JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models

Todd M. Martin,*[,†] Christopher M. Grulke,[‡] Douglas M. Young,[†] Christine L. Russom,[§] Nina Y. Wang,[⊥] Crystal R. Jackson,[∥] and Mace G. Barron[∥]

[†]National Risk Management Research Laboratory, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268, United States

[‡]National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711, United States

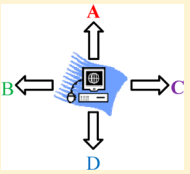
[§]National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, 6201 Congdon Boulevard, Duluth, Minnesota 55804, United States

[⊥]National Center for Environmental Assessment, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268, United States

[∥]National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, 1 Sabine Island Drive, Gulf Breeze, Florida 32561, United States

Supporting Information

ABSTRACT: The ability to determine the mode of action (MOA) for a diverse group of chemicals is a critical part of ecological risk assessment and chemical regulation. However, existing MOA assignment approaches in ecotoxicology have been limited to a relatively few MOAs, have high uncertainty, or rely on professional judgment. In this study, machine based learning algorithms (linear discriminant analysis and random forest) were used to develop models for assigning aquatic toxicity MOA. These methods were selected since they have been shown to be able to correlate diverse data sets and provide an indication of the most important descriptors. A data set of MOA assignments for 924 chemicals was developed using a combination of high confidence assignments, international consensus classifications, ASTER (ASessment Tools for the Evaluation of Risk) predictions, and weight of evidence professional judgment based an assessment of structure and literature information. The overall data set was randomly divided into a training set (75%) and a validation set (25%) and then used to develop linear discriminant analysis (LDA) and random forest (RF) MOA assignment models. The LDA and RF models had high internal concordance and specificity and were able to produce overall prediction accuracies ranging from 84.5 to 87.7% for the validation set. These results demonstrate that computational chemistry approaches can be used to determine the acute toxicity MOAs across a large range of structures and mechanisms.



1. INTRODUCTION

A critical component in the vision of 21st century toxicology is a reduction in animal testing and replacement, in part, by mechanistically based and statistically robust extrapolation models.[1] To meet this need will require a significant paradigm shift from risk assessments based on whole organism studies of well-defined apical end points to ones reliant on existing databases, in vitro testing, targeted testing, and quantitative/qualitative modeling approaches. Quantitative Structure−Activity Relationship (QSAR) models are used in risk assessments of new industrial organic chemicals within the US EPA.[2] There are three approaches for developing quantitative QSAR models for diverse data sets: (1) develop local models based on a common mode of toxic action, (2) develop local models based on sets of clearly defined congeners, and (3) develop global models using flexible machine learning techniques. Several researchers have indicated that the most toxicologically meaningful results are obtained using the first approach.[3−7]

To develop local models based on a common mode of toxic action, one must first be able to assign the mode of action for the chemicals in the overall set. There are two basic approaches for assigning mode of action: fragment rule-based and molecular descriptor-based.[8] In fragment rule-based methods, experts determine which fragments are indicative of each MOA based on prior experience. In molecular descriptor-based methods, a variety of molecular descriptors are utilized to build QSAR models such as linear discriminant analysis models.[9] Many different descriptors are used to build QSAR models, but the most common descriptors are ones that characterize partitioning (such as the octanol−water partition coefficient) or descriptors which characterize reactivity (such as the energy of the highest occupied molecular orbital, HOMO).

One advantage of fragment-based approaches is they yield a list of chemical substructures, which are easier to interpret in terms of a mechanistic basis. One disadvantage of fragment-

Received: May 6, 2013

Table 1. Data Set Breakdown by MOA Categories

MOA	MOA subcategory	example chemicals	characteristics	training (#)	validation (#)
AChE inhibition	carbamate	aldicarb, carbaryl, methiocarb	carbamate moiety	37	12
	organo phosphate	chlorpyrifos, malathion, terbufos	phosphate or thiophosphate moiety	112	38
nAChR agonism	*NA*	acetamiprid, dinotefuran, nicotine	nicotine or neconicotinoids	6	2
narcosis	ester	bromoxynil butyrate, methyl acetate, phenyl salicylate	diverse structures, mono or diesters	34	11
	nonpolar	benzene, ethanol, metolachlor, phenanthrene	diverse structures, not ionizable	230	76
	polar	4-nonylphenol, pyridine, triclosan	phenols, anilines, pyridines; single nitro or <4 halogens	42	14
neurotoxicity	organo chlorine	aldrin, DDT, fipronil, lindane	halogenated phenyls and cyclodienes	31	10
	pyrethroid	cypermethrin, ethofenprox, fenvalerate	pyrethroid analogs	62	21
reactivity	*NA*	acrolein, benzaldehyde, phenyl disulfide	electrophiles, proelectrophiles	67	23
uncoupling oxidative phosphorylation	*NA*	2,4-dinitrophenol, binapacryl, pentachloro phenol	>1 nitro or >3 halogenated phenols, anilines, pyridines	32	12
not ACHE inhibition MOA	carbamate	molinate, terbucarb	carbamate moiety	32	10
	organo phosphate	ethephon, glyphosate	phosphate or thiophosphate moiety	8	2

based methods is they can yield contradictory predictions if a chemical possesses fragments characteristic of two different MOAs. In addition fragment-based rule systems are limited because they reduce a chemical to a specified substructure and ignore other topological and electronic features of the entire compound that may influence its propensity to act under a given MOA.[10] Another disadvantage of fragment-based rule systems is that the absence of certain fragments does not necessarily indicate that the chemical does not have the given MOA. Finally, it may be difficult to assess chemicals which contain a structural fragment not contained in the training set.[9]

Descriptor-based methods rely on numerical values of descriptors that are less likely to be influenced by user bias.[8] However, some descriptors may be difficult to calculate (for example quantum descriptors) or hard to interpret in terms of mechanism of action.[8] Also, some descriptors are essentially molecular fragment counts (such as the number of hydroxyl groups attached to aromatic carbons), thus some molecular descriptor-based QSAR models can be very similar to fragment rule-based models.

In recent years several researchers have developed molecular descriptor-based QSAR methodologies to assign categories such as antibacterial activity,[11] skin sensitization,[12] aquatic toxicity mode of action,[8,9,13−20] and genotoxicity.[21] Several researchers have modeled the mode of action for a set of 221 phenols to *T. pyriformis*.[9,13,14,16,17] A variety of techniques were utilized including linear discriminant analysis (LDA),[9] binary logistic regression,[13] Adaboost,[14] decision tree models,[16] counter propagation neural networks and multinomial logistic regression,[22] and radial basis function neural networks and support vector machines (SVM).[17] The prediction accuracy for the validation data sets ranged from 84 to 93% for the different QSAR methods. Researchers[19,20] have also modeled a fathead minnow aquatic toxicity mode of action data set with more than 600 chemicals developed by Russom et al.[10,23] The researchers utilized pairwise support vector machines,[19] k nearest neighbors (k-NN), learning vector quantitization (LVQ) neural networks, and LDA.[20] The MOA prediction accuracy was approximately 75% across the different QSAR methods.

