Full Paper

www.molinf.com

molecular informatics

DOI: 10.1002/minf.201400168

Novel Uses of In Vitro Data to Develop Quantitative Biological Activity Relationship Models for in Vivo Carcinogenicity Prediction

Prachi Pradeep,*[a] Richard J. Povinelli,[b] Stephen J. Merrill,[a] Serdar Bozdag,[a] and Daniel S. Sem[c]

Abstract: The availability of large in vitro datasets enables better insight into the mode of action of chemicals and better identification of potential mechanism(s) of toxicity. Several studies have shown that not all in vitro assays can contribute as equal predictors of in vivo carcinogenicity for development of hybrid Quantitative Structure Activity Relationship (QSAR) models. We propose two novel approaches for the use of mechanistically relevant in vitro assay data in the identification of relevant biological descriptors and development of Quantitative Biological Activity Relationship (QBAR) models for carcinogenicity prediction. We demonstrate that in vitro assay data can be used to develop QBAR models for in vivo carcinogenicity prediction via two case studies corroborated with firm scientific rationale. The case studies demonstrate the similarities between QBAR and QSAR modeling in: (i) the selection of relevant descriptors to be used in the machine learning algorithm, and (ii) the development of a computational model that maps chemical or biological descriptors to a toxic endpoint. The results of both the case studies show: (i) improved accuracy and sensitivity which is especially desirable under regulatory requirements, and (ii) overall adherence with the OECD/REACH guidelines. Such mechanism based models can be used along with QSAR models for prediction of mechanistically complex toxic endpoints.

Keywords: Computational toxicology · In vitro data · Biological descriptors · Quantitative biological activity relationship (QBAR) · Carcinogenicity · OECD · REACH

1 Introduction

Chemical Risk Assessment or evaluation of the extent of toxic effects associated with chemical exposure is necessary for protection of human or environmental health. Computational toxicology is the in silico prediction of adverse or toxic effects of chemicals on living organisms. In silico models provide a less expensive, faster and more efficient alternative to otherwise time-consuming conventional animal and clinical testing methods. Quantitative Structure Activity Relationship (QSAR) models are the most widely used alternative to conventional animal and laboratory testing. They are theoretical models that relate a quantitative measure of chemical structure to a physical property or a biological effect. QSAR model development is a 3-step process: (i) generation of molecular descriptors, (ii) selection of relevant molecular descriptors, and (iii) statistical mapping of the descriptors to the toxic endpoint under consideration.[1,2]

QSAR models have been continuously improving with new machine learning algorithms, molecular descriptors and training databases.[3–5] However, several studies show that they are still not very predictive for mechanistically complex endpoints like carcinogenicity.[6,7] These limitations are primarily due to the multiple mechanisms of action associated with more complex toxic endpoints. Furthermore, the OECD principles for QSAR model development emphasize on mechanistic interpretation of results (if possible) in addition to appropriate measures of goodness-of-fit, robustness and predictivity.[8–10] Mechanistic interpretation of toxicity is complex and it is difficult to capture all the aspects of toxicity from a structural perspective. Development of new mechanism based methods and a paradigm shift towards a systems biology based approach towards toxicology is, therefore, a necessity in the future development of computational toxicology.

[a] *P. Pradeep, S. J. Merrill, S. Bozdag*
Department of Mathematics, Computer Science and Statistics, Marquette University
1313 W. Wisconsin Avenue, Milwaukee, WI 53233, USA
fax: (414) 288–5472
**e-mail: prachi.pradeep@marquette.edu*

[b] *R. J. Povinelli*
Department of Electrical and Computer Engineering, Marquette University
1515 W. Wisconsin Avenue, Milwaukee, WI 53233, USA

[c] *D. S. Sem*
School of Pharmacy, Concordia University Wisconsin
12800 N. Lake Shore Drive, Mequon, WI 53097, USA

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/minf.201400168.

2 Quantitative Biological Activity Relationships

Recent advances in the field of "omics" technologies (proteomics, metabolomics, toxicogenomics etc.) offer intriguing avenues for assessing chemical response in in vitro systems. High throughput screening methods facilitate the screening of large number of chemicals against a variety of Environmental Protection Agency's (EPA) ToxCast project[13] and the Tox21 consortium of the U.S. EPA, National Toxicology Program (NTP), National Institutes of Health Chemical Genomics Center (NCGC), and U.S. Food and Drug Administration (FDA)[14,15] are two sources of high throughput in vitro activity data for thousands of chemicals across several biochemical assays. In vitro methods are developed to provide mechanistic insight for building risk assessment strategies. In vitro data can be utilized in several ways to assist in computational modeling approaches to predict toxicity. Firstly, they can offer insight into how different chemicals can alter or perturb certain biochemical pathways that may result in toxic responses. Secondly, they can help in the identification of biological response patterns (biomarkers) associated with different toxic endpoints. Thirdly, they can help in elucidating the mechanism of action involved with various toxic endpoints.[16] Lastly, they can be used to develop biological similarity based computational models for toxicity prediction. The underlying concept is based on the assumption that mechanistically related toxic chemicals will display similar patterns of biological activity in various in vitro assays.[17] Quantitative Biological Activity Relationships (QBAR) can, thus, be defined as theoretical models that relate a quantitative measure of biological similarity to a toxic effect. The underlying principle behind QBAR models is that chemicals with similar biological responses are likely to have similar toxic effects.

Several studies have demonstrated the use of in vitro data in the development of predictive QBAR models for in vivo toxicology.[18–22] The results of these studies for carcinogenicity prediction show that all high throughput assays do not contribute equally as predictors of in vivo carcinogenicity. The report on carcinogenicity prediction trials by the U.S. National Toxicology Program (NTP) states that carcinogenicity is generally a poorly predicted endpoint and makes a guideline that best predictivemodels tend to be those that integrate biological mechanism-based data.[23] This recommendation aligns with the Organization for Economic Co-operation and Development (OECD) principles for use of (Q)SAR models in regulation, which includes a mechanistic interpretation (if possible) among other criteria for model validation.[8,9] Based on these reports, we propose the use of specific in vitro assay data in identification of relevant biological descriptors and development of QBAR models for carcinogenicity prediction. We demonstrate how in vitro data can be used independently to develop predictive models for in vivo carcinogenicity via two case studies. The case study in Section 3.1 demonstrates how we can select relevant in vitro assays as biological descriptors for developing QBAR models (analogous to selection of relevant chemical descriptors for QSAR modeling). The case study in Section 3.2 demonstrates how different in vitro assays for selected endpoints can be used together as biological descriptors for developing a QBAR model (analogous to statistical mapping of chemical descriptors to a toxic endpoint in QSAR modeling).

