

A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors

John F. Carriger¹, Todd M. Martin², Mace G. Barron^{1*}

U.S. Environmental Protection Agency, Office of Research and Development

¹Gulf Ecology Division, Gulf Breeze, FL 32561

²Sustainable Technology Division, Cincinnati, OH 45220

*Corresponding author: barron.mace@epa.gov

ABSTRACT

The mode of toxic action (MoA) has been recognized as a key determinant of chemical toxicity, but development of predictive MoA classification models in aquatic toxicology has been limited. We developed a Bayesian network model to classify aquatic toxicity MoA using a recently published dataset containing over one thousand chemicals with MoA assignments for aquatic animal toxicity. Two dimensional theoretical chemical descriptors were generated for each chemical using the Toxicity Estimation Software Tool. The model was developed through augmented Markov blanket discovery from the dataset of 1098 chemicals with the MoA broad classifications as a target node. From cross validation, the overall precision for the model was 80.2%. The best precision was for the AChEI MoA (93.5%) where 257 chemicals out of 275 were correctly classified. Model precision was poorest for the reactivity MoA (48.5%) where 48 out of 99 reactive chemicals were correctly classified. Narcosis represented the largest class within the MoA dataset and had a precision and reliability of 80.0%, reflecting the global precision across all of the MoAs. False negatives for narcosis most often fell into electron transport inhibition, neurotoxicity or reactivity MoAs. False negatives for all other MoAs were most often narcosis. A probabilistic sensitivity analysis was undertaken for each MoA to examine the sensitivity to individual and multiple descriptor findings. The results show that the Markov blanket of a structurally complex dataset can simplify analysis and interpretation by identifying a subset of the key chemical descriptors associated with broad aquatic toxicity MoAs, and by providing a computational chemistry-based network classification model with reasonable prediction accuracy.

Key words: mode of action; aquatic toxicity; Bayesian network; Markov blanket; chemical descriptors

INTRODUCTION

The mode of toxic action (MoA) has been recognized as a key determinant of chemical toxicity but MoA classification in aquatic toxicology has been generally limited to either narrow domains of applicability, highly complex or simplistic assignment schemes, or poor prediction accuracy (Russom et al. 1997; Nendza and Muller 2007; Enoch et al. 2008; Barron et al. 2015). Generally two approaches have been used to develop structure-based models to predict aquatic toxicity MoA: identification of substructural indicators based on the presence of specific molecular features (e.g., Verhaar et al. 1992; Russom et al. 1997; Nendza and Muller 2007; Enoch et al. 2008), and use of neural networks, hierarchical clustering or other machine learning methods with theoretical descriptors that describe a broad array of electronic, topological, and other features of a chemical (e.g., Yao et al. 2005; He and Jurs, 2005). Recently, Martin et al. (2013) used two dimensional chemical descriptors and machine based learning algorithms to predict aquatic toxicity MoA using a database of 924 chemicals. Linear discriminant analysis (LDA) and random forest MoA assignment models had high internal concordance and specificity and produced overall MoA prediction accuracies of 85-88% (Martin et al 2013). Barron et al. (2015) developed an integrated dataset of MoA assignments for over 1200 chemicals that included a diversity of metals, pesticides and other organic compounds that encompassed six broad and 31 specific MoAs for fish and aquatic invertebrates. Martin et al. (2015) then utilized this expanded MoA database to predict global or specific aquatic toxicity MoA with methods such as LDA and multiple linear regression. Mode of action prediction accuracy was greatest for acetylcholinesterase inhibitors and lowest for other neurotoxicants, the structurally diverse group of reactive chemicals, and MoAs with low representation in the database (Martin et al. 2015).

Bayesian networks have seen increasing use in hazard and risk assessment (Burden et al. 2000; Billoir et al. 2008; Iorio et al. 2009), but only limited application in aquatic toxicity and MoA prediction. A Bayesian network is a graphical representation of a joint probability distribution. The structure of a Bayesian network consists of individual nodes (i.e., random variables) connected by arcs (arrows) reflecting a dependence structure between the variables. Bayesian networks are built in a directed manner often reflecting causality between the variables though this is not a requirement. The absence of an arc between nodes signifies that the variables are conditionally independent. Algorithms are available for learning a Bayesian network structure from data. These algorithms can often be useful in high dimensional data where the structure and conditional dependence are not readily apparent. The automated learning process of a Bayesian network structure can either be unsupervised or supervised. In contrast to a more generalized representation of a joint probability distribution in unsupervised learning of a Bayesian network's structure, supervised learning focuses on representing the relationship of predictor nodes to a target variable. Bayesian networks have not been previously used to predict MoA, but have been applied in machine learning applications for diverse fields from cancer prediction to marketing and finance.

The objective of the current study was to elucidate the connection between molecular descriptors and MoA, and develop a Bayesian network prediction model for classifying chemicals by aquatic toxicity MoA. The model was developed following the steps of discretization, structure learning, network selection and validation, and sensitivity analysis. The

study utilized a recently published dataset containing over one thousand chemicals with MoA assignments (Barron et al. 2015), and two dimensional chemical descriptors generated for each chemical. Machine-learned Bayesian networks were chosen for their capabilities in analyzing uncertainties, and accounting for covariation among multiple variables (Barber 2012). For the current study, we focus on learning the Markov blanket of the data set with the MoA broad classifications as a target node. The Markov blanket of a target node consists of the node's parents, children, and the nodes that share children with the target node (Korb and Nicholson 2004). The target node or node in question, is screened from the rest of the nodes in the network from its Markov blanket if the value of each of the nodes in the blanket is known. Figure 1 shows an illustration of the concept. If the A_i predictor nodes in Figure 1 are known for certain, then no other node (i.e., the B_i nodes) will provide any additional information on the target variable. Identifying the Markov blanket of a complex dataset with numerous variables can simplify analytical and computational requirements by providing a subset of the key predictors in a dataset that are highly informative for making probabilistic inferences on a target variable.

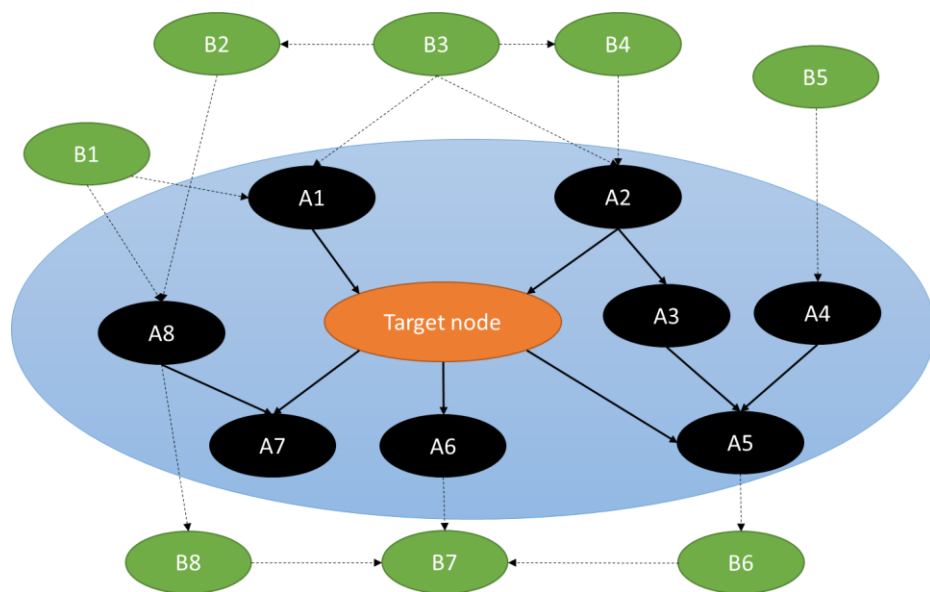


Figure 1. Representation of a Markov blanket in a Bayesian network. The orange oval is a target node or the variable of ultimate inferential interest in the analysis. The black ovals labelled A_i are variables that are used for predictions and are in the Markov blanket of the target node (i.e., either children, parents, or spouses of the target node). The green ovals labelled B_i are variables in the problem domain but not in the Markov blanket.