The goal of this study was to develop simple, fast, and high certainty chemical descriptor-based methods to assign the MOA for acute toxicity, defined as the general structural-based mode of acute toxicity of waterborne chemicals to fish. Linear discriminant analysis (LDA) and random forest (RF) modeling approaches were used to develop models to assign acute toxicity MOA. LDA was selected because it generates a series of multilinear regression models which clearly show which descriptors are associated with each MOA. RF was selected because it has been shown they have the ability to correlate large data sets using a large number of input variables (in this case molecular descriptors).[24] A data set of 924 chemicals was developed using a combination of high confidence assignments, international consensus classifications, ASTER (ASessment Tools for the Evaluation of Risk)[25] predictions, and weight of evidence professional judgment based on an assessment of structure and literature information. LDA and RF models were fit to the training set (75% of the overall set) and validated with the remaining chemicals (25% of the overall set). In the future, these models may be utilized in a two-tiered approach for toxicity estimation. In the first tier, the mode of action will be determined. In the second tier, the toxicity will be quantitatively estimated using the QSAR model for the assigned MOA.

2. MATERIALS AND METHODS

2.1. Mode of Action Data Set. A data set of 924 chemicals was developed for six MOAs based on their mode of acute toxic action in fish (see Table 1). The data set was developed using a combination of high confidence assignments, including biological responses in fish acute toxicity assays,[10] international consensus classifications (e.g., Insecticide Resistance Action Committee's (IRAC) MOA classification tool[26]), ASTER (ASessment Tools for the Evaluation of Risk)[25] predictions, and weight of evidence professional judgment based on assessment of structure and literature information (see Table S1 in the Supporting Information). For example, the MOA database developed for this study (Table S1) included 188 high confidence (level A or B) MOA assignments from an existing fathead minnow (*Pimphales promelas*) acute toxicity and MOA database of primarily industrial organic chemicals.[10] MOAs in the fathead minnow database had been assigned using a weight of evidence approach, based on information gleaned from acute lethality studies with fish, joint toxicity of chemical mixture

dx.doi.org/10.1021/ci400267h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

studies, and respiratory-cardiovascular responses studies as well as supporting evidence from the literature.[10] The chemical space was further expanded to include underrepresented chemicals, including organophosphates, carbamates, and chlorinated phenols and anilines. MOA was assigned for the pesticide actives using information from several sources including the IRAC MOA classification tool[26] and the Herbicide Resistance Action Committee (HRAC 2012).[27] MOA was confirmed using secondary sources and an analysis of the structure rules via the ASTER tool.[28] The final data set was compiled using only those structures that had a high level of confidence in the MOA based on sources of MOA information and best professional judgment by the authors. 63% of the chemicals had either highly confident sources or consensus among moderate to highly confident sources. The remaining 37% relied on a single determination. Chemicals with lower confidence or conflicting MOA assignments were excluded from the final data set used in the analyses. If the mode of action for a chemical could not be assigned via experimental data or by the best professional judgment of the authors it was omitted.

The molecular structures for the chemicals in the overall data set were obtained using the procedure detailed by Young et al.[29] Chemicals were removed from the overall data set if they contained elements other than carbon, hydrogen, oxygen, nitrogen, fluorine, chlorine, bromine, iodine, sulfur, phosphorus, silicon, or arsenic. In addition, chemicals were removed if they were salts, polymers, or mixtures.

One of six broad MOAs was assigned to each chemical: acetylcholinesterase (AChE) inhibition, nicotinic acetylcholine receptor (nAChR) agonism, narcosis, neurotoxicity, reactivity, and uncoupling oxidative phosphorylation (see Table 1). MOA assignments included a subcategory based on chemical structure for AChE inhibition (carbamates, organophosphates), narcosis (ester, nonpolar, polar), and neurotoxicity (organochlorine, pyrethroid). The first column in Table 1 indicates "broad" MOAs, while the first and second columns taken together indicate "specific" MOAs. Models were developed for both broad and specific MOAs. The MOA for reactive chemicals was inclusive of a number proelectrophile and electrophile mechanisms, including acylation, isocyanate reactivity, nucleophilic substitution, alkylation/arylation, and sulfhydryl reactivity, but these specific reaction classes were not used in model development.

The MOA data set were randomly subdivided into training (75% of chemicals) and validation set (25%;) within each MOA and MOA subcategory (see Table S1 in the Supporting Information). Prior to dividing into sets, narcotic chemicals within each specific MOA subcategory were further categorized as either simple or complex and aliphatic or aromatic to ensure a balance of structures in both the training and validations sets. Similarly, organochlorine neurotoxicants were categorized as either phenyl or cyclodiene prior to assignment to training and validation sets.

2.2. Multiclass Modeling Approaches. There are three general ways to develop QSAR models for multiclass data sets: 1) apply a machine learning technique that can differentiate multiple categories, 2) develop a series of models to distinguish the differences between each pair of classes, or 3) generate a series of models that separate the members of each one of the classes from the rest. LDA is intended for binary modeling and cannot differentiate multiple classes in a single learning, so one of the latter two modeling approaches was required.

In a pairwise modeling approach, regression models are developed for all possible pairs of results categories. The predicted results category (or in this case MOA) is the category which is selected the most often. In "one against the rest" LDA, separate discriminant models are developed for each results category. The predicted category is the category which yields the highest score from all the different models. "One against the rest" modeling was chosen for this paper because we wished to have a method that does not always assign a MOA (for example if the highest score does not exceed a certain cutoff value). This may occur for chemicals that are very dissimilar to those in the training set. Also, having a single model associated with each MOA allows a straightforward interpretation of the resultant models. The "one against the rest" approach was used with both LDA and RF modeling

The training set included several carbamate and organophosphate chemicals that were structurally similar to AChE inhibitors but were known to not to have an AChE inhibition MOA. These chemicals were not inhibitors since they did not have the ability to bind via electrostatic interactions to the AChE receptor's anionic center and by hydrogen bonding to the esteratic site of the enzyme.[30] These chemicals were assigned a score of 0. These chemicals were included to facilitate the training of discriminant models to not predict all carbamates and organophosphates to be AChE inhibitors. Similar chemicals were also added to the validation set to assess whether the models would incorrectly assign them an AChE inhibition MOA.

2.2.1. Linear Discriminant Analysis (LDA). An LDA model is essentially a multilinear regression model[13]

$$Score = \sum_{i=1}^{n} a_i x_i + a_0 \tag{1}$$

where a_i is the fitted constant for parameter i, x_i is the ith molecular descriptor, and a_0 is the model intercept. Chemicals with a given MOA are assigned a score of 1. Chemicals that do not have the given MOA are assigned a score of 0. Descriptors are selected (and regression constants fitted) so that the model can match the experimental values for the training set. Typically, if the predicted score for a given MOA model is greater than or equal to 0.5, then that MOA is indicated to be positive. If the score is less than 0.5, then that MOA is not indicated.

