JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

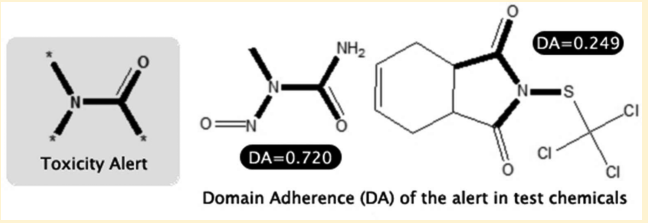
Optimizing Predictive Performance of CASE Ultra Expert System Models Using the Applicability Domains of Individual Toxicity Alerts

Suman K. Chakravarti,*[,†] Roustem D. Saiakhov,[†] and Gilles Klopman[†,‡]

[†]Multicase Inc., 23811 Chagrin Boulevard, Suite 305, Beachwood, Ohio 44122, United States

[‡]Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, United States

ABSTRACT: Fragment based expert system models of toxicological end points are primarily comprised of a set of substructures that are statistically related to the toxic property in question. These special substructures are often referred to as toxicity alerts, toxicophores, or biophores. They are the main building blocks/classifying units of the model, and it is important to define the chemical structural space within which the alerts are expected to produce reliable predictions. Furthermore, defining an appropriate applicability domain is required as part of the OECD guidelines for the validation of quantitative structure−activity relationships (QSARs). In this respect, this paper describes a method to construct applicability domains for individual toxicity alerts that are part of the CASE Ultra expert system models. Defining applicability domain for individual alerts was necessary because each CASE Ultra model is comprised of multiple alerts, and different alerts of a model usually represent different toxicity mechanisms and cover different structural space; the use of an applicability domain for the overall model is often not adequate. The domain for each alert was constructed using a set of fragments that were found to be statistically related to the end point in question as opposed to using overall structural similarity or physicochemical properties. Use of the applicability domains in reducing false positive predictions is demonstrated. It is now possible to obtain ROC (receiver operating characteristic) profiles of CASE Ultra models by applying domain adherence cutoffs on the alerts identified in test chemicals. This helps in optimizing the performance of a model based on their true positive−false positive prediction trade-offs and reduce drastic effects on the predictive performance caused by the active/inactive ratio of the model's training set. None of the major currently available commercial expert systems for toxicity prediction offer the possibility to explore a model's full range of sensitivity−specificity spectrum, and therefore, the methodology developed in this study can be of benefit in improving the predictive ability of the alert based expert systems.



1. INTRODUCTION

In silico techniques for predicting toxic manifestations of chemicals have gathered considerable momentum in recent years mainly as a result of increased availability and accessibility of toxicity data e.g. DSSTox,[1] ChemIDplus,[2] etc., important legislative changes, e.g. REACH,[3] involvement of regulatory bodies,[4] and significant developments in the computer hardware and predictive algorithms. In silico techniques cover a range of methods from single equation quantitative structure−activity relationships (QSARs) to complex artificial intelligence based expert systems and are well covered in multiple reviews.[5] A detailed review of available techniques is not the subject of this publication, we will however mention few crucial points.

Simple QSARs are often adequate to explain variation of toxicity or medicinal properties of focused training sets of limited sizes.[6] However, predicting toxicity of chemicals is much more challenging and is very different from modeling therapeutic properties.[5] Usually a chemical database of a particular toxicological end point contains diverse chemical structures and more importantly multiple mechanisms leading to the same end point. Poor quality of the experimental data of the toxicity end points also sometimes adds to the complexity. Computer based expert systems address such inherent complexities and difficulties of modeling toxicological end points (e.g., carcinogenicity,[7] mutagenicity,[8,9] reproductive and developmental toxicity,[10,11] hepatotoxicity,[12] etc).

A computer expert system is an artificial intelligence based computer program with the ability to learn and make decisions much like a human expert within a particular knowledge domain.[13] In the field of QSAR, expert systems consist of a knowledge base comprised of structural features or fragments of chemicals that are statistically or mechanistically related to the biological end point.[14] Several commercial computer based expert systems are available and used for safety assessment of drugs and chemicals (e.g., MC4PC,[15] Derek,[16] and Leadscope Products[17]). To conduct the current study, we have used the CASE Ultra expert system which is mainly influenced by the MCASE methodology (e.g., CASE,[18] MC4PC,[15,19] CaseTox[20]) and falls under the category of fragment based QSAR expert systems. MCASE methodology[15] and its applications have been discussed in detail in several publications.[9,21,22]

Received: February 29, 2012
Published: September 4, 2012

   dx.doi.org/10.1021/ci300111r | J. Chem. Inf. Model. 2012, 52, 2609−2618

Being the product of a fragment based QSAR expert system, a CASE Ultra model mainly consists of a set of "positive alerts" or "biophores", i.e. those fragments that were identified as statistically related to activity, and "deactivating alerts" or "biophobes", i.e. those fragments that are identified as statistically related to the inhibition of activity. The positive alerts are the main building blocks of the model and are responsible for identifying active chemicals during prediction.[23] The ability of any CASE Ultra model to correctly differentiate toxic compounds from nontoxic ones critically depends on the discriminating power of these positive alerts.