METHODS

General Approach

The approach to database and network model development is described in detail in the subsections below. The database used in developing and validating the Bayesian classification network was comprised of a dataset of known MoA assignments and a dataset of two dimensional theoretical chemical descriptors for over 1000 chemicals. Six preliminary models were initially constructed with two different discretization levels and three Markov blanket algorithms. An optimal structural coefficient was selected for each model from trade-offs between internal precision and structural complexity. A 10-fold cross validation was implemented for each candidate model and a final network was selected with the highest empirical performance from overall precision. Sensitivity analyses on the selected model examined the strengths of the relationships between the variables, and the capabilities of maximizing MoA probabilities using individual and multiple descriptors.

MoA and Descriptor Database Development

The MoA dataset was based on the MOAtox database of Barron et al. (2015) that encompassed the six broad acute aquatic toxicity categories of acetylcholinesterase inhibition (AChEI), narcosis, electron transport inhibition (ETI), ionoregulatory/osmoregulatory/circulatory impairment (IOCI), other neurotoxicity, and reactivity. Those MoA assignments had been determined for aquatic invertebrates and fish toxicity from over 1200 chemicals using a combination of international consensus classifications, QSAR predictions, and weight of evidence professional judgment based on an assessment of structural features and literature information (Barron et al. 2015). Up to 790 two dimensional theoretical molecular descriptors were then calculated for each chemical following the method of Martin et al. (2015) using the computational chemistry application TEST (USEPA 2016). The descriptor classes included E-state values and E-state counts, constitutional descriptors, topological descriptors, walk and path counts, connectivity, information content, two-dimensional autocorrelation, Burden eigenvalue, molecular properties, Kappa, hydrogen bond acceptor/donor counts, molecular distance edge and molecular fragment counts (USEPA 2008; Zhu et al. 2009; Martin et al. 2013, 2015). Definitions of key descriptors are in Supplemental Table S1 and available for download from the TEST internet site (USEPA 2008). Chemicals were removed from the database that included metals, other inorganic compounds, and organic salts because of the inability to compute descriptors for these ionic compounds leaving a total of 1098 chemicals used in the Bayesian network.

Bayesian Network Development

-Discretization

All Bayesian network analysis was conducted with Bayesialab 5.4.3 (Bayesia S.A.S. 2015). Each chemical entry in the MoA and descriptor dataset was used except several descriptors only containing zeroes were removed prior to analysis. Continuous variables were discretized using the decision tree algorithm with the broad MoA variable as the target node for choosing levels. The decision tree algorithm maximizes the information content between a target node (i.e., MoA) and predictor (i.e., a descriptor) from the discretization levels. In case of failure of

the decision tree discretization algorithm to identify intervals, we used density approximation, k-means discretization, and normalized equal distance in that order of priority for selecting a discretization algorithm. Discretization was run twice using both three and four discretization intervals as initial specifications in the algorithms. The number of intervals for the descriptors was chosen as a compromise between representation of the distribution and data availability for estimating probabilities (Conrady and Jouffe 2015).

-Structure learning

After discretizing the continuous variables, three supervised structure learning methods were applied. The first was a Markov blanket (MB) analysis and the second was an augmented Markov blanket (AMB) analysis and the third was a minimal augmented Markov blanket (MAMB) analysis. The MB analysis searches for the parents, children, and spouses of the target node. The AMB first identifies the MB and then the relationships among the parent, son, and spouse nodes (Fig. 1). The MAMB prunes the MB by eliminating nodes from the blanket through an additional equivalence class unsupervised learning step.

Bayesialab uses the minimum description length of candidate networks in its score-based algorithms for comparing Bayesian network structures (Conrady and Jouffe 2015). The structural coefficient (SC) modifies the complexity of the network through weighting the structure encoding (in bits) of the Bayesian network. In a Markov blanket analysis, the SC can influence the number of descriptors selected in the blanket as well as their connectivity. We adjusted the SC by finding the optimal value between complexity and data representation/precision similar to Harris et al. (2014) and Conrady and Jouffe (2015). The SC analysis generated networks with each of the supervised learning algorithms (i.e., MB, AMB, MAMB) and discretization levels with SCs ranging from 0.5 to 6.0 at 0.1 intervals. Bayesian networks with lower (<0.5) SC values were found to be difficult to computationally manage because of higher model complexity. Plotting the SC values against the structure:target precision ratio for each of the generated networks helped to visually identify candidate SCs. The inflection of the structure:target precision ratio plot was located to identify a SC that optimally did not underfit or overfit the model (Bayesia 2010; Harris et al. 2014). The target precision was also examined as a function of the SC before final selection where the height of the curve was used to prioritize the SC. In case of constant values for the precision across the range of SCs, the structure/data ratio was plotted against the SC to further enhance descriptor representation or prevent overfitting. A single SC value was identified for all candidate models except for the MAMB model with three discretization levels where two different SCs were tested.

-Network selection and validation

Targeted cross validation (k fold) was used to estimate the performance of the networks in out of sample predictions (Stone 1974; Conrady and Jouffe 2015). Prior to cross-validation work, a virtual data point was added to every conditional and unconditional probability table cell to prevent impossible assignments in the cross-validation and to improve later inferential analysis. The k-folds cross validation method splits the data set into k equal sized sets and uses k-1 sets for learning the model and one for validation. This is repeated for k rounds and synthesized results are reported (Conrady and Jouffe 2011). The data were randomly split into 10 sets for a 10 fold cross validation. The k-folds cross validation method was used to compare the precision of these candidate models and the final selected model was built with the AMB algorithm with three discretization levels and a SC of 1. Cross validation results are discussed

for this model including reliability and precision statistics from an occurrence matrix. The reliability statistic related the number of correct predictions for each MoA to the total occurrences that were predicted to be that or another MoA. The precision statistic related the number of occurrences that were predicted for a MoA to the total occurrences that should have been predicted as that MoA. Total precision sums the correct predictions for each MoA and relates this value to the total number of predictions that were made. All precision and reliability statistics were expressed as percentages.

For the AMB network with three discretization levels, a final check on the connections identified among the descriptors was made for overfitting based on the Kullback-Leibler divergence for the arcs between descriptors (Bayesia 2010). This check for lack of dependence was done for each parent-child relationship in the AMB network that was previously identified with the minimum description length score-based algorithm. The validity of the connections among the descriptors that was identified by the learning algorithm was examined using a statistical independence test (G_{KL} -test; p-values greater than 0.05 provide evidence of independence) (Bayesia 2010). From the tests, the p-values or independence probabilities, were used to check the significance of each identified relationship between the descriptors or between the descriptors and the target node (Thai et al. 2012; Harris et al. 2014).

-Sensitivity analysis

A comprehensive sensitivity analysis examined the relationship among the descriptors in the selected network and helped hone in on the influence of individual descriptors on MoA prediction. The sensitivity analysis was conducted using higher, moderate or lower values of the descriptors with each MoA. For a broad understanding of the relationships between variables, the highest and lowest values of mutual information, Pearson's correlation, Kullback-Leibler divergence, and node force between variables were examined globally on the network. Mutual information was used to examine the probabilistic dependence between the nodes in the network. Pearson's correlation gave a measure of the linear strength of the relationship between the variables. The Kullback-Leibler divergence can be used as a measure of the information gain from assuming a joint relationship between two variables in a network compared to an assumption of independence. Node force sums the Kullback Leibler divergence input to a node and output by the node. The nodes that have more direct relationships and greater dependence with other nodes have the highest node force. Sensitivity of each MoA to each descriptor was examined in tornado plots that displayed the influence of knowledge of the descriptor values on the probability of each MoA and provided information on the maximum strength of the individual relationships between each MoA and descriptor. The tornado plots display the range between the lowest and highest probability that can be achieved for each MoA from hard evidence placed on the corresponding descriptor states.

A target dynamic profile analysis (TDP) (Bayesia 2010) identified a pathway of states of the descriptors most likely to reduce uncertainty on the MoAs. The TDP checked the sensitivity of each MoA to multiple descriptors by maximizing their probability as the search criteria optimization specification. This analysis uses a search that can take into account the joint probability of the evidence to diminish the impact of low probability scenarios. Only hard evidence was used to establish the TDP under the assumption that there would be minimal uncertainty on the descriptors' states for a given chemical.