To build the LDA model for each MOA, a genetic algorithm technique was used to select the optimal descriptor set. This was done to select descriptors in an unbiased fashion (for example the octanol−water partition coefficient is not automatically selected). In building our QSAR models, we investigated using different ratios of chemicals to descriptors. It was determined that using a minimum ratio of 80 chemicals per descriptor (about eight descriptors per model) yielded the best results in terms of resubstitution sensitivity (the fraction of positive chemicals correlated accurately to the number of positive chemicals in the training set). Using additional descriptors per model did not improve the internal model statistics. The 80:1 ratio far exceeds the minimum recommended ratio of 5:1.[31]

The genetic algorithm used in this study was taken from the Weka statistical package, version 3.5.1.[32,33] The objective function that is maximized is given by[34]

$$objective = \exp(q^2_{adj,LMO}) \tag{2}$$

dx.doi.org/10.1021/ci400267h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

where $q^2_{adj,LMO}$ is the adjusted 5-fold leave many out cross-validation coefficient

$$q^2_{adj,LMO} = 1 - \left[\frac{\sum_{i=1}^{n_k} (\hat{y}_i - y_{exp,i})^2/(n - p - 1)}{\sum_{i=1}^{n_k} (y_{exp,i} - \bar{y}_{exp})^2/(n - 1)} \right] \quad (3)$$

where $\hat{y}_i$ and $y_{exp,i}$ are the predicted and experimental toxicity values for chemical i, $\bar{y}_{exp}$ is the average score for the chemicals in the training set, n is the number of chemicals in the training set, and p is the number of parameters in the model. The 5-fold q^2 was used instead of the traditional q^2 LOO (leave one out) inside the genetic algorithm because it yields a significant degree of computational savings for large training sets. The n_k-p-1 term penalizes models that include extra parameters that do not significantly increase the predictive power of the model. For the "nAChR agonists" end point, only 6 descriptors were selected by the genetic algorithm (as opposed to 8 which is the maximum number for an 80:1 ratio). Traditionally Wilk's lambda is used to assess the fitness of discriminant models.[12,21] In this study, it was determined that similar training statistics could be obtained if either $q^2_{adj,LMO}$ or Wilk's lambda was used for the objective function (not shown).

To build the LDA models, the molecular descriptors from T.E.S.T.[35] were utilized. The descriptors consisted of 790 descriptors in the following classes: E-state values and E-state counts, constitutional descriptors, topological descriptors, walk and path counts, connectivity, information content, 2D autocorrelation, Burden eigenvalue, molecular property (such as the octanol−water partition coefficient), kappa, hydrogen bond acceptor/donor counts, molecular distance edge, and molecular fragment counts.[36]

Before any QSAR model can be used to make a prediction for a test chemical, it must be determined whether the test chemical falls within the domain of applicability (or AD) for the model. The first constraint, the model ellipsoid constraint, checks if the test chemical is within the multidimensional ellipsoid defined by the ranges of descriptor values for the chemicals in the training set (for the descriptors appearing in the model). The model ellipsoid constraint is satisfied if the leverage of the test compound (h_{00}) is less than the maximum leverage value for all the compounds used in the model.[37] The second constraint, the Rmax constraint, checks if the distance from the test chemical to the centroid of the training set is less than the maximum distance for any chemical in the training set to the its centroid. The final constraint is the training set must contain at least one example of each of the molecular fragments that are present in the test chemical.[34]

A filter was employed in which the maximum score from all the LDA models must exceed 0.5 (denoted from here on as the min score filter). The prediction accuracy for the compounds that are excluded by this criterion (and by the AD discussed above) was assessed to determine if they improve the overall prediction accuracy (by omitting mostly incorrect predictions).

2.2.2. Random Forest (RF) Analysis. RF is a technique developed by Breiman and Cutler[24] that builds a series of decision trees utilizing subsets of compounds and available descriptors. The developed models are then used as an ensemble predictor. Each decision tree is formulated to only predict whether a chemical does or does not belong to a particular MOA. The final score for a compound from RF models for each MOA is simply the percentage of trees contained within the forest that predicts the chemical to belong to that MOA. If more than half of the trees (i.e., a score of 0.5 or 50%) predict a particular MOA, it is considered a positive indication.

Each decision tree within the forest is optimized only on a subset of compounds from the data set. To select the tree nodes, an algorithm identifies which descriptor (and threshold) from a randomly selected subset of descriptors will result in the lowest Gini impurity. This measure of the mingling between classes is given by

$$I_G = 1 - \Sigma_i f_i^2 \quad (4)$$

where f_i is the fraction of members belonging to class i in the set of items. The algorithm selecting criteria is greedy, always selecting the split point that provides the purest branches.

While the modeling in this paper was inspired by the RF technique developed by Breiman, modeling was carried out through use of the Chembench Web portal.[38,39] The RF procedure in Chembench, while quite similar to the traditional application of random forest, is technically only "RF-like" in that it varies the way that modeling set selection is done. Rather than a new training set being selected for each new tree grown, a manageable number of internal training sets are defined and then multiple trees (a grove) grown for each of these sets. Additionally, these sets are selected without replacement. The generation of grooves is done using the "randomForest" package for R available from http://cran.r-project.org/web/packages/randomForest/index.html. "RF-like" models were generated using 50 training/test set splits with 50 trees generated per split. Each tree was optimized using a subset of 25 chemical descriptors selected from the 900 hydrogen-depleted dragon descriptors generated. The Dragon software[40] is the underlying engine used to calculate all 2D Dragon descriptors within ChemBench. These included topological descriptors, constitutional descriptors, walk and path counts, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, Burden eigenvalues, topological charge indices, eigenvalue-based indices, functional group counts, atom-centered fragments, and molecular properties.

As noted above there are limitations in the extrapolation of a model prediction beyond the realm of its knowledge base (training set). While multiple restrictions on the AD can be made using a variety of paradigms, since the goal is to limit extrapolation beyond the information contained in the training set, perhaps the simplest and most straightforward approach is requiring chemicals meet a lower limit of similarity to compounds in the training set. In this case, global tanimoto similarity of the MACCS keys[41] (as calculated using Molecular Operation Environment[42]) for each predicted compound to its nearest neighbor in the training set was required to be greater than 0.6. A minimum score filter similar to that used for the LDA models was also applied.

3. RESULTS AND DISCUSSION

3.1. Results. *3.1.1. Internal Model Statistics.* The internal model statistics were defined as follows

$$Concordance = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

dx.doi.org/10.1021/ci400267h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

All of the LDA and RF models had high correlations (84−100%) in terms of concordance and specificity (see Table 2).

Table 2. Modeling Statistics for Training Set

	LDA (resubstitution)			RF (5-fold CV)		
specific MOA	conc[a]	sens	spec	conc	sens	spec
AChE inhibition-carbamates	0.99	1.00	0.99	0.97	0.46	1.00
AChE inhibition-organo phosphates	0.98	0.97	0.98	0.99	0.97	0.99
nAChR agonism	1.00	0.50	1.00	0.99	0.00	1.00
narcosis-esters	0.99	0.97	0.99	0.96	0.23	1.00
narcosis-nonpolar	0.84	0.92	0.80	0.84	0.72	0.91
narcosis-polar	0.96	0.50	0.99	0.96	0.43	0.99
neurotoxicity-organo chlorine	0.98	0.68	1.00	0.99	0.71	1.00
neurotoxicity-pyrethroid	0.99	0.92	0.99	1.00	0.98	1.00
reactivity	0.94	0.45	0.99	0.93	0.39	0.99
uncoupling oxidative phosphorylation	0.98	0.66	1.00	0.98	0.69	1.00
	LDA (resubstitution)			RF (5-fold CV)		
broad MOA	conc	sens	spec	conc	sens	spec
AChE inhibition	0.97	1.00	0.96	0.96	0.86	0.99
nAChR agonism	1.00	0.50	1.00	0.99	0.00	1.00
narcosis	0.87	0.94	0.81	0.89	0.91	0.87
neurotoxicity	0.98	0.84	1.00	0.98	0.89	0.99
reactivity	0.94	0.45	0.99	0.93	0.39	0.99
uncoupling oxidative phosphorylation	0.98	0.66	1.00	0.98	0.69	1.00

[a]Abbreviations: conc = concordance, sens = sensitivity, spec = specificity.