3 Case Studies

3.1 Identification of a Novel Biological Descriptor Based on Xenobiotic Induced Cytochrome P450 Transcription for Carcinogenicity Prediction

3.1.1 Cytochrome P450 Enzyme System

Cytochrome P450 (CYP) enzymes are the most important enzymes in the metabolism process in mammals and are primarily responsible for the metabolism (degradation and elimination) of xenobiotics. CYP enzymes are subdivided into various families based on the percentage of amino acid sequence identity. The major families are CYP1, CYP2 (with five subfamilies CYP2A-E), and CYP3. There are about 57 identified CYP enzymes that are found to be involved in metabolism reactions. Approximately 75% of the drugs are metabolized by P450s. Out of those, five major isoforms viz., CYP2D6, CYP3A4, CYP2C9, CYP2C19 and CYP1A2 are involved in ~75–90% metabolic reactions. CYP2D6 alone is involved in the metabolism of ~70% marketed drugs.[24–26]

Xenobiotic metabolizing enzymes can help in detoxification by elimination of potential carcinogens or facilitate toxicity by conversion of primary non-carcinogens (procarcinogens) into secondary carcinogenic metabolites. Procarcinogens usually require transformation into a more electrophilic from to cause DNA damage and cancer. Thus, they can be classified into two categories. The first class includes enzymes that are more involved in drug metabolism, such as CYP2A6, CYP2B6, CYP2C9, CYP2C19 and CYP2D6. The second class includes CYP1A1, CYP1A2, CYP2E1 and CYP3A4 which are found to be involved in the metabolism of procarcinogens. Significant effort has been spent in characterization of the mechanism of activation of procarcinogens and toxicants by P450 enzymes.[27,28]

3.1.2 Cytochrome P450 Induction and Carcinogenicity

Cytochrome P450 enzymes are either expressed constitutively in fixed amounts or induced by certain substrates. Induction is usually a protective mechanism and helps in detoxification, but can also lead to an increase in production of carcinogenic, mutagenic and/or cytotoxic metabolites.[29] Several clinical studies have shown significantly increased or decreased levels of certain P450s in tumor tissue as compared to normal tissue suggesting a relationship between CYP induction and tumor development.

Polycyclic aromatic hydrocarbons (PAHs) are known carcinogens, which are distributed everywhere in the environment.[30] PAHs are usually metabolized by CYP1A1 and CYP1B1 enzymes. Many studies have demonstrated that CYP1As are highly inducible by carcinogenic (PAHs).[31] Such feedback cycle enables the PAHs to induce their own metabolism into carcinogenic forms. CYP1B1 has been found to be expressed at abnormally high levels in tumors (122 out of 127) under investigation.[32,33] It is the most expressed form of CYP1 family in breast cancer tissue. CYP1B1 is hypothesized to be involved in tumor growth and progression. CYP1B1 bears ~40% homology with both CYP1A1 and CYP1A2 enzymes. CYP3A enzymes play an important role in catalyzing the metabolism of different drugs, carcinogens and endogenous substances.

Variation in expression of different P450 enzymes leads to significant changes in carcinogenic response. Notable agreement has been seen between the Ames test for genotoxicity and ENACT enzyme induction assay; and they seem to align with the potential carcinogenicity of test chemicals. Induction of CYP enzymes has been hypothesized to be associated with potential toxicity and tumor occurrences at certain sites.[34,35] The observation of such prominent induction of P450 enzymes by the PAHs and their increased expression in tumor tissue raises concerns for the safety of humans and animals in general. The impact of these studies led to profound influence on the drug development, cancer research, and toxicology. Pharmaceutical companies employ a general policy in the drug development process to discontinue drug development if the drug shows CYP1 inducibility, for fear of possible toxic or carcinogenic effects.[36] P450 enzymes that are involved in procarcinogen activation and metabolism are reasonably well conserved in their expression among different species. Therefore, P450 enzyme induction is an important system for analyzing the interrelations between induction of drug metabolism and chemical toxicity in general. In this study, we investigate the role of simultaneous induction of three P450 enzymes in identification of carcinogens.

3.1.3 Methods

In Vitro Assay Data: Cellzdirect enzyme induction data for CYP1A1, CYP1A2 and CYP3A4 was obtained from the phase I of U.S. EPA's ToxCast database.[13,37] CellzDirect assay reports fold-change in expression (above basal levels) of the enzymes in an in vitro test after exposure to chemicals for 6, 24, and 48 h. The data set consists of 320 chemicals across the three enzymes. We selected the chemicals which had fold-change data for all three enzymes for 6hr (dataset 1) and 24 h (dataset 2) time points. The filtering reduced the number of chemicals to 17 in dataset 1 and 16 in dataset 2.

Carcinogenicity Data: The experimental in vivo carcinogenicity data for test chemicals is obtained from publicly available carcinogenic potency database (CPDB)[38] and chemical carcinogenesis research information system (CCRIS).[39] The distribution of carcinogens to non-carcinogens is 4:13 for dataset 1 and 8:8 for dataset 2.

Chemical Diversity: Diversity of the chemical dataset is an important measure for model validation and robustness. Diversity of chemicals in the two datasets was evaluated by the AP Tanimoto coefficient. Tanimoto coefficient ranges between 0 and 1, where 0 indicates completely dissimilar and 1 indicates completely similar. Chemicals with a Tanimoto coefficient of 0.7 and greater are considered biologically similar molecules.[40] Figure 1 shows the distribution of chemicals with respect to each other indicating structurally diverse nature of chemicals in both the datasets.