A target optimization tree (Bayesia 2010) was also used to identify additional pathways for reducing uncertainty on the MoAs from knowledge of descriptors in the Markov blanket. This allowed a fuller consideration of different evidence from multiple variables (descriptors). The tree presented a visualization of the hard evidence set for predicting a MoA through an exhaustive search that can also take into account the joint probability of the evidence to diminish the impact of low probability states. We limited the target optimization tree to three descriptors for depth of evidence and a maximum of two alternate trees corresponding to the breadth of evidence and utilized hard evidence for its derivation. Both the TDP and target optimization tree analyses were run twice, once using the joint probability of the evidence to weight the search methods and once not considering it but only considering the maximization of the probability of the MoA. The probability changes from initial marginal probability to final probability for the MoAs were used to score the pathways in the target optimization trees. When the joint probability of the evidence was considered in the tree construction, the probability change to the MoA was weighted by the joint probability of the evidence to calculate the optimization score.

RESULTS

Network Precision and Reliability

The total precision for the selected AMB model (Fig. 2) with three discretizations was 80.2% (Table 1). The best precision was for the AChEI MoA (93.5%) where 257 chemicals out of 275 were correctly classified. Model precision was poorest for the reactivity MoA (48.5%) where only 48 out of 99 reactive chemicals were correctly classified. Narcosis represented the largest class within the MoA dataset and had a precision of 80.0%, reflecting the total precision across all of the MoAs. False negatives for chemicals with the narcosis MoA most often fell into the ETI, neurotoxicity or reactivity MoAs. False negatives for all other MoAs were most often narcosis.

Global Model Results

The structure of the selected AMB model contained over a dozen connections between the descriptors in the Markov blanket of the MoA variable (Fig. 2). The p-values for each of the variable relationships were all 0.0000% from the G_{KL} -test indicating that the evidence is not strong enough to disprove the links identified by the AMB algorithm. The highest Kullback-Leibler divergence was between the MAXDN and Broad MoA nodes (0.5256) and the lowest was between SdsCH_acnt and SsssCH_acnt (0.0523). The highest mutual information was found between the broad MoA and SRW10 nodes (0.4913). The lowest was between xc4 and SdsCH_acnt (0.0495). The highest positive Pearson's correlation was between Broad and SdsssP (0.6839). The lowest positive correlation was between Broad and SdssNp (0.0055). The highest negative Pearson's correlation was between SdsssP and GATS1v (-0.6065) and the lowest negative Pearson's correlation was between Broad and DELS (-0.0044). Aside from the node force of the Broad MoA variable (5.2611), the highest node force was for MAXDN (1.2074). The lowest was for StsC (0.0553). The highest outgoing force for a descriptor was MAXDN (1.2074) followed by xc4 (0.3767). Thirteen descriptors had zero values for outgoing force indicating that these are leaf variables in the network. The lowest non-leaf descriptor outgoing force was for

SdsCH_acnt (0.0523). The highest incoming force for a descriptor was for xc4 (0.5539) followed by DELS (0.5405) and SRW10 (0.5069). The lowest incoming force was for StsC (0.0553). MAXDN had a zero for incoming force indicating that it is a root variable with no parents.

Table 1. Global cross validation results for mode of toxic action (MoA) predicted from chemical descriptors (n=number of chemicals in dataset). Data are: % reliability (R%) and % precision (P%). Row titles and numerical values in parentheses pertain to predicted cases and column titles and numerical values are for actual cases.												
MoA	AChEI ¹ (275)		ETI ² (61)		IOCI ³ (18)		Narcosis (460)		Neurotoxicity (185)		Reactivity (99)	
	R %	P %	R %	P %	R %	P %	R %	P %	R %	P %	R %	P %
AChEI ¹ (276)	93	93	0	0	0	0	5.1	3.0	1.5	2.2	0.4	1.0
ETI ² (76)	0	0	55	69	0	0	33	5.4	0	0	12	9.1
IOCI ³ (14)	0	0	0	0	86	67	7.1	0.2	7.1	0.5	0	0
Narcosis (471)	3.8	6.6	3.4	26	1.1	28	78	80	5.5	14	8.1	38
Neurotoxicity (184)	0	0	1.1	3.3	0.54	5.6	13	5.2	84	83	1.6	3.0
Reactivity (77)	0	0	1.3	1.6	0	0	36	6.1	0	0	62	48
1. acetylcholinesterase inhibition 2. electron transport inhibition 3. iono/osmoregulatory/circulatory impairment												

Mode of Toxic Action Sensitivity to Individual Descriptors

The sensitivity of each MoA to the individual descriptors is based on changes to the probability of each broad MoA from changes to each descriptor individually while allowing the other descriptors to co-vary with the state changes for a descriptor of concern (Fig. 3). The supporting information contains separate tornado plots of the sensitivity of each MoA to each individual descriptor (Figs. S1-S6).

AChEI had the greatest sensitivity to changes in SdssP with ranges from 0.081 to 0.95 probability depending on the state, indicating the importance of a phosphate group. GATS1v was the second most influential individual descriptor but J was also influential and is capable of raising the probability of AChEI to the second highest level below SdssP. ETI was most sensitive to SdssNp, suggesting that the number of nitro groups is critical in defining the MoA. Based on the range of probabilities for ETI, Hmax was the second most influential descriptor and Qv also has the potential to raise the probability of ETI to greater than 0.30. Weak positive relationships were found between ETI and most of the descriptors. Electron transport inhibitors were poorly represented in the database.

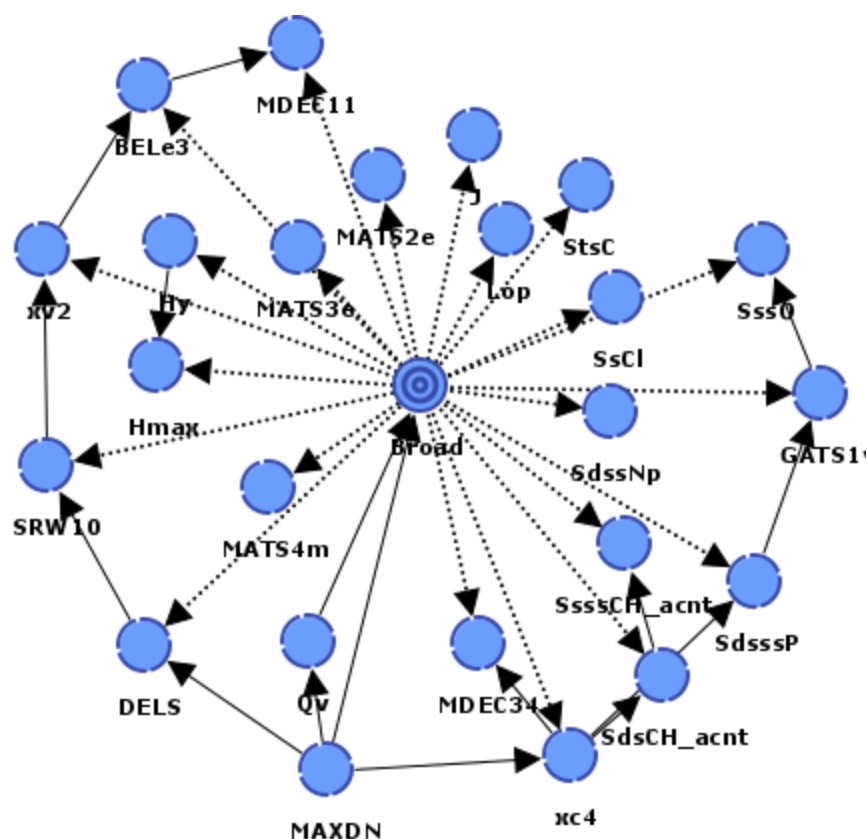


Figure 2. Structural components of the augmented Markov blanket model used for classifying modes of toxic action using chemical descriptors. The target (predicted) node (*Broad*) contains the broad mode of toxic action categories in its variable states. Dashed arcs delineate relationships where *Broad* is a parent node of a descriptor. Inferences made within the model are omnidirectional so that observing a descriptor changes the probability of the other descriptors as well as the target node depending on all other observations in the model and the conditional independence relationships contained in the structure.