However, for several of the specific MOAs (reactivity, nAChR agonism, and narcosis-polar) we were unable to achieve high sensitivity values with either LDA (based on resubstitution) or RF models (based on 5-fold cross-validation). Even utilizing additional descriptors in the LDA models did not improve the sensitivity for these MOAs. On the other hand, some of the problems with the RF models may be related to overfitting. Resubstitution statistics indicate perfect prediction for all training set compounds; however, the poor sensitivity during cross-validation indicates that specific rules are likely being identified to target individual compounds. This appears to be more likely when the number of "active" compounds is low for a particular MOA modeling effort.

Other researchers have reported that 3D or quantum descriptors can improve models for predicting MOAs such as reactivity.[43,44] Quantum descriptors were not investigated because the goal is to develop computationally efficient models that can quickly assign mode of action without relying on proprietary molecular descriptor generation software. The reactivity MOA contained a variety of structural classes (and thus different reactive mechanisms are at play), which may explain why it was difficult to consistently flag reactive compounds. Though the reactivity models included fragments that are often associated with increased chemical reactivity, such as the number of carbon atoms that are double bonded (SdCH2_acnt), the number of carbon atoms that are triple bonded (#C [aliphatic attach]), the number of nitrogens attached to other nitrogens (−NH− [nitrogen attach])), the number of aldehyde groups (−CHO [aliphatic attach] and −CHO [aromatic attach]), and the number of acrylate groups (CH2═CHC(═O)O), it seems difficult to capture the breadth of biochemical reaction fingerprints within a single binary model.

The inability to effectively model nAChR agonism is almost certainly related to the small number of representative chemicals in the training set. The high ratio of inactives to actives (100:1) likely inhibited the ability of the machine learning algorithms to identify an appropriate solution. The difficulty of modeling imbalanced, skewed, or biased binary data sets has been well documented.[45−48]

Finally, the narcosis-polar models may have failed due to ambiguity about the nature of difference between the narcosis-polar and narcosis-nonpolar categories. For example, it has been demonstrated that ortho substitutions on a pyridine can shift the MOA from polar narcosis (e.g., pyridine) to nonpolar narcosis (2-ethylpyridine), presumably due to steric hindrance.[10]

3.1.2. Validation Set Results. The prediction results for the validation set (with no AD or min score filter constraints applied) are given in Table 3. The overall prediction accuracy for the different QSAR techniques was excellent (85 and 87% for the specific and broad MOAs, respectively). The prediction accuracy is defined as the fraction of chemicals for which the assigned MOA (the MOA for the model with the maximum score) matched the experimental MOA.

Table 3. Validation Set Results for Each MOA[a]

		percent correct	
specific MOA	count	LDA (TEST)	RF (Dragon)
AChE inhibition-carbamates	12	100.0	91.7
ACHE inhibition-organophosphates	38	97.4	100.0
nAChR agonism	2	50.0	50.0
narcosis-esters	11	90.9	72.7
narcosis-nonpolar	76	92.1	93.4
narcosis-polar	14	64.3	57.1
neurotoxicity-organochlorine	10	70.0	90.0
neurotoxicity-pyrethroid	21	95.2	100.0
reactivity	23	47.8	43.5
uncoupling oxidative phosphorylation	12	66.7	75.0
total	219	84.5	84.9
Cohen's kappa		0.805	0.810
		percent correct	
broad MOA	count	LDA (TEST)	RF (Dragon)
AChE inhibition	50	100.0	94.0
nAChR agonism	2	50.0	50.0
narcosis	101	96.0	95.0
neurotoxicity	31	83.9	96.8
reactivity	23	47.8	43.5
uncoupling oxidative phosphorylation	12	50.0	66.7
total	219	87.2	87.7
Cohen's kappa		0.811	0.818

[a]All chemicals are predicted; no AD or min score filter was applied.

These results are slightly better than the 73.3% prediction accuracy (using 5-fold cross-validation) reported by Michielan et al.[19] for a smaller fathead minnow data set. The results were

dx.doi.org/10.1021/ci400267h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

comparable to the 2-fold prediction results achieved for a 221 chemical *T. pyriformis* data set (85−89%,[9] 86−91%,[13] 93%,[14] 72−93%[16]). For the LDA and RF methods, the Cohen's kappa values were about 0.81 which indicates almost perfect performance.[15]

The total concordance values were slightly higher for the broad MOAs than for the specific MOAs. However, the kappa values were nearly identical for each QSAR method, which indicates that the performance was essentially equivalent. This is somewhat surprising since it should be easier to get better results with four fewer possible results categories.

The prediction accuracies for each MOA were highly correlated with the training set sensitivities. For example for the LDA method, the AChE inhibition-carbamates were predicted with perfect accuracy (the training set sensitivity was 100%), while the reactive chemicals were predicted with 47.8% accuracy (the training set sensitivity was 45%). Similar trends were observed for the random forest method. This similarity is not unexpected since the overall data set was divided into training and prediction sets randomly with equal distributions of each mode of action.

3.1.3. Model Descriptors. The LDA models for each end point are given in the Supporting Information (Tables S2 and S3). The most common descriptors selected include molecular fragment counts,[36] autocorrelation,[49] molecular distance edge,[50] Burden eigenvalue,[51] and walk and path count descriptors.[52] The descriptors included in RF models are primarily fragment counts, autocorrelation descriptors, and Burden eigenvalues (not shown). The definitions for the descriptors appearing in the LDA models are given in the T.E.S.T. user's guide,[53] while those associated with the RF models can be obtained from the Dragon user's guide.[40] The molecular fragment count descriptors are given by Martin et al.[34] Some of the fragments were initially developed so that they captured specific modes of action as detailed by Russom et al.[10]

One benefit of LDA models is they are inherently interpretable. It was found that the fragment counts selected in the LDA models are similar to those in the Verhaar scheme[54] or ASTER.[25] For example, the model for AChE inhibition-organophosphate included the count of sulfur atoms with 2 single bonds (SssS_acnt), and the count of phosphorus atoms double bonded to sulfur (P═S). For the narcosis-nonpolar MOA, the genetic algorithm selected a series of molecular fragments and Qsv, which gives a measure of the polarity in a molecule.[36] The model coefficients for the fragment descriptors were negative which indicates that nonpolar narcosis MOA is less likely to be assigned if a given fragment is present. The coefficient for Qsv was positive which indicates that the nonpolar narcosis MOA is less likely if the polarity increases (since Qsv decreases with increasing polarity). Similar understanding can be drawn from the other linear models that were developed.

