0
svm smo	92.4	89.9	**63.0**	52.8	47.9	51.3	53.2	51.3	68.4	**43.1**	52.6	63.0

Table 5. Performance of the regression algorithms. Results obtained in the regression task by different learning algorithms using 1D, 1D plus finger prints (1D+FP) and 1D+FP plus 2D features (1D+FP+2D). Values represent RAE.

HR0: 1D descriptors contain sufficient information for Machine Learning algorithms to construct accurate and simple predictive models of the degree of toxic activity of molecules.

HR1: Extending the 1D set of descriptors of molecules with 2D descriptors improves the models constructed by Machine Learning algorithms to predict the degree of toxic activity of molecules.

The results of Table 5 show that the confirmation of HR0 depends on the data set. For EPAFHM the m5rules algorithm achieves a RAE of 26.6% that is quite good, whereas in CPDBAS data set the best score is 72%. A similar conclusion may be reached as HR1 is concerned. In CPDBAS data set we see a systematic reduction in RAE in all algorithms when the set of descriptors is enriched with 2D descriptors (with SVM algorithm it drops from 92.4% to 63.0%).

4   Related Work

In this study machine learning algorithms were applied to the task of predicting toxicity endpoints. Other solutions to the prediction toxicology problem have

been published. In this section we report in related work. Notice that, to the best of our knowledge, we are the first the report extensive comparison results for these recent datasets.

In [22] decision tree models were constructed to select from a set of 20 attributes the ones whose values best discriminate a set of 904 rodent bioassay experiments. A classification system TIPT (Tree Induction for Predictive Toxicology) based on the tree was then applied and compared with neural networks models in terms of accuracy and understandability. The classification problem was also the subject of investigation in [23] where the Support vector machine (SVM) proved to be reliable in the classification of 190 narcotic pollutants (76 polar and 114 nonpolar). A selection algorithm was also used to identify three necessary attributes for the compounds discrimination and the leave-one-out cross-validation was the evaluated procedure. Again in [24] the problem is to predict potential toxicity of compounds depending on their physico-chemicals properties. It was used a wide variety of machine learning algorithms with Weka (machine learning software), including classical algorithms, such as k-nearest neighbours and decision trees, as well as support vector machines and artificial immune systems. SVMs proved to be the best algorithm followed by a neural network model. In [10] the ILP (Inductive logic programming) approach was used with support vector machines to extends the essentially qualitative ILP-based SAR to quantitative modelling. In this study a data set of 576 compounds with known fathead minnow fish toxicity was used and the results were compared with the commercial software TOPKAT. Furthermore, in [25] other machine learning approaches was analysed, such as Particle Swarm Optimisation (PSO) and Genetic Programming (GP), they are suitable for use with noisy, high dimensional data, as in commonly used in drug research studies. In [26] a literature review was done focus in predictive models such as partial-least square (PLS), support vector machines, neuronal nets, multiple linear regression and decision trees. A novel model to simulate complex chemical-toxicology data sets was reported in [27] and used to evaluate the performance of different machine learning approach, neuronal networks, k-nearest neighbours, linear discriminant analysis (LDA), naive Bayes, recursive partitioning and regression trees (RPART), and support vector machines.

5 Conclusions

The work reported in this paper addresses the problem of constructing predictive models of toxicity in a drug design setting. We have evaluated the usefulness of Machine Learning algorithms to construct accurate and simple models for toxicity. The study compared the usefulness of 1D and 2D molecular descriptors not only in the prediction of the degree of toxic activity but also the classification problem of predicting if a drug is toxic or not.

The results indicate that Machine Learning algorithms can effectively use 1D molecular descriptors to construct accurate and simple models to predict compound toxicity. The experiments also show that extending the set of 1D

descriptors with 2D descriptors may improve the accuracy of the models, but that further work is required to take full advantage of these features.