Machine Learning Algorithm: Support Vector Machines (SVM), a supervised machine learning algorithm, is used in this study for classification and regression analysis. It is a linear binary classifier which calculates an optimal hyper plane for categorizing data, which consist of pairs of values $(x_i:y_i) : i=1,..,n$ where x_i is the data point with k features ($f_j : j=1,\ldots,k$) and y_i is the corresponding class label. A hyper plane separates all data points of one class from those of the other class and is used to classify any new data points.[41,42] SVM models are especially suited for this problem because they were originally designed for training data with small size and binary classifiers.

Svmtrain,[43] a Matlab SVM implementation is used for this analysis. The svmtrain function is used with default parameters and the linear kernel function. Fold-change in expression of CYP450 enzymes is used as features in model classification and the actual experimental value is used as the class label. A new chemical with enzyme induction data can be classified using the svmclassify function based on the hyper plane generated using the training data set as explained in Section 3.1.4.

Model Validation: We perform external model validation using leave one out cross validation (LOOCV). *N* SVM models were developed each with (*N*–1) chemicals as the training set and *1* chemical as the test set. The following standard metrics were then calculated for the performance assessment of the model:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

where, *TP* is true positives, *TN* is true negatives, *FP* is false positives and *FN* is false negatives reported in the tests. Accuracy or concordance is a measure of correctness of overall predictions. Sensitivity is a measure of correctness in prediction of positives or carcinogenic chemicals and specificity is a measure of correctness in prediction of nega-

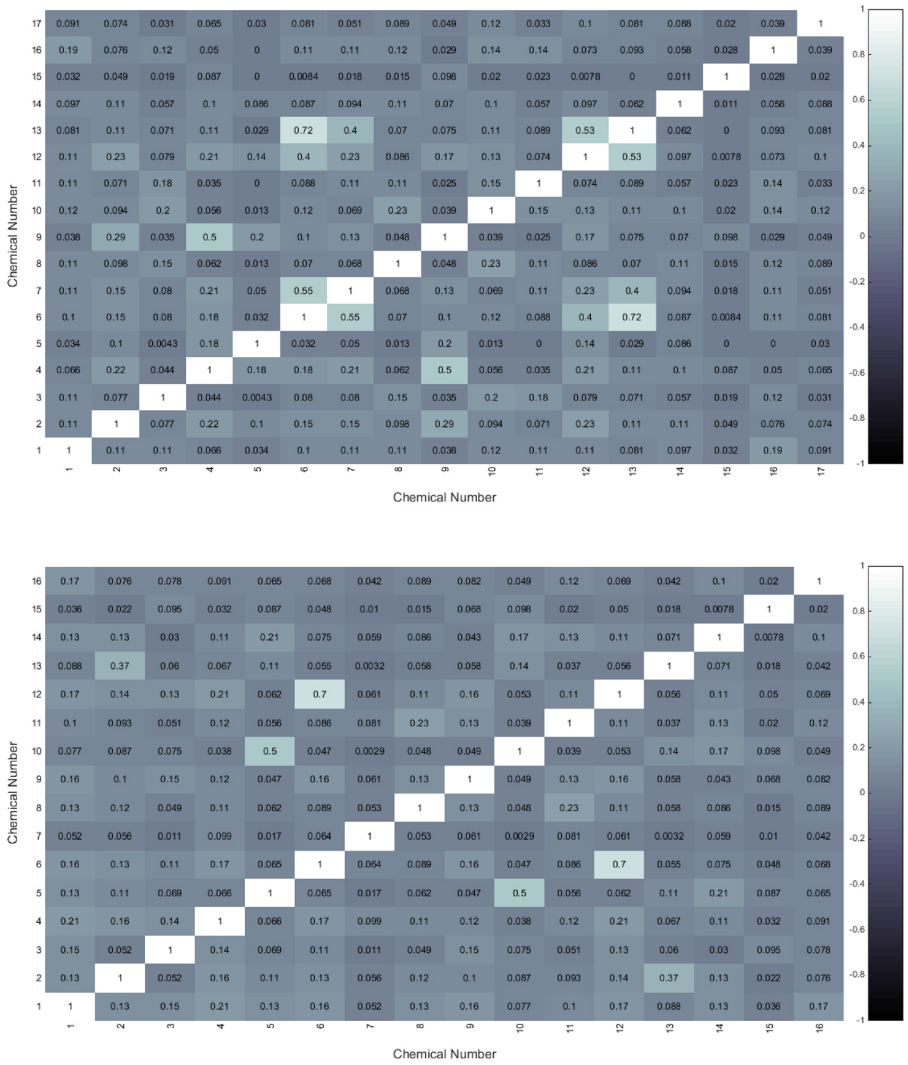

Figure 1. HeatMap representation of the chemical diversity of the two datasets measured in terms of Tanimoto distance. The annotations in each cell correspond to the distance between the two chemicals (numbers). The colorbar on the right shows mapping of the distance (range: 0 – 1) to a gray colorscale.

tives or non-carcinogenic chemicals. We also calculate the Receiver Operating Characteristics (ROC) which is a plot of true positive rate (sensitivity) versus false positive rate (1-specificity). The ROC plot demonstrates how the performance of a binary classifier changes as the threshold parameters are varied.[44]

Performance Comparison with In Silico Tools: The performance of the SVM classifier is compared with three standard in silico QSAR tools viz., Toxtree (expert knowledge-based),[45] OECD Toolbox[46] (statistical) and Vega[47] (hybrid). The tools make a binary prediction about carcinogenic potential of the test chemicals.

3.1.4 Results

The SVM separates the two classes (carcinogens and non-carcinogens) by generating a hyper plane for each training dataset in the LOOCV analysis. A new test chemical is evaluated based on the fold-change in the expression of CYP1A1, CYP1A2 and CYP3A4 and classified as carcinogenic or non-carcinogenic depending upon its distance from the separating hyperplane. Statistical performance of the SVM classifier in comparison to the various in silico tools is summarized in Table 1. As shown, the accuracy is improved to as high greater than 80% for both the datasets. Sensitivity and specificity are also greatly improved as compared to the in silico tools. The results are more relevant for dataset 2 which is more balanced with an equal distribution of positives and negatives.

Figure 2 shows the receiver operating characteristics of the SVM classifier with reference to the in silico tools. An ideal binary predictor would have zero false predictions and so the desired point on the ROC plot is the top left corner where sensitivity is one and (1-specificity) is zero.