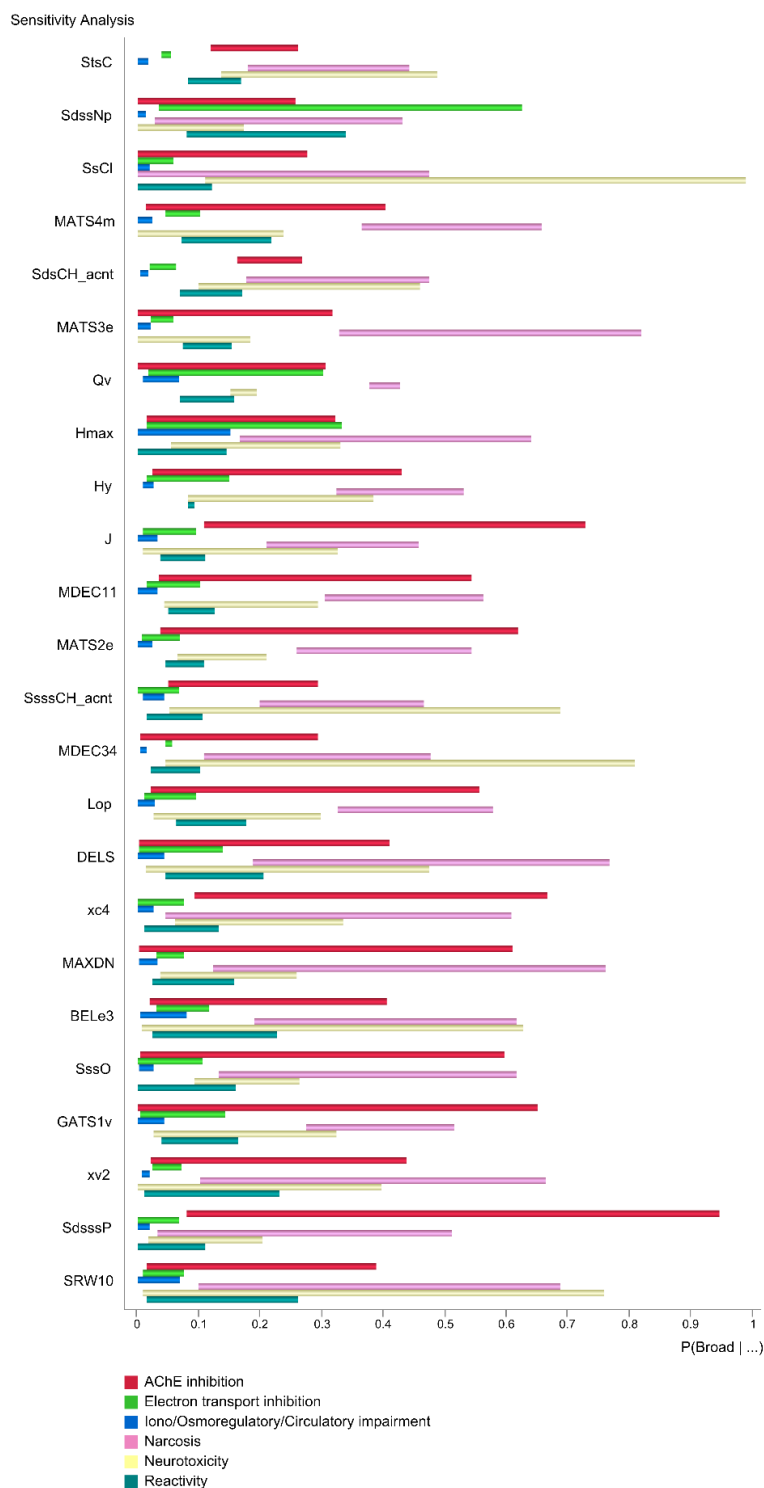


Figure 3. Sensitivity of the chemical modes of toxic action to each of the descriptors. Horizontal colored bars represent the range of probabilities for each mode of toxic action (i.e., the conditional probability of the broad mode of toxic action ($P(\text{Broad} | \dots)$) (x-axis)) from findings at the individual descriptors listed on the y-axis.

The number of data points available for IOCI was the lowest in comparison with other MoAs. Iono/osmoregulatory/circulatory impairment had Hmax as the most influential descriptor followed by BELe3, SRW10, and Qv. None of the states of the individual descriptors could raise the probability of IOCI above 0.20 on their own. The low overall probabilities for IOCI in the tornado plot indicates weak positive relationships with individual descriptors and the IOCI MoA.

In contrast, narcosis was the best represented MoA in the database. The descriptor with the greatest influence on the narcosis MoA was MAXDN with a probability range of 0.12 to 0.76. DELS and MATS3e could raise the probability of narcosis to greater values than MAXDN but MAXDN can lower the probability more giving it an overall greater range of potential probability changes than DELS and MATS3e. Numerous descriptors could lower the narcosis MoA to lower probabilities such as SRW10, which was the second most sensitive descriptor, xv2, xc4, SdsssP, SsCL, and SdssNP.

The neurotoxicity MoA was most sensitive to SsCL which has the potential to raise the probability of neurotoxicity from 0.11 to 0.99 alone. Although individual findings at multiple descriptors can lower the probability of neurotoxicity to less than 0.1, five descriptor states can raise the probability of its occurrence to greater than 0.5 (SsCL, MDEC34, SRW10, SsssCH_acnt, and BELe3).

The probability of reactivity was most sensitive to SdssNP. The range of probabilities of reactivity from this one descriptor extended from slightly lower than reactivity's initial probability (0.090) to 0.34. Several other descriptors could raise the initial probability of reactivity from 0.090 to a posterior probability of greater than 0.20. Like IOCI, reactivity had relatively weak positive relationships with the individual descriptors and several descriptor states that could lower the posterior probability of reactivity to close to zero from individual findings at their states.

Mode of Toxic Action Optimization by Multiple Descriptors

The results from the TDP analyses are displayed in the supporting information. Findings on multiple descriptors could lead to equal or greater than 0.99 probability of a MoA both when the joint probability of the evidence was considered in developing the pathway and when it was not. For the models that did not consider the joint probability of the evidence when selecting descriptors, fewer descriptors were necessary to reach the final probabilities of a MoA. The first nodes in the tornado plots that have the potential to affect the greatest range of probabilities were always the first nodes chosen in the TDP with the exception of ETI, IOCI, neurotoxicity, and reactivity when the joint probability was taken into account, and narcosis without taking into account the joint probability of the evidence. These initial descriptors chosen for the TDP were within the top six descriptors that individually created greater range variation with the target state or MoA's probability.

From the existing database alone, multiple descriptors are able to identify MoAs with high certainty even when constrained to follow paths that reflect the available data. For AChEI, only two descriptors were needed to reach or exceed 0.99 probability when considering the joint probability of the evidence and when not considering it. This was largely due to the influence of SdsssP which was the first descriptor for each and which achieved a greater than

0.94 probability of AChEI from one of its states alone. After SdsssP, different descriptors were identified in each of AChEI's target dynamic profiles for achieving a probability of 0.99. For ETI, only two descriptors were needed to reach or exceed a 0.99 probability of its occurrence without the joint probability considered but six descriptors were needed with the joint probability considered. For IOCI, five descriptors were needed to reach or exceed 0.99 probability without the joint probability considered and seven with the joint probability considered. For narcosis, eleven were needed with the joint probability considered and two without the joint probability considered. For neurotoxicity, an outcome on one descriptor was the minimum needed without the joint probability considered and three descriptor outcomes with the joint probability considered. For reactivity, nine descriptors were needed with the joint probability considered and four were needed without the joint probability considered.

Several descriptors that were used in the TDP to reach or exceed 0.99 probability for a single MoA were found to be unique to certain MoAs or were found multiple times across MoAs. For unique descriptors prioritized to reach 0.99 probability, Lop was only found in AChE joint, MAXDN was unique to narcosis joint, MATS3e and SssO were unique to narcosis and MDEC34 and xv2 were only in neurotoxicity joint. Hmax was in both ETI and ETI joint as well as IOCI and IOCI joint. SdssNp was in ETI, ETI joint, narcosis joint, and reactivity. Qv was in ETI joint, IOCI, narcosis joint, and reactivity joint. Bele3, GATS1v, Hy, J, SdsCH_acnt, SdsssP, SRW10, SsCl, SsssCH_acnt, and xc4 were found in at least three MoA's TDPs and used prior to or at the point of reaching 0.99. The rest of the descriptors that were prioritized for optimally reaching 0.99 (i.e., DELS, MDEC11, StsC) were found twice across the MoA's TDPs.