Acknowledgements

References

1. Plewczynski, D.: Tvscreen: Trend vector virtual screening of large commercial compounds collections. Biocomputation, Bioinformatics, and Biomedical Technologies, 2008. BIOTECHNO '08. International Conference on (2008) 59–63
2. Graham, J., Page, C., Kamal, A.: Accelerating the drug design process through parallel inductive logic programming data mining. Computational Systems Bioinformatics Conference, International IEEE Computer Society (2003) 400
3. Tiwari, A., Knowles, J., Avineri, E., Dahal, K., Roy, R., eds.: Advances in the Application of Machine Learning Techniques in Drug Discovery, Design and Development. In Tiwari, A., Knowles, J., Avineri, E., Dahal, K., Roy, R., eds.: Applications of Soft Computing: Recent Trends. Advances in Soft Computing, Springer (2006)
4. Duch, W., Swaminathan, K., Meller, J.: Artificial intelligence approaches for rational drug design and discovery. Current Pharmaceutical Design **13** (2007) 1497–1508(12)
5. van de Waterbeemd, H., Gifford, E.: Admet in silico modelling: towards prediction paradise? Nat Rev Drug Discov **2**(3) (2003) 192–204
6. Neagu, D., Craciun, M., Stroia, S., Bumbaru, S.: Hybrid intelligent systems for predictive toxicology - a distributed approach. Intelligent Systems Design and Applications, International Conference on (2005) 26–31
7. Hansch, C., Maloney, P., Fujita, T., Muir, R.: Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. Nature (194) (1962) 178–180
8. White, A., Mueller, R., Gallavan, R., Aaron, S., Wilson, A.: A multiple in silico program approach for the prediction of mutagenicity from chemical structure. Mutation Research/Genetic Toxicology and Environmental Mutagenesis **539** (2003) 77–89(13)
9. Richard, A.: Future of toxicology-predictive toxicology: An expanded view of "chemical toxicity". Chem. Res. Toxicol. **19**(10) (2006) 1257–1262
10. Amini, A., Muggleton, S., Lodhi, H., Sternberg, M.: A novel logic-based approach for quantitative toxicology prediction. J. Chem. Inf. Model. **47**(3) (2007) 998–1006
11. Dearden, J.: In silico prediction of drug toxicity. Journal of computer-aided molecular design **17**(2-4) (2003) 119–127
12. Ekins, S.: Computational Toxicology: Risk Assessment for Pharmaceutical and Environmental Chemicals (Wiley Series on Technologies for the Pharmaceutical Industry). Wiley-Interscience (2007)

13. Kazius, J., Mcguire, R., Bursi, R.: Derivation and validation of toxicophores for mutagenicity prediction. J. Med. Chem. **48**(1) (2005) 312–320
14. Russom, C., Bradbury, S., Broderius, S., Hammermeister, D., Drummond, R.: Predicting modes of toxic action from chemical structure : Acute toxicity in the fathead minnow (pimephales promelas). Environmental toxicology and chemistry **16**(5) (1997) 948–967
15. Richard, A., Williams, C.: Distributed structure-searchable toxicity (dsstox) public database network: a proposal. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **499** (2002) 27–52(26)
16. Gold, L., Manley, N., Slone, T., Ward, J.: Compendium of chemical carcinogens by target organ: Results of chronic bioassays in rats, mice, hamsters, dogs, and monkeys. Toxicologic Pathology **29**(6) (2001) 639–652(14)
17. Fang, H., Tong, W., Shi, L., Blair, R., Perkins, R., Branham, W., Hass, B., Xie, Q., Dial, S., Moland, C., Sheehan, D.: Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. Chem.Res.Toxicol. (14) (2001) 280–294
18. Woo, Y., Lai, D., McLain, J., Manibusan, M., Dellarco, V.: Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products. Environ. Health Perspect (110) (2002) 75–87
19. Todeschini, R., Consonni, V., Mannhold, R., Kubinyi, H., Timmerman, H.: Handbook of Molecular Descriptors. Wiley-VCH (2000)
20. Guha, R., Howard, M., Hutchison, G., Murray-Rust, P., Rzepa, H., Steinbeck, C., Wegner, J., Willighagen, E.: The blue obelisk – interoperability in chemical informatics. J. Chem. Inf. Model. **3**(46) (2006) 991–998
21. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. 2nd edition edn. Morgan Kaufmann (2005)
22. Bahler, D., Stone, B., Wellington, C., Bristol, D.: Symbolic, neural, and bayesian machine learning models for predicting carcinogenicity of chemical compounds. J. Chemical Information and Computer Sciences (8) (2000) 906–914
23. Ivanciuc, O.: Aquatic toxicity prediction for polar and nonpolar narcotic pollutants with support vector machines. Internet Electronic Journal of Molecular Design (2) (2003) 195–208
24. Ivanciuc, O.: Weka machine learning for predicting the phospholipidosis inducing potential. Current Topics in Medicinal Chemistry (8) (2008)
25. Pugazhenthi, D., Rajagopalan, S.: Machine learning technique approaches in drug discovery, design and development. Information Technology Journal **5**(6) (2007) 718–724
26. Muster, W., Breidenbach, A., Fischer, H., Kirchner, S., Mller, L., Phler, A.: Computational toxicology in drug development. Drug Discovery Today **8**(7) (2008)
27. Judson, R., Elloumi, F., Setzer, R., Li, Z., Shah, I.: A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. BMC Bioinformatics (2008)