Table 1. Performance metrics for the in silico QSAR tools and the in vitro data based SVM Classifier.

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Dataset 1 ($n=17$)			
Toxtree	64.7	**100.0**	53.9
Vega	82.4	50.0	**92.3**
OECD Toolbox	52.9	**100.0**	38.5
SVM Classifier	**88.2**	75.0	**92.3**
Dataset 2 (n=16)			
Toxtree	56.3	50.0	62.5
Vega	43.8	12.5	**75.0**
OECD Toolbox	62.5	62.5	50.0
SVM Classifier	**81.3**	**87.5**	**75.0**

The red line corresponds to the performance of a random classifier which does not have any preferences in a binary outcomes. Since the predictions are binary in nature, each classifier is represented as a point on the ROC plot. The closer the prediction is to the ideal point, the greater is the predictive ability of the classifier. As seen, SVM classifier offers better balance between sensitivity and specificity and out performs the in silico tools for both the datasets.

3.1.5 Discussion

The SVM classification QBAR model suggests a strong correlation between carcinogenic potential and the ability of test chemicals to simultaneously induce the transcription of CYP1A1, CYP1A2 and CYP3A4 enzymes. The ROC curve demonstrates a better trade-off between sensitivity and specificity in SVM classification versus in silico tools used. SVM classification also has better performance metrics than in silico QSAR tools demonstrating the advantage of using biological data as descriptors for predictive modeling.

Figure 1 shows how structurally diverse the chemical datasets are with reference to Tanimoto similarity index. It is interesting to observe that even with such a diverse dataset there is an apparent correlation between chemical carcinogenicity and the ability to simultaneously induce the three enzymes. This demonstrates that even without structural similarity we can successfully predict a toxic response based on biological similarity. This observation validates the concept behind QBAR modeling. The findings illustrate that xenobiotic induced cytochrome P450 expression (in vitro data) can be successfully used as a descriptor in QBAR modeling for carcinogenicity prediction.

3.2 QBAR Model of In Vitro Genotoxicity Assays for Carcinogenicity Prediction

3.2.1 Carcinogenicity, Mutagenicity and In Vitro Genotoxicity Assays

Carcinogenic chemicals can be broadly categorized as genotoxic and non-genotoxic based on their mechanism of action. Genotoxic carcinogens exert their carcinogenic ability by direct damage or alteration of the DNA. Mutagenic toxicity is the ability of a physical or chemical agent to cause mutations by damage to the DNA.[48,49] Owing to the correlations between mutagenicity and carcinogenicity, mutagenic toxicity is widely used as an indicator of possible carcinogenicity. Short term in vitro mutagenicity tests are, therefore, widely used to assess genotoxic carcinogenicity.[50]

Experimentally, mutagenicity is routinely assessed by the Ames test, which is an in vitro bacterial reverse mutation assay to test genotoxicity.[51] The Ames test is a benchmark method for mutagenicity testing by virtue of its well established standard protocol and acceptance within the regulatory agencies. Over the past decades, several other bacterial mutagenicity tests have been developed which are now being used worldwide because of their concordance with the Ames test. In vitro genotoxicity assays are particularly gaining importance because they: (i) present themselves as a short term and an effective alternative to long term in vivo rodent cancer studies, (ii) offer an insight into the mechanism behind genotoxic mode of action of chemicals, and (iii) can be used in the quantification of risk associated with genotoxic chemicals.[52,53]

Unlike genotoxic carcinogens, there is no clear understanding of the mechanism of action of non-genotoxic carcinogens. Carcinogenesis by non-genotoxic carcinogens can occur due to chronic cell injury, immunosuppression, increased secretion of trophic hormones, receptor activation, or CYP450 induction (also inferred in Section 3.1).[48,54] Given the complex nature of non-genotoxic carcinogenicity, the results of in vitro genotoxicity assays are not sufficient and could well be over-conservative and mechanistically unjustifiable. For instance, negative result in the Ames test cannot necessarily be translated into a negative result for carcinogenicity, which leads to increased false negative predictions. The National Toxicology Program (NTP) conducted a study on the ability of the Ames test to predict carcinogenicity and reported good accuracy but low sensitivity (~45%). The Ames test is also reported to have ~85% reproducibility rate and ~70% concordance with structural alerts for carcinogenicity.[55]

In general, in vitro genotoxicity assays are reported to have low sensitivity for prediction of carcinogenicity. The use of genotoxicity testing strategy for carcinogenicity prediction, thus, comes with a caveat of misleading false positive and false negative predictions. The latter case of false negatives is especially important under REACH regulations for regulatory acceptance of computational toxicology models.[10] The performance of different assays varies quite widely and, therefore, no single test should be considered as a gold standard for carcinogenicity prediction. A stepwise approach using a battery of in vitro genotoxicity assays should be performed to overcome the weaknesses of a single test.[56–58] We propose that this protocol be adjusted to mathematically combine the results of different