Understanding the results of RF models is not so simple. The generation of multiple trees leads to model complexity that is not easy to disentangle. A single tree is simple to understand, but the forest is nearly indecipherable. Commonly, rather than try to understand the model itself, RF modelers look at the descriptor that occurred most frequently within the forest and then how the values of that descriptor were distributed among the actives and inactives within the data set. From that they draw conclusions regarding the effect of the descriptor on the activity of a chemical. For example, the most frequently selected descriptors in the AChE inhibition-organophosphate RF model were the counts of phosphorus atoms (nP), phosphate atoms forming 3 single bonds and 1 double bond (P-117), the number of oxygens bonded to two aliphatics (O-059). Hence, the descriptors that were captured segregate the organophosphates in the data set from the remaining compounds.

3.2. Discussion. *3.2.1. Effect of Prediction Filters.* In this study, we investigated whether model constraints (or filters) improve the prediction accuracy or just decrease the prediction coverage (the fraction of chemicals which can be predicted). In general, it was found that the min score filter removed predictions which have a lower accuracy (see Table 4) than the overall accuracy reported in Table 3. For example, for the LDA method, the accuracy of the omitted chemicals (for broad MOAs) was 2/9 (22%) which is considerably lower than the accuracy without filters (87.2%). For specific MOAs, the accuracy of the omitted chemicals for the LDA method was 13/22 (59%). Similarly with RF, the accuracies for compounds omitted due to this filter were 7/15 (46.7%) and 25/41 (61%) for the broad and specific MOA models, respectively. While the addition of the min score filter decreases the prediction coverage at a slightly greater percentage than it does increase the prediction accuracy, we recommend including this constraint because it is illogical to place confidence in a prediction with a score less than 0.5 (since typically scores less than 0.5 indicate that a chemical does not possess a given MOA).

Table 4. Effect of Filters on Validation Results

Specific MOAs		
	fraction of omitted that have correct predictions	
method (descriptors)	max score	AD
LDA (TEST)	13/22	8/9 0/1[a]
RF (Dragon)	25/41	19/31
Broad MOAs		
	fraction of omitted that have correct predictions	
method (descriptors)	max score	AD
LDA (TEST)	2/9	8/9 0/1[a]
RF (Dragon)	7/15	21/31

[a]Fragment constraint removed from AD filter.

For the LDA method, it was found that the descriptor-based AD removed predictions that were very accurate (8/9 correct), which is not desirable. In the T.E.S.T. software, the fragment constraint was not applied to binary end points such as mutagenicity and developmental toxicity since it was not found to improve prediction concordance.[53] When the fragment constraint is omitted from the AD, the correctly predicted chemicals are not excluded from being predicted. A possible explanation for this is that the LDA models already account for key molecular fragments so that including the fragment constraint just reduces prediction coverage. The one chemical that is still excluded by the applicability domain is incorrectly predicted and also violates the min score filter.

For the RF method, the nearest neighbor similarity applicability domain omitted predictions that were correct for roughly two-thirds of the chemicals which is somewhat lower than the overall prediction accuracy of 85%. As the acceptable

level of error is user dependent, it is recommended that when applying the model the user specify a similarity threshold that meets their comfort level. It is important to note that the similarity threshold applied in this case is model independent, and other model independent techniques for AD definition may elicit comparable or superior results.[55]

3.2.2. Prediction Accuracy for "not AChE" Chemicals. In order to increase the accuracy of the AChE inhibition models, we included 40 compounds that were known not to be AChE inhibitors (but had a similar structure) to the training set. These chemicals were included to determine if the models are selective or just assign AChE inhibition to any chemical with certain structural fragments (such as a P═S fragment). The exact MOA in fish was not precisely known for all of these chemicals, but the assignment of "not-AChE" was based on best professional judgment. In order for a substance to inhibit the AChE enzyme, it must block the serine hydroxyl group where acetylcholine would normally attach. The electronic and steric configuration of AChE enzyme and the pesticide as well as the leaving group on the pesticide are critical to the likelihood that a substance will inhibit to the enzyme.[30] The results for the twelve "not AChE" chemicals are given in Table 5. The MOAs assigned by ASTER[25] are also given in Table 5. With the exception of bensulide (CAS#741-58-2), all of the chemicals were predicted by ASTER to not be AChE inhibitors. Although the primary MOA for bensulide is as a fungicide, it does have moderate AChE inhibition based on measures in rat brain.[56] The model could be improved by expanding the data set to include other similar structures that do not bind to the AChE enzyme.

For LDA and RF models, the majority of the "not AChE" chemicals were predicted to have a narcosis MOA which matches the predictions from ASTER. It is not surprising that this MOA is predicted since narcosis is usually predicted in the absence of known toxophores.

3.2.3. Analysis of Incorrectly Predicted Chemicals. Several compounds were incorrectly predicted by both the LDA and RF methods. The compounds that did not violate the applicability domain of either method and were predicted incorrectly are given in Table 6 (for specific MOAs) and Table 7 (for broad MOAs). The predicted MOA for the training set example chemicals were taken from the LDA method. In many cases there were insufficient training set compounds which contained a similar structure to the test compound (i.e., nicotine, 1,3-dibromopropane, phenyl disulfide, flufenamic acid, pentachloropyridine, and isopimaric acid). For example there was only 3 disulfides in the training set so it would be difficult for a model to determine whether to predict that phenyl sulfide should be narcosis-nonpolar like morpholine, 4,4-dithiobis or 4,5-dichloro-1,2-dithio-2-one or reactivity like propyl disulfide. In addition there are no examples of brominated alkanes that are reactive as in the case of 1,3-dibromopropane.

In several cases, the models failed to account for certain structural features despite adequate examples in the training set (i.e., 1,3-diaminopropane, 1,3,5-trinitrobenzene, 2-methyl-1,4-naphthalenedione, and 2,4,5-trichlorophenol). For example, 1,3,5-trinitrobenzene was incorrectly predicted to be an uncoupler of oxidative phosphorylation even though trinitrotoluene was present in the training set. This can be attributed to the fact that sometimes machine based learning methods fail to develop hard and fast rules for certain structural features.

In some cases, training set compounds had incorrect or lower certainty MOA assignments which would cause improper training of the models. A post analysis quality assurance review of the 924 chemical revealed an error rate in MOA assignments of less than 1%. For example, n-butyl amine was assigned a MOA of narcosis-nonpolar. However, the toxicity of n-butyl amine to the Golden Orfe fish[57] was reported to be additive with phenol (a polar narcotic). Russom and co-workers had difficulty assigning MOA to aliphatic amines;[10] and others have proposed a separate narcosis MOA for amines.[8] These uncertainties in MOA assignment would have impacts to model predictions. For example if n-butylamine was assigned to be narcosis-polar, then 1,3-diaminopropane would have had a greater chance of being predicted to be polar as well.