To examine multiple pathways for identifying the MoAs, target optimization trees were developed. Figure 4 contains the target optimization tree for AChEI while taking into account the joint probability of the evidence and Figure 5 contains the target optimization tree for AChEI without the joint probability of the evidence taken into account in the development of the pathways. All pathways in Figures 4 and 5 led to a probability of AChEI that is greater than 0.98. The red highlighted box at the leaf of the tree contains the final scenario with the highest overall score based on the joint probability (if considered in the evidence determination) and the change in probability from the initial probability of the MoA displayed in the root of the tree. Both of the primary (first) pathways in the target optimization trees for AChEI duplicate the steps in the corresponding TDPs. The two nodes in the TDP that optimally led to a probability of AChEI that was greater than 0.99 probability were SdsssP and J when the joint probability was not considered. In the same target optimization tree, SdsssP only appeared in one pathway as the initial variable and J appeared in two pathways, as an initial and secondarily selected variable. When the joint probability was considered (Fig. 4), J was no longer in the target optimization tree but SdsssP was still selected early and raised the probability of AChEI the most in the first branching. The highest scoring pathway with the joint probability contained SdsssP, MAXDN, and Qv leading to a 0.99 probability of AChEI. The highest scoring pathway without consideration of the joint probability contained SdsssP, J, and MATS2e leading to a probability of 0.998 (near certainty) of AChEI.

Neurotoxicity had the second highest precision from cross validation and its target optimization trees are displayed in Figures 6 and 7, with and without the joint probability of the evidence weighted in the pathways, respectively. Like AChEI, multiple scenarios with at most three predictors can lead to a probability equal to or greater than 0.99 for the neurotoxicity

MoA. The lowest final probability was 0.98 in a few lineages with the joint probability of the evidence considered. An outcome on only one descriptor was necessary to reach a 0.99 probability of neurotoxicity (SsCl) in the target optimization tree that did not consider the joint probability. For the target optimization tree that did consider the joint probability, SsCl was not selected at all. When the joint probability was considered, the 0.99 probability was still commonly reached but required several additional descriptors. The highest scoring pathway in the tree that considered the joint probability contained MDEC34, xv2, and J in the same order of priority. The highest scoring pathway in the tree that did not consider the joint probability contained SsCl, MDEC34, and SssO, duplicating the TDP. Both led to a probability of neurotoxicity greater than 0.98.

As can be seen, for AChEI and neurotoxicity several possible descriptor groupings can lead to a high probability of the MoA being true based on a few descriptors alone including the number of phosphate groups and chlorine atoms, respectively. Figures 8 and 9 contain the target optimization trees for narcosis. By contrast, some scenarios contained final probabilities that were lower than AChEI and neurotoxicity target optimization tree pathways, especially when the joint probability of the evidence was considered in the tree development. From the TDP analysis above, it was also observed that narcosis took as many or more descriptors when compared to the corresponding TDPs of neurotoxicity and AChEI to raise the probability to near certainty (0.99 or above). The final probabilities range from 0.76 to 0.91 at the terminal evidence set for scenarios with consideration of the joint probability. Additional descriptors could raise these probabilities further but the trees were constrained for examining the capabilities of only a handful of descriptors in predicting the MoAs. Without consideration of the joint probability, probabilities of narcosis being true ranged from 0.93 to near certainty. The highest scoring pathway with the joint probability considered contained MAXDN, SdsCH_acnt, and Qv, duplicating the TDP. The highest scoring pathway without consideration of the joint probability contained MATS3e, SssO, and DELS, also duplicating the initial evidence set of the TDP. Narcosis and neurotoxicity both contained SssO in their highest scoring pathways without considering the joint probability. Narcosis and AChEI both contained MAXDN and Qv in their highest scoring pathways with the joint probability considered. Neurotoxicity and AChEI both contained J in highest scoring pathways. For neurotoxicity, J was in the pathway that considered the joint probability and for AChEI it was in the pathway without consideration of the joint probability. Additional target optimization trees for the other modes of action are found in the supporting information.

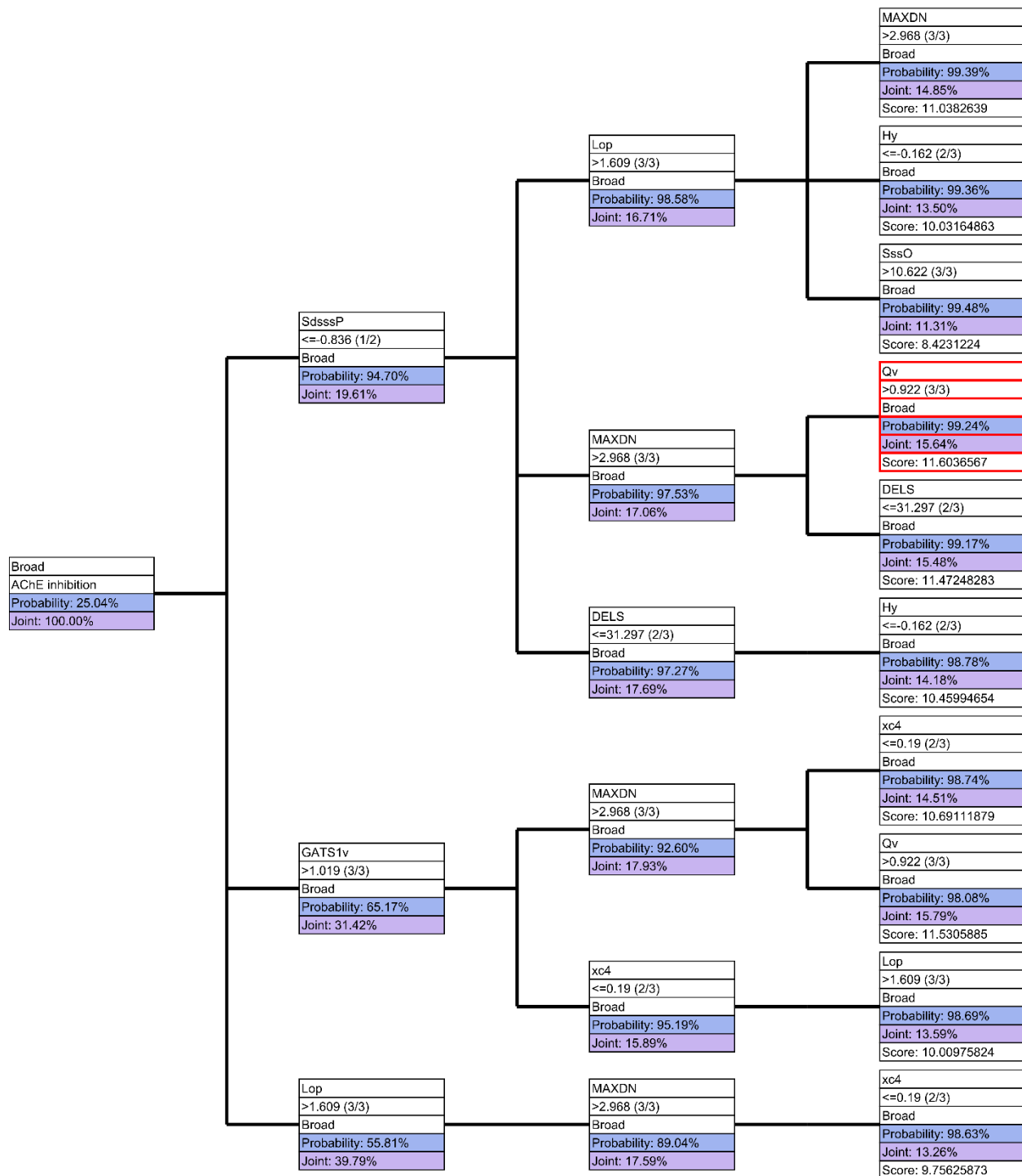


Figure 4. Target optimization tree for acetylcholinesterase inhibition (AChEI) with the joint probability of the evidence taken into consideration. The rows on the leaves contain the score for the pathway built on the probability change from the prior probability of the mode of toxic action weighted by the final joint probability. The red highlighted descriptor (Qv) is the terminal node of the pathway with the highest score.