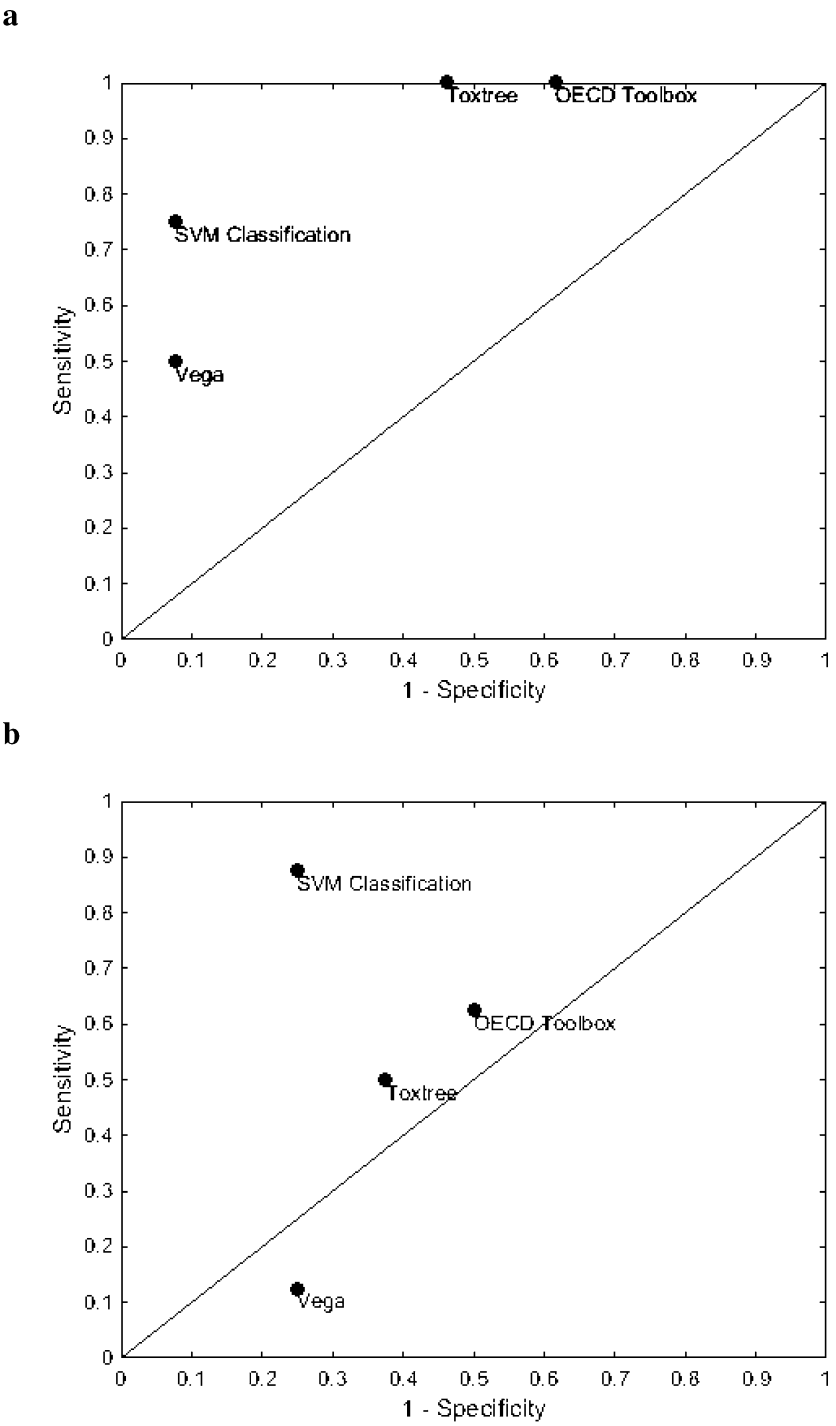

Figure 2. ROC plot of in vitro data based SVM classification, Toxtree, Vega and OECD Toolbox. (a). Dataset 1 ($n = 17$); (b). Dataset 2 ($n = 16$).

genotoxicity assays to arrive at a final prediction. Such a combination is expected to improve the sensitivity and overall concordance while still preserving the mechanistic insight from each of the in vitro assays. In this project, we use in vitro genotoxicity assay data as biological descriptors for carcinogenicity prediction as a proof-of-concept for development of proposed QBAR models.

3.2.2 Methods

In Vitro Genotoxicity Assay Data: The European Centre for the Validation of Alternative Methods (ECVAM), released a list of 22 genotoxic and 42 non-genotoxic chemicals for the evaluation of the ability of various in vitro tests to predict rodent carcinogenicity. We identified the results of 9

high throughput in vitro genotoxicity assays (Ames, micronucleus, H2AXISV, Vitotox, Radarscreen, RAD51, Cystatin, p53, Nrf2)[59–61] in open literature for the ECVAM set to develop a QBAR model for carcinogenicity prediction.

Carcinogenicity Data: The experimental in vivo carcinogenicity data for test chemicals is obtained from publicly available carcinogenic potency database (CPDB)[38] and chemical carcinogenesis research information system (CCRIS).[39] Chemicals with both chemical in vivo carcinogenicity and in vitro assay data were finally selected for classification analysis. This filtering led to a total of 56 chemicals in the dataset. The distribution of carcinogens to non-carcinogens in the dataset is 31:25.

Machine Learning Algorithm: Random Forests,[62] a machine-learning classification algorithm that produces an ensemble of unpruned decision trees for classification is used in this study. Each tree is developed by (i) selecting a bootstrap sample from the training data with replacement, (ii) randomly selecting the best descriptor variables at each node and growing the tree, and then (iii) estimating the classification error by testing the tree on the remaining data. The new data is classified based on the majority prediction of all the trees in the ensemble. The implementation is relatively simple since only two parameters need to be specified: the number of trees in the forest and the number of predictor variables at each node. The number of trees is generally proportional to the number of predictor variables, so that each predictor is likely enough to be selected. The number of predictor variables is generally defaulted to square root of the total number of variables.[63–65]

The RF algorithm is especially suited for this problem because: (i) the algorithm can assess the importance of the different predictor variables (in vitro assays) and select them accordingly at different decision nodes incorporating multiple modes of action, (ii) it does an internal performance assessment on the left out training data, thus, strengthening the analysis, and (iii) it is robust against over fitting. In general, the error rate (strength) of a RF depends upon the correlation between the trees and the strength of individual trees. Higher correlation leads to increased error rates and higher strength of each tree leads to decreased error rates.[66,67] We used the RF implementation, Treebagger,[68] in Matlab for our analysis. The Treebagger algorithm uses bagging to develop an ensemble of decision trees for classification. There is no recommended threshold for the number of trees and usually the number is varied to observe any performance changes. Based on different articles on using RFs, we varied the number of trees between 5 and 500 and used default values for other parameters.

3.2.3 Results

We perform external model validation using leave one out cross validation (LOOCV) technique and evaluate the metrics defined in Section 3.1.3. Table 2 summarizes the correlation analysis of in vitro genotoxicity assays to rodent carcinogenicity tests. The benchmark Ames assay has a sensitivity of only ~49% whereas the H2AXIS assay has the highest overall accuracy or concordance of ~70%. In general, all the genotoxicity assays have high specificity but low sensitivity (<52%) for the given ECVAM dataset. The corresponding statistics for RF classification results are summarized in Table 3. Similar to reports in a study[69] increasing the number of trees did not lead to improved prediction accuracy. The best classification metrics are obtained at generating only 5 trees. RF classification with 5 trees improves the sensitivity to ~61%.