A positive alert may however perform poorly (i.e., may cause false positive predictions) in correctly identifying a toxic chemical during prediction if the positive alert is found in a test chemical which is completely different from the training set chemicals from which the alert was originally derived. In this respect, a toxicity alert is similar to a QSAR and a domain of applicability[24] needs to be defined for the alert. The applicability domain of an alert defines the range of compounds for which the alert can be confidently held responsible as a reason for their toxicity. The process and consequences of defining the alert applicability domains in CASE Ultra is the subject of this paper.

In a recent publication, Ellison et al.[24] have mentioned that the structural alert models often fulfill only four of the five OECD principles and they lack adequately defined applicability domains despite it being a requirement in the OECD guidelines for the validation of QSARs.[25−27] Ellison et al. described various methods of defining applicability domain for the mutagenicity model of Derek for Windows.[24] However, the described methods are implemented for the whole model which we consider inadequate because each model consists of multiple alerts and the alerts are often very different from each other from the perspective of toxicity mechanisms and structural diversity of the covered training set chemicals (member chemicals), and therefore, a single applicability domain defined for the overall model is too general and not sufficient to minimize false positive predictions.

Another toxicity prediction tool called *lazar*[28] (lazy structure−activity relationships) has successfully shown the importance of defining applicability domain for the query structures. In *lazar*, the applicability domain of a query structure (test chemical) is defined using similarity determined with respect to the biological activity. However, the present work differs significantly from *lazar*. In the present study, the domain of applicability of the individual positive alerts identified in the query compounds are being defined as opposed to the domain of applicability for the query compound itself. A modified *k*-nearest neighborhood was employed by *lazar* for the prediction of activity whereas CASE Ultra uses a stepwise hierarchical learning procedure to identify a set of positive and deactivating alerts, and the presence and absence of these alerts in query compounds are the main criteria in activity prediction. These differences have a marked effect on the prediction results. In *lazar*, use of the applicability domain results in a significant drop in the coverage (the percentage fraction of the test chemicals that have valid predictions) whereas in CASE Ultra the coverage does not decrease as the allowed domain adherence of the positive alerts in the query compounds is increasingly tightened.

The present work also demonstrates the use of the domain adherence measurement of positive alerts to obtain ROC (receiver operating characteristic) profiles during prediction. A gradually increasing cutoff to the domain adherence value of the alerts can be imposed that systematically decreases a model's tendency for false positive predictions and enables the study of the true positive−false positive trade-off tendency of the model (drop in the test chemical coverage would have been a disadvantage in this regard). The ROC profile is a better measure of a model's performance and helps to identify the optimal settings for a model's predictive ability. We think that this work is also a right step forward in resolving concerns[29] that MCASE methodology cannot be adapted to a specific chemical space.

2. DATA SETS

For the purpose of this study, we have obtained mutagenicity data from different sources and used them as multiple training sets and as an external test set. For building CASE Ultra models, mutagenicity assay data mentioned in the two publications by Matthews et al.[30,31] was used. The results of the original mutagenicity tests were transformed to numerical activity units using the method described in detail in the above-mentioned publications. For example, when a chemical was evaluated in only one study and was reported as inactive (negative), with marginal activity (weakly positive) or active (positive), the results were converted to 10, 25, and 35 activity units, respectively. In a case when a chemical is evaluated in more than one related mutagenicity end point (composite data sets), a weight of evidence method was used to calculate an overall activity. If the chemical was active in one, two, three, or more studies, it was assigned 35, 45, 55, or more activity units, respectively. Chemicals that were positive in only one study, but inactive in one, two, three, or more studies, were assigned 30, 25, 19, or less units, respectively. In the present study the compounds that have assigned activity units between 20 and 29 are considered marginally active (M), compounds with activity units more than 29 are considered active (A), and compounds with activity units below 20 are considered inactive (I). The following three data sets were used to build models in this study:

1. **A7B**: This data set contains 3535 chemicals and the mutagenicity data is from the *S. typhimurium* histidine reversion gene mutation test using tester strains TA97, TA98, TA100, TA1535, TA1536, TA1537, and TA1538. The distribution of active, marginal and inactive chemicals (A/M/I) is 1568, 82, 1885, respectively.
2. **A7E**: This data set contains 599 chemicals with fungal mutagenicity data from all gene mutation tests using *Saccharomyces cerevisiae*, *Aspergillus nigans*, and *Neurospora crassa* (A/M/I = 334/7/258).
3. **A7H**: This data set contains 584 chemicals with *Drosophila melanogaster* heritable (reciprocal) translocation (ht) and sex-linked recessive lethal (slrl) gene mutation tests (A/M/I = 279/1/304).