Table 5. Predictions for "not AChE" Chemicals in the Validation Set

CAS	specific MOA	ASTER MOA	LDA (TEST)	RF (Dragon)
741-58-2	NA-OP	AI-OP	AI-OP	AI-OP
1071-83-6	NA-OP	N-N	R	N-N
1134-23-2	NA-C	N-N	N-N	N-N
1929-77-7	NA-C	N-N	N-N	N-N
2303-17-5	NA-C	N-N	N-N	N-N
3337-71-1	NA-C	N-P	AI-C	AI-C
13684-63-4	NA-C	N-N	AI-C	N-E
19622-08-3	NA-C	N-N	N-N	N-N
52888-80-9	NA-C	N-N	N-N	N-N
66952-49-6	NA-C	R	N-N	AI-C
85785-20-2	NA-C	N-N	N-N	N-N
283159-90-0	NA-C	N-E	AI-C	N-E
CAS	broad MOA	ASTER MOA	LDA (TEST)	RF (Dragon)
741-58-2	NA	AI	AI	AI
1071-83-6	NA	N	AI	N
1134-23-2	NA	N	N	N
1929-77-7	NA	N	N	N
2303-17-5	NA	N	N	N
3337-71-1	NA	N	AI	AI
13684-63-4	NA	N	AI	N
19622-08-3	NA	N	N	N
52888-80-9	NA	N	N	N
66952-49-6	NA	R	N	AI
85785-20-2	NA	N	N	N
283159-90-0	NA	N	AI	N

Legend

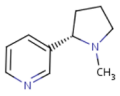
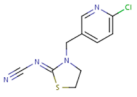
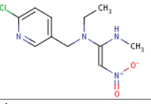
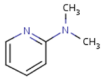
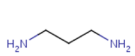
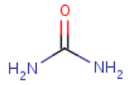
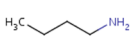
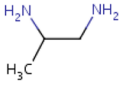
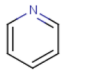
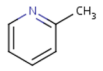
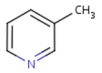
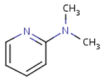
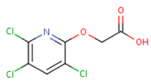
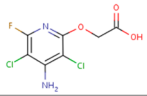
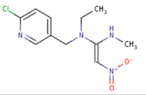
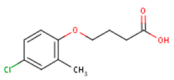
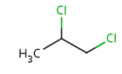
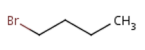
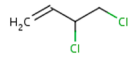
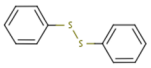
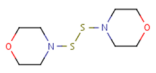
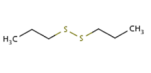
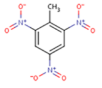
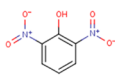
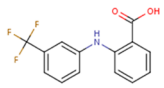
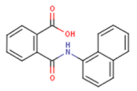
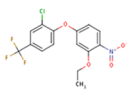
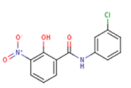
specific MOA	abbrev	broad MOA	abbrev
AChE inhibition-carbamates	AI-C	not AChE	NA
AChE inhibition-organophosphates	AI-OP	AChE inhibition	AI
not AChE-carbamates	NA-C	narcosis	N
not AChE-organophosphates	NA-OP	reactivity	R
narcosis-nonpolar	N-N		
narcosis-esters	N-E		
narcosis-polar	N-P		
reactivity	R		
neurotoxicity-organochlorine	N-O		

4. CONCLUSIONS

In this study different classifier models were derived to estimate acute toxicity mode of action. The overall prediction accuracy (85%) exceeded that obtained in the literature (73%) for a smaller fathead minnow mode of action data set. The majority of the modes of action were predicted accurately with the exception of reactive compounds and nAChR agonists. The prediction performance for specific and broad modes of action

dx.doi.org/10.1021/ci400267h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 6. Incorrectly Predicted Specific MOAs by Both the LDA and RF Methods

	Chemical	Training Example 1	Training Example 2	Training Example 3
Structure				
Name	Nicotine	Thiacloprid	nitenpyram	2-Dimethylaminopyridine
CAS	54-11-5	111988-49-9	150824-47-8	5683-33-0
Exp	nAChR agonism	nAChR agonism	nAChR agonism	Narcosis-Nonpolar
Pred	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar
Structure				
Name	1,3-Diaminopropane	Urea	Butylamine	1,2-Propanediamine
CAS	109-76-2	57-13-6	109-73-9	78-90-0
Exp	Narcosis-Polar	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Polar
Pred	Narcosis-NonPolar	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar
Structure				
Name	Pyridine	2-Methylpyridine	3-Methylpyridine	2-Ethylpyridine
CAS	110-86-1	109-06-8	108-99-6	100-71-0
Exp	Narcosis-Polar	Narcosis-Nonpolar	Narcosis-Polar	Narcosis-Nonpolar
Pred	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar
Structure				
Name	Triclopyr acid	Fluroxypyr acid	Picloram	MCPB
CAS	55335-06-3	69377-81-7	1918-02-1	94-81-5
Exp	Narcosis-Polar	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar
Pred	Narcosis-Nonpolar	Narcosis-Polar	Narcosis-Polar	Narcosis-Nonpolar
Structure				
Name	1,3-Dibromopropane	1,2-Dichloropropane	1-Bromobutane	3,4-Dichloro-1-butene
CAS	109-64-8	78-87-5	109-65-9	760-23-6
Exp	Reactivity	Narcosis-Nonpolar	Narcosis-Nonpolar	Reactivity
Pred	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar	Reactivity
Structure				
Name	Phenyl disulfide	Morpholine, 4,4-dithiobis-	Propyl disulfide	4,5-Dichloro-1,2-dithio-3-one
CAS	882-33-7	103-34-4	629-19-6	1192-52-5
Exp	Reactivity	Narcosis-Nonpolar	Reactivity	Narcosis-Nonpolar
Pred	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar
Structure				
Name	1,3,5-Trinitrobenzene	2,4,6-trinitrophenol	Trinitrotoluene	2,6-Dinitrophenol
CAS	99-35-4	88-89-1	118-96-7	573-56-8
Exp	Reactivity	Uncoupling Oxidative Phosphorylation	Reactivity	Uncoupling Oxidative Phosphorylation
Pred	Uncoupling Oxidative Phosphorylation	Uncoupling Oxidative Phosphorylation	Uncoupling Oxidative Phosphorylation	Uncoupling Oxidative Phosphorylation
Structure				
Name	Flufenamic acid	Naptalam	Oxyfluorfen	3-Chloro-3-nitrosalicylanilide
CAS	530-78-9	132-66-1	42874-03-3	6491-02-7
Exp	Uncoupling Oxidative Phosphorylation	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar
Pred	Narcosis-Nonpolar	Narcosis-Nonpolar	Narcosis-Nonpolar	Uncoupling Oxidative Phosphorylation

dx.doi.org/10.1021/ci400267h | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 6. continued