Table 2. Performance metrics of genotoxicity assays.

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ames	67.9	45.2	**96.0**
MN	64.3	41.9	92.0
H2AXISV	**69.6**	51.6	92.0
Vitotox	64.3	41.9	92.0
Radarscreen	62.5	45.2	84.0
RAD51	60.7	35.5	92.0
Cystatin A	66.1	41.9	**96.0**
P53	66.1	48.4	88.0
Nrf2	62.5	**54.8**	72.0

Table 3. Performance metrics of the in vitro data based RF classifier (QBAR model) with varying number of trees.

Number of Trees	Accuracy (%)	Sensitivity (%)	Specificity (%)
5	67.9	61.3	76.0
10	62.5	51.6	76.0
20	62.5	58.1	68.0
30	58.9	54.8	64.0
40	58.9	54.8	64.0
50	64.3	58.1	72.0
100	58.9	54.8	64.0
110	58.9	54.8	64.0
120	62.5	58.1	68.0
130	58.9	54.8	64.0
140	62.5	54.8	72.0
150	62.5	54.8	72.0
200	62.5	54.8	72.0
300	58.9	51.6	68.0
400	60.7	54.8	68.0
500	62.5	54.8	72.0
600	64.3	58.1	72.0
700	64.3	58.1	72.0

Figure 3 shows the receiver operating characteristics of the RF classifiers with reference to the in vitro assays. The red line corresponds to the performance of a random classifier that does not have any preferences in a binary outcomes. As seen, RF classifiers have higher sensitivity as compared to the genotoxicity assays and show improved rate of false negatives.

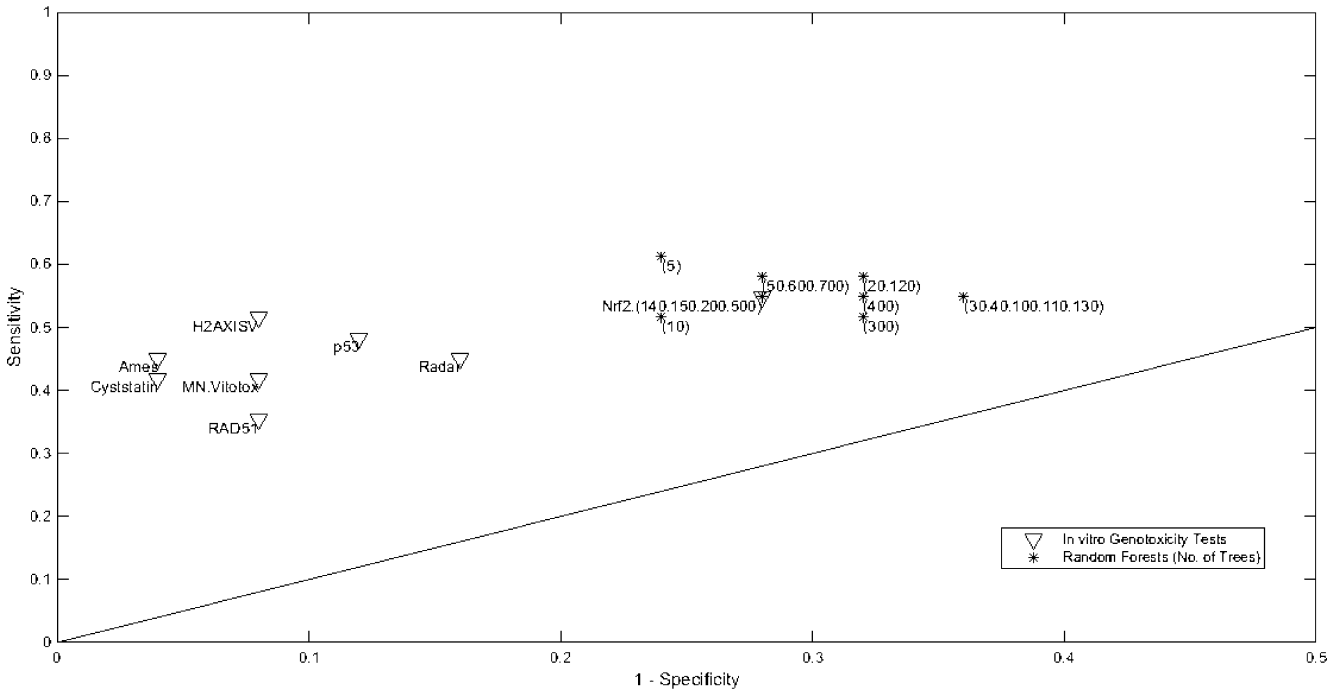

Figure 3. ROC plot of in vitro data based RF classification and in vitro genotoxicity assays.

3.2.4 Discussion

The results of the example case study demonstrate that RF classification addresses the issue of low sensitivity of in vitro genotoxicity assays as discussed in Section 3.2.1. High sensitivity is especially important under REACH requirements for regulatory purposes i.e., to protect environment and human health. Gain in sensitivity happens at the expense of specificity or higher rate of false positives which also affects the overall accuracy. It is important for a classifier to have high sensitivity in order to reduce the number of false negatives. RF classification does not result in any improvement in overall accuracy but it still maintains the accuracy of the best in vitro assay with the additional benefit of lower number of false positives. In terms of genotoxicity assays, false negatives most likely include non-genotoxic carcinogens.[50] Thus, improved sensitivity is probably an indication of higher rate of identification of non-genotoxic carcinogens using genotoxicity assays.

The results of the RF classification also illustrate that: (i) the threshold parameter in the model (number of trees in the random forest) can be changed to adjust the desired trade-off between false positives and false negatives. However, if any in vitro assay were to be used independently, there is no reference or protocol to change the threshold for each new chemical, and (ii) the choice of number of trees in RF implementation creates only minor variation in the classifier performance which demonstrates the robustness and consistency in performance of RF algorithm for developing classification models. The results demonstrate how RF classification results based on combination of in vitro genotoxicity assays can improve the identification of true carcinogens. Further analysis can also be done to identify the most important assays to assist in the design and selection of an in vitro battery of genotoxicity tests for improved carcinogenicity prediction.