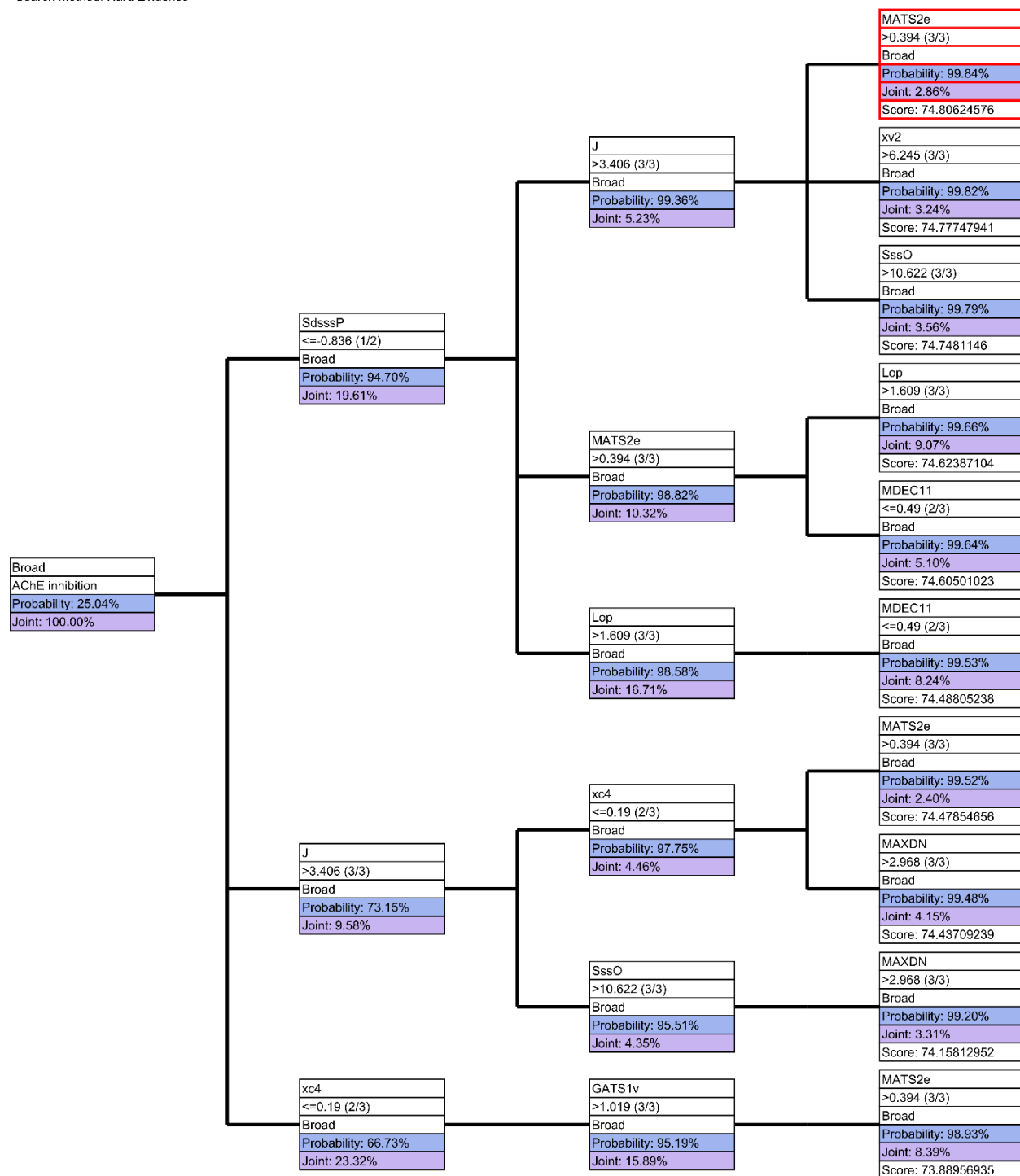


Figure 5. Target optimization tree for acetylcholinesterase inhibition (AChEI) without taking into account the joint probability of the evidence. The rows on the leaves contain the score for the pathway built on the probability change from prior probabilities before findings on descriptors are entered. The red highlighted descriptor (MATs2e) is the terminal node state of the pathway with the highest score.

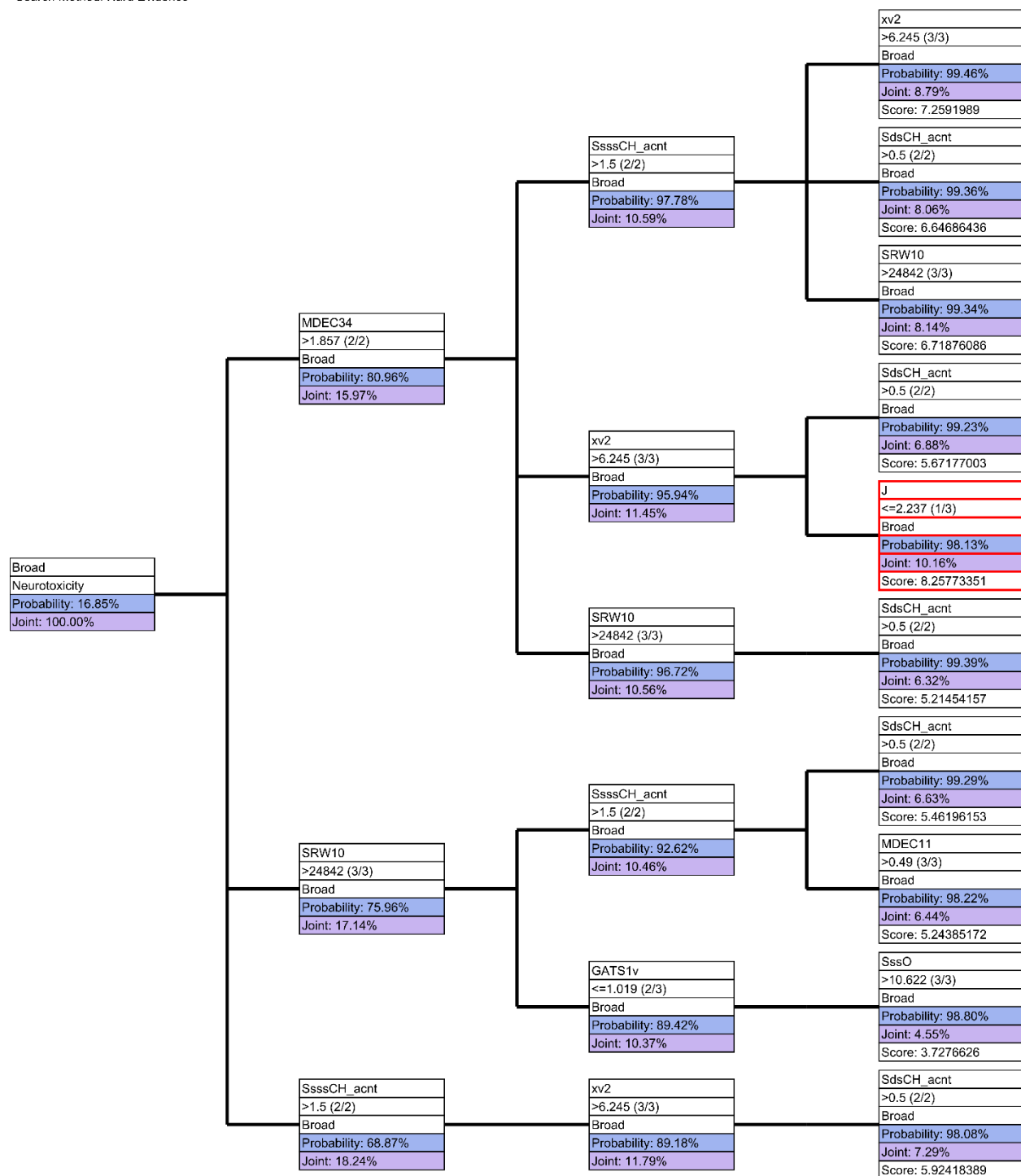


Figure 6. Target optimization tree for neurotoxicity with the joint probability of the evidence taken into consideration. The rows on the leaves contain the score for the pathway built on the probability change weighted by the final joint probability. The red highlighted descriptor (J) is the terminal node state of the pathway with the highest score.