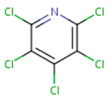
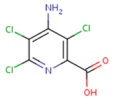
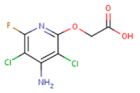
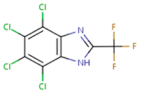
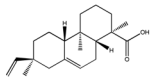
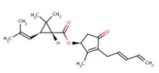
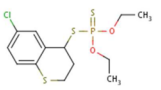
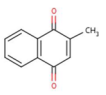
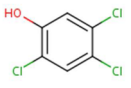
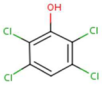
	Chemical	Training Example 1	Training Example 2	Training Example 3
Structure				
Name	Pentachloropyridine	Picloram	Fluroxypyr acid	TTFP
CAS	2176-62-7	1918-02-1	69377-81-7	2338-29-6
Exp	Uncoupling Oxidative Phosphorylation	Narcosis-Nonpolar	Narcosis-Nonpolar	Uncoupling Oxidative Phosphorylation
Pred	Narcosis-Nonpolar	Narcosis-Polar	Narcosis-Polar	Uncoupling Oxidative Phosphorylation

Table 7. Incorrectly Predicted Broad MOAs by Both the LDA and RF Methods[a]

	Chemical	Training Example 1	Training Example 2	Training Example 3
Structure				
Name	Isopimaric acid	pyrethrin I	thicrofos	dithicrofos
CAS	5835-26-7	121-21-1	41219-32-3	41219-31-2
Exp	Narcosis	Neurotoxicity	AChE inhibition	AChE inhibition
Pred	Neurotoxicity	Neurotoxicity	AChE inhibition	AChE inhibition
Structure				
Name	2-Methyl-1,4-naphthalenedione	Juglone	Acetophenone	2,4-Dichloroacetophenone
CAS	58-27-5	481-39-0	98-86-2	2234-16-4
Exp	Reactivity	Reactivity	Narcosis	Narcosis
Pred	Narcosis	Narcosis	Narcosis	Narcosis
Structure				
Name	2,4,5-trichlorophenol	2,3,4,6-Tetrachlorophenol	3,4,5-trichlorophenol	2,3,5,6-Tetrachlorophenol
CAS	95-95-4	58-90-2	609-19-8	935-95-5
Exp	Uncoupling Oxidative Phosphorylation	Uncoupling Oxidative Phosphorylation	Uncoupling Oxidative Phosphorylation	Uncoupling Oxidative Phosphorylation
Pred	Narcosis	Narcosis	Narcosis	Narcosis

[a]Excluding the chemicals appearing in Table 6.

was similar. The use of min score rule prediction filters enabled one to exclude less accurate predictions. The models were able to achieve good prediction accuracy for compounds possessing AChE inhibitor-like fragments (but are not actually AChE inhibitors). Incorrect predictions can sometimes be explained by a lack of sufficient structural analogs in the training set. Future research could include investigating specific reactivity categories and better differentiation between polar and nonpolar narcosis. These may be achieved by subdividing the training sets into smaller structural-mechanism based groupings.

■ ASSOCIATED CONTENT

Supporting Information

The mode of action data sets and the LDA models. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 513-569-7682. E-mail: martin.todd@epa.gov.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Rocky Goldsmith and Danny Chang for technical discussions and insight. The opinions expressed in this work are those of the authors and do not represent the policies or opinions of the U.S. EPA.

■ REFERENCES

(1) National Research Council, *Toxicity testing in the 21st century: A vision and a strategy. National Academy of Sciences, Washington, DC*; National Academy of Sciences: Washington, DC, 2007.

(2) Auer, C. M.; Nabholz, J. V.; Baetchke, K. P. Mode of action and the assessment of chemical hazards in the presence of limited data: Use of structure-activity relationships (SAR) under TSCA, section 5. *Environ. Health Perspect.* **1990**, *87*, 183−197.

(3) Benigni, R.; Richard, A. M. QSARS of mutagens and carcinogens: Two case studies illustrating problems in the construction of models for noncongeneric chemicals. *Mutat. Res., Genet. Toxicol.* **1996**, *371*, 29−46.

(4) Schultz, T. W.; Sinks, G. D.; Bearden, A. P. QSAR in aquatic toxicology: A mechanism of action approach comparing toxic potency to *Pimephales promelas*, *Tetrahymena pyriformis*, and *Vibrio fischeri*. In *Comparative QSAR*; Devillers, J., Ed.; Taylor & Francis: New York, 1998; pp 51−109.

(5) Abernethy, S. G.; MacKay, D.; McCarty, L. S. "Volume fraction" correlation for narcosis in aquatic organisms: The key role of partitioning. *Environ. Toxicol. Chem.* **1988**, *7*, 469−481.