4 Conclusions

The availability of high throughput in vitro assay data offers a unique opportunity of deriving knowledge about a chemical's mechanism of toxic action. Mechanistically relevant in vitro assays can be used as a powerful tool for identification of biomarkers of chemical toxicity and uncover novel biochemical pathways underlying complex toxic endpoints. We proposed the use of specific in vitro assays data in identification of relevant biological descriptors and development of QBAR models for carcinogenicity prediction. The main objective of the approach described in this paper is to demonstrate a strategy for development of quantitative biological activity relationships with carcinogenicity as an example endpoint.

We have presented two case studies supported by theory to highlight similarities between QBAR and QSAR modeling techniques. Case studies in Section 3.1 and 3.2 demonstrate an analogy between QSAR and QBAR modeling in: (i) the selection of relevant descriptors to be used in different machine learning algorithm, and (ii) the development of a computational model which maps chemical/biological descriptors to a toxic endpoint, respectively. Both the case studies show increased sensitivity or lower rates of

false negatives, which is desirable for regulatory purposes and are corroborated with theoretical knowledge to address the OECD concerns as well. Our case studies demonstrate that in vitro data can be sufficiently used to develop QBAR models for carcinogenicity prediction. Such mechanism based models can be used along with QSAR models for mechanistically complex toxic endpoints to successfully advance the development of toxicology and risk assessment studies.

Acknowledgements

This project was completed under a volunteer status at the Center for Devices and Radiological Health (CDRH) at the U.S. Food and Drug Administration. Sincere thanks to Ronald Brown (Toxicologist, CDRH) for providing the datasets used in the case studies, discussions and constructive suggestions in the development of this work.

References

[1] T. W. Schultz, M. T. Cronin, T. I. Netzeva, *J. Mol. Struct.: THEOCHEM* **2003**, *622*, 23 – 38.
[2] T. W. Schultz, M. T. Cronin, J. D. Walker, A. O. Aptula, *J. Mol. Struct.: THEOCHEM* **2003**, *622*,1 – 22.
[3] C. Hansch, D. Hoekman, A. Leo, L. Zhang, P. Li, *Toxicol. Lett.* **1995**, *79(1)*, 45 – 53.
[4] Y. J. Tseng, A. J. Hopfinger, E. X. Esposito, *J. Comput.-Aided Mol. Des.* **2012**, *26*, 39 – 43.
[5] Thomas Ferrari, Giuseppina Gini, *Chem. Cent. J.* **2009**, *4(Suppl 1)*, S2.
[6] R. Serafimova, M. F. Gatnik, A. Worth, *Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity*, Publications Office, **2010**.
[7] J. Jaworska, N. Nikolova-Jeliazkova, and T. Aldenberg, *Review of Methods for QSAR Applicability Domain Estimation by The Training Set*, Technical Report, The European Commission – Joint ResearchCentre Institute for Health & Consumer Protection - ECVAM, **2005**.
[8] *OECD principles for the validation of (Q)SARs*, https://eurl-ecvam.jrc.ec.europa.eu/laboratoriesresearch/predictive_toxicology/background/oecd-principles, last accessed, October **2013**.
[9] A. P. Worth, *Recent Advances in QSAR Studies*, Springer, Heidelberg, **2010**, 367 – 382.
[10] E. Benfenati, *Practical Aspects of Computational Chemistry*, Springer, Heidelberg, **2010**, 183 – 200
[11] G. Eisenbrand, B. Pool-Zobel, V. Baker, M. Balls, B. J. Blaauboer, A. Boobis, A. Carere, S. Kevekordes, J.-C. Lhuguenot, R. Pieters, J. Kleiner, *Food Chem. Toxicol.* **2002**, *40(2)*, 193 – 236.
[12] Hao Zhu, *Computational Toxicology*, Humana Press, New York, **2013**, 53 – 65.
[13] D. J. Dix, K. A. Houck, M. T. Martin, A. M. Richard, R. W. Setzer, R. J. Kavlock, *Toxicol. Sci.* **2007**, *95(1)*, 5 – 12.
[14] F. S. Collins, G. M. Gray, J. R. Bucher, *Science* **2008**, *319(5865)*, 906 – 907.
[15] R. R. Tice, C. P. Austin, R. J. Kavlock, J. R. Bucher, *Environ. Health Perspect.* **2013**, *121(7)*, 756 – 765.
[16] U. Bernauer, A. Oberemm, S. Madle, U. Gundert-Remy, *Basic Clin. Pharmacol. Toxicol.* **2005**, *96(3)*, 176 – 181.
[17] Keith A. Houck, Robert J. Kavlock, *Toxicol. Appl. Pharmacol.* **2008**, *227(2)*, 163 – 178.
[18] I. Rusyn, A. Sedykh, Y. Low, K. Z. Guyton, A. Tropsha, *Toxicol. Sci.* **2012**, *127(1)*, 1 – 9.
[19] R. S. Thomas, M. Black, L. Li, E. Healy, T. M. Chu, W. Bao, M. Andersen, R. Wolfinger, *Toxicol. Sci.* **2012**, *128(2)*, 398 – 417.
[20] A. Sedykh, H. Zhu, H. Tang, L. Zhang, A. Richard, I. Rusyn, A. Tropsha, *Environ. Health Perspect.* **2011**, *119(3)*, 364 – 370.
[21] Bas J Blaauboer, *Environ. Toxicol. Pharmacol.* **2002**, *11(3)*, 213 – 225.
[22] H. Zhu, I. Rusyn, A. Richard, A. Tropsha, *Environ. Health Perspect.* **2008**, *116(4)*, 506 – 513.
[23] A. M. Richard, R. Benigni, *SAR QSAR Environ. Res.* **2002**, *13(1)*, 1 – 19.
[24] S. Rendic, F. P. Guengerich, *Chem. Res. Toxicol.* **2012**, *25(7)*, 1316 – 1383.
[25] F. P. Guengerich, *AAPS J.* **2006**, *8(1)*, E101 – E111.
[26] C. J. Omiecinski, R. P. Remmel, V. P. Hosagrahara, *Toxicol. Sci.* **1999**, *48(2)*, 151 – 156.
[27] I. Johansson, M. Ingelman-Sundberg, *Toxicol. Sci.* **2011**, *120(1)*, 1 – 13.
[28] F. P. Guengerich, *Chem. Res. Toxicol.* **2007**, *21(1)*, 70 – 83.
[29] M. S. Denison, J. P. Whitlock, *J. Biol. Chem.* **1995**, *270(31)*, 18175 – 18178.
[30] D. W. Nebert, D. W. Russell, *The Lancet* **2002**, *360(9340)*, 1155 – 1162.
[31] T. Shimada, Y. Fujii-Kuriyama, *Cancer Sci.* **2004**, *95(1)*, 1 – 6.
[32] F. J. Gonzalez, S. Kimura, *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2001**, *477(1)*, 79 – 87.
[33] G. I. Murray, M. C. Taylor, M. C. McFadyen, J. A. McKay, W. F. Greenlee, M. D. Burke, W. T. Melvin, *Cancer Res.* **1997**, *57(14)*, 3026 – 3031.
[34] David F. Lewis, Costas Ioannides, Dennis V. Parke, *Environ. Health Perspect.* **1998**, *106(10)*, 633–641.
[35] F. P. Guengerich, T. Shimada, *Chem. Res. Toxicol.* **1991**, *4(4)*, 391 – 407.
[36] Qiang Ma, Anthony YH Lu, *Drug Metab. Dispos.* **2007**, *35(7)*, 1009 – 1016.
[37] *Toxcast database*, www.actor.epa.gov, last accessed: September **2013**.
[38] Lois S. Gold, *The Carcinogenic Potency Project*, **1984**, last accessed: October **2013**.
[39] *Chemical Carcinogenesis Research Information System*, National Library of Medicine, **2001**, last accessed: March **2014**.
[40] P. Willett, J. M. Barnard, G. M. Downs, *J. Chem. Inf. Model.* **1998**, *38(6)*, 983 – 996.
[41] M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning*, The MIT Press, Cambridge, MA, **2012**.
[42] C. J. Burges, *Data Mining and Knowledge Discovery*, **1998**, *2(2)*, 121 – 167.
[43] Train Support Vector Machine Classifier, *MATLAB and Statistics Toolbox Release 2012b*, The MathWorks, Inc., last accessed: October **2013**.
[44] K. H. Zou, A. J. O'Malley, L. Mauri, *Circulation* **2007**, *115(5)*, 654 – 657.
[45] *Toxtree* 2.5.0, last accessed: 2013.
[46] OECD QSAR Toolbox website, last accessed: April **2013**.
[47] N. Fjodorova, M. Vračko, M. Novič, A. Roncaglioni, E. Benfenati, *Chem. Central J.***2010**, *4(Suppl 1)*, S3.
[48] B. Silva Lima, V. der Laan, J. Willem, *Regul. Toxicol. Pharmacol.* **2000**, *32(2)*, 135 – 143.