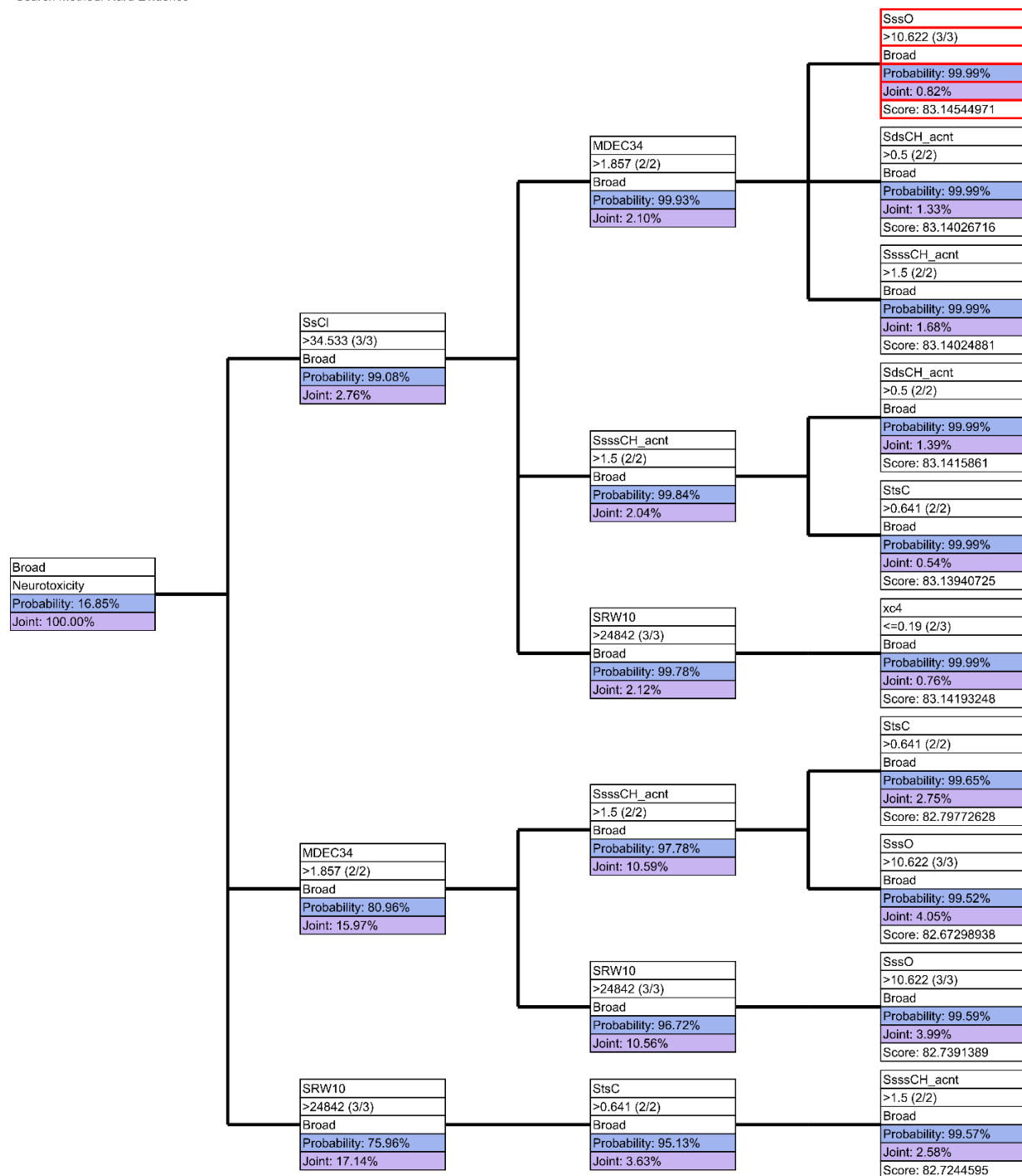


Figure 7. Target optimization tree for neurotoxicity without taking into account the joint probability of the evidence. The rows on the leaves contain the score for the pathway built on the probability change. The red highlighted descriptor (Ssso) is the terminal node state of the pathway with the highest score.

19

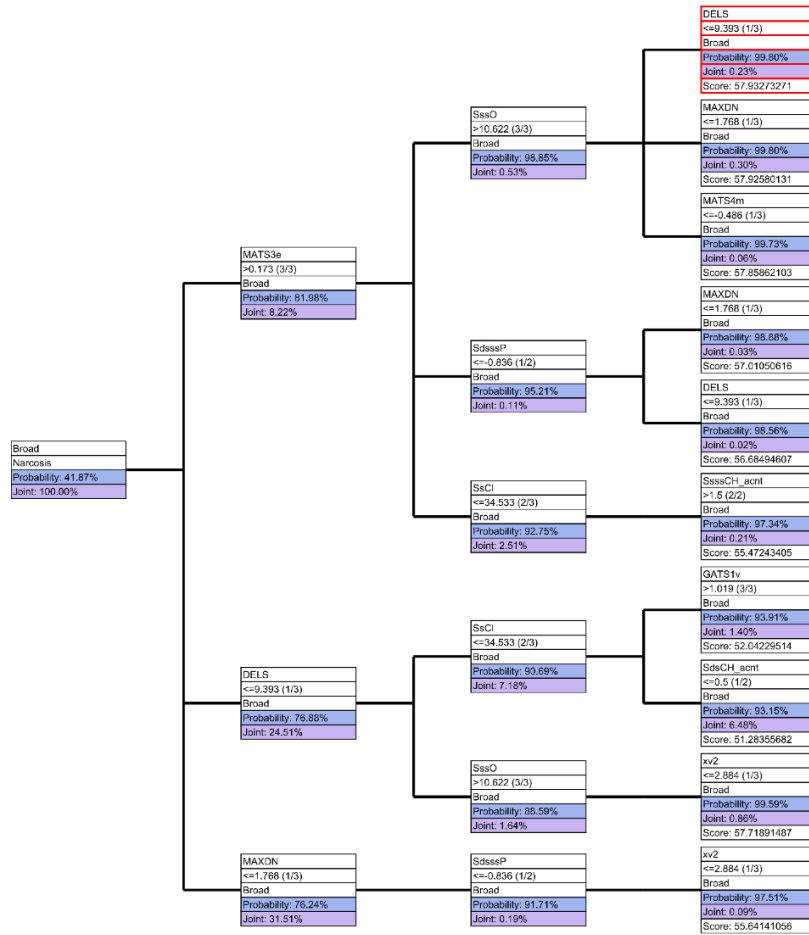


Figure 9. Target optimization tree for narcosis without taking into account the joint probability of the evidence. The rows on the leaves contain the score for the pathway built on the probability change. The red highlighted box (DELS) is the terminal node state of the pathway with the highest score.

DISCUSSION

Compound classification methods offer the ability to associate structural features and chemical properties with specific types of biological activities and to screen for and predict biologically active compounds (Bajorath 2001). Molecular similarity is the basis of these approaches, which typically requires the use of chemical descriptors that capture a broad range of molecular characteristics and definition of the theoretical chemical space of the model (Martin et al. 2002; Bajorath 2001). Hundreds of molecular descriptors are available in the literature that can be calculated using computational chemistry software such as Dragon which provides more than 1600 theoretical descriptors (e.g., Mauri et al. 2006). In the current study, a diversity of two dimensional chemical descriptors were calculated and assigned using the computational chemistry application TEST (USEPA 2008, 2016), including E-state values and E-state counts, constitutional descriptors, topological descriptors, walk and path counts, connectivity, information content, two-dimensional autocorrelation, Burden eigenvalue, molecular property, Kappa, hydrogen bond acceptor/donor counts, molecular distance edge and molecular fragment counts (Martin et al. 2013, 2015). Previous studies have demonstrated high performance of two dimensional descriptors in both molecular similarity analysis and compound classification (Bajorath 2001), although they have seen only limited application in aquatic toxicology (e.g., Papa et al. 2005; Martin et al. 2013, 2015).

Interpretation of computational chemistry descriptors is typically challenging because they are based on hundreds of calculated attributes of a molecule that may have positive or negative values. For example, the electrotopological state index represents the electronic state of an atom as perturbed from electronic influence of other atoms in the molecule (Hall and Kier, 1995). The index reflects the contribution of a specific group of atoms (e.g., -CH₃, =CH₂-) and depends on the environment of that group of atoms within a specific molecule (Hall and Kier, 1995). As another example, Q_v is a whole molecule polarity index that decreases in value as polarity increases (Votano et al. 2004). Of the over 700 descriptors assigned by the TEST computational chemistry application, only 24 were used to establish a Markov blanket around the MoA variable and 22 of these 24 were in a pathway prior to reaching or at the point of exceeding 0.99 probability of a MoA in the TDP analysis. The Markov blanket analysis used in the current network provided a method that reduced the inherent complexity of numerous chemical descriptor assignments by determining the most important predictors in terms of probabilistic dependence with the MoAs.