(6) Cronin, M. T. D.; Schultz, T. W. Structure-toxicity relationships for phenols to Tetrahymena pyriformis. *Chemosphere* **1996**, *32*, 1453−1468.
(7) Mekenyan, O. G.; Veith, G. D. Relationships between descriptors for hydrophobicity and soft electrophilicity in predicting toxicity. *SAR QSAR Environ. Res.* **1993**, *1*, 335−344.
(8) Ren, S. Predicting three narcosis mechanisms of aquatic toxicity. *Toxicol. Lett.* **2002**, *133*, 127−139.
(9) Aptula, A. O.; Netzeva, T. I.; Valkova, I. V.; Cronin, M. T. D.; Schultz, T. W.; Kühne, R.; Schüürmann, G. Multivariate discrimination between modes of toxic action of phenols. *Quant. Struct.-Act. Relat.* **2002**, *21*, 12−22.
(10) Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). *Environ. Toxicol. Chem.* **1997**, *16*, 948−967.
(11) Mishra, R. K.; Garcia-Domenech, R.; Galvez, J. Getting discriminant functions of antibacterial activity from physicochemical and topological parameters. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 387−393.
(12) Ren, Y.; Liu, H.; Xue, C.; Yao, X.; Liu, M.; Fan, B. Classification study of skin sensitizers based on support vector machine and linear discriminant analysis. *Anal. Chim. Acta* **2006**, *572*, 272−282.
(13) Schüürmann, G.; Aptula, A. O.; Kühne, R.; Ebert, R.-U. Stepwise discrimination between four modes of toxic action of phenols in the Tetrahymena pyriformis assay. *Chem. Res. Toxicol.* **2003**, *16*, 974−987.
(14) Niu, B.; Jin, Y.; Lu, W.; Li, G. Predicting toxic action mechanisms of phenols using AdaBoost Learner. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 43−48.
(15) Spycher, S.; Nendza, M.; Gasteiger, J. Comparison of different classification methods applied to a mode of toxic action data set. *QSAR Comb. Sci.* **2004**, *23*, 779−791.
(16) Norinder, U.; Lidén, P.; Boström, H. Discrimination between modes of toxic action of phenols using rule based methods. *Mol. Diversity* **2006**, *10*, 207−212.
(17) Yao, X. J.; Panaye, A.; Doucet, J. P.; Chen, H. F.; Zhang, R. S.; Fan, B. T.; Liu, M. C.; Hu, Z. D. Comparative classification study of toxicity mechanisms using support vector machines and radial basis function neural networks. *Anal. Chim. Acta* **2005**, *535*, 259−273.
(18) Nendza, M.; Müller, M. Discriminating toxicant classes by mode of action: 2. Physico-chemical descriptors. *Quant. Struct.-Act. Relat.* **2000**, *19*, 581−598.
(19) Michielan, L.; Pireddu, L.; Floris, M.; Moro, S. Support vector machine (SVM) as alternative tool to assign acute aquatic toxicity warning labels to chemicals. *Mol. Inf.* **2010**, *29*, 51−64.
(20) Basak, S. C.; Grunwald, G. D.; Host, G. E.; Niemi, G. J.; Bradbury, S. P. A comparative study of molecular similarity, statistical, and neural methods for predicting toxic modes of action. *Environ. Toxicol. Chem.* **1998**, *17*, 1056−1064.
(21) Du, H.; Wang, J.; Watzl, J.; Zhang, X.; Hu, Z. Classification structure−activity relationship (CSAR) studies for prediction of genotoxicity of thiophene derivatives. *Toxicol. Lett.* **2008**, *177*, 10−19.
(22) Spycher, S.; Pellegrini, E.; Gasteiger, J. Use of structure descriptors to discriminate between modes of toxic action of phenols. *J. Chem. Inf. Model.* **2005**, *45*, 200−208.
(23) US EPA. DSSTOX, EPA Fathead Minnow Acute Toxicity Database File. http://www.epa.gov/ncct/dsstox/sdf_epafhm.html (accessed 11/2/2012).
(24) Breiman, L.; Cutler, A. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed 1/2/2013).
(25) US EPA. ASTER (ASsessment Tools for the Evaluation of Risk). http://www.epa.gov/med/Prods_Pubs/aster.htm (accessed 11/27/12).
(26) Insecticide Resistance Action Committee (IRAC). IRAC MoA Classification Scheme. Prepared by the IRAC International MoA Working Group. http://www.irac-online.org/eClassification/ (accessed 3/20/13).
(27) Herbicide Resistance Action Committee (HRAC). Classification of herbicides according to site of action. . http://www.hracglobal.com/Publications/ClassificationofHerbicideSiteofAction.aspx (accessed 3/20/13).
(28) Russom, C. L.; Anderson, E. B.; Greenwood, B. E.; Pilli, A. ASTER: an integration of the AQUIRE data base and the QSAR system for use in ecological risk assessments. *Sci. Total Environ.* **1991**, *109−110*, 667−670.
(29) Young, D. M.; Martin, T. M.; Venkatapathy, R.; Harten, P. Are the chemical structures in your QSAR correct? *QSAR Comb. Sci.* **2008**, *27*, 1337−1345.
(30) Fukuto, T. R. Mechanism of action of organophosphorus and carbamate insecticides. *Environ. Health Perspect.* **1990**, *87*, 245−254.
(31) Eriksson, L.; Jaworska, J. S.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361−1375.
(32) The University of Waikato. WEKA - The Waikato Environment for Knowledge Analysis. http://www.cs.waikato.ac.nz/~ml/weka/ (accessed 10/24/12).
(33) Witten, I. H. *Data Mining: Practical machine learning tools and techniques*; Morgan Kaufmann: San Francisco, 2005.
(34) Martin, T. M.; Harten, P.; Venkatapathy, R.; Das, S.; Young, D. M. A hierarchical clustering methodology for the estimation of toxicity. *Toxicol. Mech. Methods* **2008**, *18*, 251−266.
(35) US EPA. T.E.S.T. Version 4.1. http://www.epa.gov/nrmrl/std/qsar/qsar.html#TEST (accessed 9/26/12).
(36) US EPA. Molecular Descriptors Guide. http://www.epa.gov/nrmrl/std/qsar/MolecularDescriptorsGuide-v102.pdf (accessed 9/26/12).
(37) Montgomery, D. C. In *Introduction to linear regression analysis*; John Wiley and Sons: New York, 1982; pp 141−143.
(38) Walker, T.; Grulke, C. M.; Pozefsky, D.; Tropsha, A. Chembench: a cheminformatics workbench. *Bioinformatics* **2010**, *26*, 3000−3001.
(39) UNC MML. ChemBench Web Portal. http://chembench.mml.unc.edu (accessed 3/19/2003).
(40) Talete SRL. DRAGON for Windows and Linux. http://www.talete.mi.it/help/dragon_help/ (accessed 9/18/2010).
(41) Symyx MACCS Structural Keys; MDL Information Systems Inc.: San Ramon, CA, 2005.
(42) Chemical Computing Group Inc. Molecular Operating Environment (MOE); 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.
(43) Mekenyan, O. G.; Veith, G. D. The electronic factor in QSAR: MO-parameters, competing interactions, reactivity and toxicity. *SAR QSAR Environ. Res.* **1994**, *2*, 129−143.
(44) Schwoöbel, J. A. H.; Koleva, Y. K.; Enoch, S. J.; Bajot, F.; Hewitt, M.; Madden, J. C.; Roberts, D. W.; Schultz, T. W.; Cronin, M. T. D. Measurement and estimation of electrophilic reactivity for predictive toxicology. *Chem. Rev. (Washington, DC, U. S.)* **2011**, *111*, 2562-2596.
(45) Kubat, M.; Holte, R.; Matwin, S. Learning when negative examples abound. In *Machine Learning: ECML-97*; Someren, M., Widmer, G., Eds.; Springer: Berlin, Heidelberg, 1997; Vol. *1224*, pp 146−153.
(46) Dacheng, T.; Xiaoou, T.; Xuelong, L.; Xindong, W. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *Pattern Analysis Machine Intelligence, IEEE Trans.* **2006**, *28*, 1088−1099.
(47) Pugazhenthi, D.; Rajagopalan, S. P. Unbalance quantitative structure activity relationship problem reduction in drug design. *J. Comput. Sci.* **2009**, *5*, 764−772.
(48) Kondratovich, E.; Baskin, I. I.; Varnek, A. Transductive support vector machines: Promising approach to model small and unbalanced datasets. *Mol. Inf.* **2013**, *32*, 261−266.
(49) Todeschini, R.; Consonni, V. In *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000; p 18.
(50) Liu, S.; Cao, C.; Li, Z. Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular

distance-edge (MDE) vector, λ. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 387−394.
(51) Todeschini, R.; Consonni, V. In *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000; p 132.
(52) Todeschini, R.; Consonni, V. In *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000; p 384.
(53) US EPA. TEST User's Guide. http://www.epa.gov/nrmrl/std/qsar/TEST-user-guide-v41.pdf (accessed 11/27/12).
(54) Verhaar, H. J. M.; van Leeuwen, C. J.; Hermens, J. L. M. Classifying environmental pollutants. *Chemosphere* **1992**, *25*, 471−491.
(55) Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR applicabilty domain estimation by projection of the training set descriptor space: a review. *ATLA, Altern. Lab. Anim.* **2005**, *33*, 445−459.
(56) US EPA Organophosphorus Cumulative Risk Assessment − 2006 Update; U.S. Environmental Protection Agency, Office of Pesticide Programs. Docket ID: EPA-HQ-OPP-2006-0618-0002; 2006; p 522.
(57) Kamlet, M. J.; Taft, R. W.; Abraham, M. H.; Veith, G. D.; Abraham, D. J. Solubility properties in polymers and biological media. 8. An analysis of the factors that influence toxicities of organic nonelectrolytes to the Golden Orfe fish (Leuciscus idus melanotus). *Environ. Sci. Technol.* **1987**, *21*, 149−155.