[49] J. H. M. Van Delft, E. Van Agen, S. G. J. Van Breda, M. H. Herwijnen, Y. C. M. Staal, J. C. S. Kleinjans, *Carcinogenesis* **2004**, *25(7)*, 1265 – 1276.
[50] D. Clive, *Regul. Toxicol. Pharmacol.* **1985**, *5(1)*, 79 – 100.
[51] B. N. Ames, J. McCann, E. Yamasaki, *Mutat. Res.* **1975**, *31*, 347 – 364.
[52] S. Pfuhler, M. Fellows, J. van Benthem, R. Corvi, R. Curren, K. Dearfield, P. Fowler, R. Frötschl, A. Elhajouji, L. Le Hégarat, T. Kasamatsu, H. Kojima, G. Ouédraogo, A. Scott, G. Speit, *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2011**, *723(2)*, 101 – 107.
[53] R. Benigni, C. Bossa, *Mutagenesis* **2011**, *26(3)*, 455 – 460.
[54] J. C. Barrett, *Environ. Health Perspect.* **1993**, *100*, 9 – 20.
[55] R. Benignia, A. Giulianib *J. Toxicol. Environ. Health* **1988**, *25*, 135 – 148.
[56] L Müller, *Hum. Exp. Toxicol.* **2009**, *28(2–3)*, 131 – 133.
[57] R. M. Walmsley, N. Billinton, *Br. J. Pharmacol.* **2011**, *162(6)*, 1250 – 1258.
[58] F. Madia, P. Phrakonkham, R. Corvi, *In Vitro Toxicology Systems*, Springer, New York, **2014**, pp. 307 – 332.
[59] W. Westerink, J. C. Stevenson, G. J. Horbach, W. G. Schoonen, *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2012**, *696(1)*, 21 – 40.
[60] L. Le Hégarat, A. Mourot, S. Huet, L. Vasseur, S. Camus, C. Chesné, V. Fessard, *Toxicol. Sci.* **2014**, *138(2)*, 300 – 309.
[61] W. Westerink, J. C. Stevenson, A. Lauwers, G. Griffioen, G. J. Horbach, W. G. Schoonen, *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2009**, *676(1)*, 113 – 130.
[62] L. Breiman, *Machine Learning* **2001**, *45(1)*, 5 – 32.
[63] W. Tong, H. Hong, H. Fang, Q. Xie, R. Perkins, *J. Chem. Inf. Model.* ***2003***, *43(2)*, 525 – 531.
[64] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, B. P. Feuston, *J. Chem. Inf. Model.* **2003**, *43(6)*, 1947 – 1958.
[65] A. L. Boulesteix, S. Janitza, J. Kruppa, I. R. König, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2012**, *2(6)*, 493 – 507.
[66] A. Statnikov, L. Wang, C. F. Aliferis, *BMC Bioinf.* **2008**, *9(1)*, 319.
[67] M. Liu, M. Wang, J. Wang, D. Li, *Sens. Actuators, B* **2013**, *177*, 970 – 980.
[68] Treebagger, *MATLAB and Statistics Toolbox Release 2012b, The* MathWorks, Inc., last accessed: October **2014**.
[69] T. M. Oshiro, P. S. Perez, J. A. Baranauskas, *Machine Learning and Data Mining in Pattern Recognition*, Springer, Heidelberg, **2012**, 154–168.

Received: November 16, 2014
Accepted: February 24, 2015
Published online: March 25, 2015