The Bayesian network had an overall prediction accuracy (model precision) of 80%, which is similar to the 86% accuracy (correct fraction) of the LDA model of Martin et al. (2015) using two dimensional theoretical chemical descriptors and the same MoA dataset of Barron et al. (2015). AChEI MoA had the highest prediction accuracy in both the network (93.5%) and LDA (98 %), and reactivity the lowest (network: 48%; LDA: 33%). Prediction accuracy of chemicals with an MoA of neurotoxicity was highest in the network model (83%; LDA: 65%), whereas narcosis prediction accuracy was highest in LDA (98%) compared to the network (80%) (Martin et al. 2015). These results suggest that MoA prediction accuracy may be a function of the intrinsic uncertainty in the theoretical descriptors and MoA classification, rather than fundamental differences in the statistical modeling approach. As knowledge of MoA

assignments and theoretical relationships between descriptors and MoAs advances, model performance and predictive capabilities should likewise improve.

For machine learned Bayesian networks, the minimum description length scores are useful for establishing the relationships among variables from the available data. With deterministic models, rules can be difficult to establish and interpret and complexity can overwhelm understanding of the model and the processes being represented. However, the nonparametric quantitative methods used in the current approach require larger amounts of data than might be found in some datasets. For the current model, MoAs with low numbers of data generally had relatively diminished predictive accuracy and data coverage for establishing conditional probabilities. In contrast, Martin et al. (2013, 2015) attributed poor predication accuracy of the reactivity MoA to a large range in reactivity mechanisms and general structural diversity. Within the current model, pathways constrained to a relatively small number of descriptors still led to high probabilities for each MoA despite varying data coverage.

Bayesian approaches can decompose complex multivariate problems into graphical networks that can be visualized and that support bidirectional inferences under various levels of uncertainty. The probabilistic focus inherent in a Bayesian network is also conducive to improvements in structuring and quantification as the knowledge and evidence improves over time. Coupling a Bayesian network with subjective interpretations of probability maintains focus on missing information and the imprecision from insufficient knowledge of true relationships. In the current study, the MOATox database of Barron et al. (2015) was taken as the best compendium of knowledge currently available regarding the acute aquatic toxicity MoA for a diversity of chemicals. The ability to predict unknown MoAs from existing data will become more precise as the knowledge base of MoA increases. Bayesian frameworks can provide useful pathways to improved predictive and diagnostic tools for aquatic toxicology and other problems with high dimensional data and multiple uncertain outcomes.

Acknowledgments

This research was supported in part by an appointment to the ORISE participant research program supported by an interagency agreement between the U.S. EPA and the U.S. Department of Energy. We thank Tom Stockton and Chris Grulke for review of a draft of the manuscript. The conclusions may not necessarily reflect the views of EPA and no official endorsement should be inferred.

References

- Bajorath, J. 2001. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J Chem Inf Comp Sci* 41:233-245.
- Barber D. 2012. Bayesian Reasoning and Machine Learning. Cambridge University Press, Cambridge, UK.
- Barron MG, Lilavois CR, Martin TM. 2015. A comprehensive mode of action and acute aquatic toxicity database for predictive model development. *Aquat Tox* 16:102-107.

- Bayesia. 2010. *Bayesialab User Guide*. Bayesia S.A.S., Laval Cedex, France.
- Bayesia S.A.S. 2015. Bayesialab software version 5.4.3. <http://www.bayesia.com/>. Accessed 05/16/2016.
- Billoir E, ML Delignette-Muller, ARR Péry, S Charles. 2008. A Bayesian Approach to Analyzing Ecotoxicological Data. *Environ. Sci. Technol.* 42:8978–8984.
- Burden FR, DA Winkler. 2000. A Quantitative Structure–Activity Relationships Model for the Acute Toxicity of Substituted Benzenes to *Tetrahymena pyriformis* Using Bayesian-Regularized Neural Networks. *Chem. Res. Toxicol.* 13: 436–440.
- Conrady S, Jouffe L. 2011. Microarray analysis with Bayesian networks: Using Bayesialab for cancer type classification. Bayesia USA, Franklin, TN.
- Conrady S, Jouffe L. 2015. *Bayesian Networks and Bayesialab: A Practical Introduction for Researchers*. Bayesia USA, Franklin, TN.
- Enoch SJ, M. Hewitt, M.T.D. Cronin, S. Azam, J.C. Madden. 2008. Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere* 73:243-248
- Hall LH, Kier LB. 1995. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J Chem Inf Comput Sci* 35:1039-1045.
- Harris M, Bhuvaneshwar K, Natarajan T, Sheahan L, Wang D, Tadesse MG, Shoulson I, Filice R, Steadman K, Pishvaian MJ, Madhavan S, Deeken J. Pharmacogenomic characterization of gemcitabine response – a framework for data integration to enable personalized medicine. *Pharmacogenetics and Genomics* 24 (2): 81-93.
- He L, P.C. Jurs. 2005. Assessing the reliability of QSAR model's predictions *J. Mol. Graph. Model.* 23: 503–523.
- Iorio F, R Tagliaferri, D di Bernardo. 2009. Identifying Network of Drug Mode of Action by Gene Expression Profiling. *Journal of Computational Biology* 16: 241-251.
- Korb KK, Nicholson AE. 2004. *Bayesian Artificial Intelligence*. Chapman & Hall/CRC, Boca Raton, FL.
- Martin YC, JL Kofron, LM Traphagen. 2002. Do structurally similar molecules have similar biological activity? *J Med Chem* 45:4350-4358.
- Martin T.M., Grulke, C.M., Young, D.M., Russom, C.L., Wang, N., Jackson, C.R., Barron, M.G., 2013. Prediction of aquatic toxicity mode of action using linear discriminant and random forest models. *J. Chem. Info. Model.* 53, 2229-2239.

Martin TM, Young DM, Lilavois CR, Barron MG. 2015. Comparison of global and mode of action-based models for aquatic toxicity. *SAR QSAR Environ Res* 26:245-262.

Mauri A, V Consonni, M Pavan, R Todeschini. 2006. MATCH Commun Math Comput. Chem. 56: 237-248.

Nendza M, Muller M. 2007. Discriminating toxicant classes by mode of action: 3. Substructure indicators. *SAR and QSAR in Environmental Research* 18:155-168.

Papa E, F Villa, P Gramatica. 2005. Statistically validated QSARs, based on theoretical descriptors, for modeling aquatic toxicity of organic chemicals in *Pimephales promelas* (fathead minnow). *J Chem Inf Model* 45:1256-1266.

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ. Toxicol. Chem.* 16, 948-967.

Stone M. 1974. Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*: 36(2):111-147.

Thai H, Campo DS, Lara J, Dimitrova Z, Ramachandran S, Xia G, Ganova-Raeva L, Teo C-G, Lok A, Khudyakov Y. 2012. Convergence and coevolution of Hepatitis B virus drug resistance. *Nature Communications* 3:789.

USEPA. 2008. Molecular Descriptors Guide. Description of the molecular descriptors appearing in the toxicity estimation software tool. V. 1.0.2. U.S. Environmental Protection Agency. [\[www.epa.gov/chemical-research/molecular-descriptors-guide\]](http://www.epa.gov/chemical-research/molecular-descriptors-guide)

USEPA. 2016. User's guide for T.E.S.T. (version 4.2) (Toxicity Estimation Software Tool). A program to Estimate Toxicity from Molecular Structure. EPA/600/R-16/058. U.S. Environmental Protection Agency, Cincinnati, OH. 63 pp. [www.epa.gov/chemical-research/toxicity-estimation-software-tool-test]

Verhaar, H.J.M., van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental pollutants part 1, Structure activity relationships for prediction of aquatic toxicity. *Chemosphere* 25, 471–491.

Votano JR, Parham M, Hall LH, Kier LB, Oloff S, Tropsha A, Xie Q, Tong W. 2004. Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis* 19:365-377.

Yao, X.J., Panaye, A., Doucet, J.P., Chen, H.F., Zhang, R.S., Fan, B.T., Liu, M.C., Hu, Z.D., 2005. Comparative classification study of toxicity mechanisms using support vector machines and radial basis function neural networks. *Anal. Chim. Acta* 535, 259–273.

Zhu, H., Martin, T.M., Ye, L., Sedykh, A., Young, D.M., Tropsha, A., 2009. Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. *Chem. Res. Toxicol.* 22:1913-1921.