The nonproprietary training data used to build the models were available through a Research Collaboration Agreement entitled "Enhancement of in Silico Decision Support Tools for the Evaluation of Drug Safety" between MultiCASE Inc. and The US Food and Drug Administration's Center for Drug Evaluation and Research.[32]

For external validation tests, we have used the Ames mutagenicity benchmark data set from the publication of Hansen et al.[29] This mutagenicity data set originally contained 6512 chemicals from a variety of sources. According to

Hansen's paper, a compound is considered as Ames positive if it significantly induces revertant colony growth in one out of five strains with or without S9. Whereas, a compound is considered negative if it does not induce significant revertant colony growth in any strain with or without S9 mix.[29] For our purposes, we have matched this data set with the 3535 chemicals from the A7B data set mentioned above and used only the subset of 4072 chemicals (A/I = 2328/1744) that are not present in the A7B data set.

3. METHODS

Main Methodology of CASE Ultra. CASE Ultra's algorithm is mainly influenced by the MCASE methodology[15,18] and can automatically generate a predictive model from a training set of noncongeneric compounds with associated biological or toxicity data. The training set ideally should contain examples of both active and inactive chemicals. CASE Ultra has some noteworthy advancements over its predecessor, e.g. the identified alerts are no longer limited to linear paths of limited size or limited branching pattern, the training sets could be larger than 8000 molecules and can take advantage of computers with multiple core processors.

To build a model, CASE Ultra picks up one active chemical at a time from the training set and systematically generates a list of fragments for that chemical. Each fragment's relevance for activity is then determined using a two-objective criteria comprised of Shannon's entropy[33] as a fitness measure and the number of the active training set molecules containing this fragment (fragments that are optimal based on the two objectives, i.e. the ones that cannot be replaced by any other fragment without degrading one or both objectives, are selected and then sorted in descending order of the number of their active chemicals). Shannon's entropy for a particular fragment is calculated using the following formula:

$$\text{Shannon's entropy} = -1(pA \log_2(pA) + pI \log_2(pI)) \quad (1)$$

where pA is the fraction of the active molecules among the training set chemicals that contain that fragment and pI is the fraction of the inactive training chemicals for the same fragment. If any marginally active training chemicals contain the alert, their count is divided in active and inactive components using their locations on the activity scale prior to computing pA and pI. A top few fragments (based on the aforementioned two-objective criteria, e.g. fragments that have low entropy as well as supported by higher number of active training chemicals) are selected. These fragments are considered as *potential positive alerts*. The fragment generation procedure produces simple linear chains of varying lengths and branched fragments as well as complex substructures generated by combining simple fragments. When the algorithm has finished scanning all the active chemicals, a search is made in the accumulated list of the potential positive alerts to find the alert that covers the highest number of active chemicals, and it is added to the final list of positive alerts. This step is repeated until enough positive alerts were identified to cover all the active chemicals in the training set. It is important to note here that the number of the potential positive alerts generated during model building is usually much higher than the number of the positive alerts that ultimately were able to find their place in the final list. However, as we will show in the later sections, the potential positive alerts that did not make in the final list could be effectively utilized to build domain of applicability for the positive alerts instead of rejecting them completely.

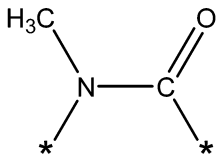
Once a final set of positive alerts is identified, CASE Ultra attempts to build separate local QSARs for each positive alert in order to explain the variation in activity within the training set chemicals covered by that alert. In addition, deactivating alerts are found using a very similar process but by scanning inactive chemicals and finding fragments that occur mainly in inactive chemicals.

This collection of positive and deactivating alerts constitutes a model for a particular end point and can be used for predicting activity in test chemicals. During prediction, a test chemical is scanned against the list of the model's positive and deactivating alerts, and if no positive alerts could be identified in it, the chemical is considered inactive. In general, if the test chemical contains one or more positive alerts, it is predicted as "active". However, this active prediction call can be changed if the local QSAR of the positive alert modifies the prediction. The presence of a deactivating alert alongside a positive alert renders the prediction call as "inactive". If more than one positive alerts are present, then the one with the highest number of active chemicals is used, and in the case of more than one deactivating alert, the one with the highest number of inactive chemicals is used. If a test chemical contains a positive alert that has been seen in just one or two active training set chemicals, the prediction result is considered "inconclusive" because of the alert's low statistical confidence. CASE Ultra has the ability to recognize unusual features/fragments in test chemicals that do not match training data (unknown structural fragments), and this imposes a type of global applicability domain for the overall model. The presence of more than three unknown structural fragments in the test chemical results in an "out of domain" call. The percentage of the test chemicals with either an active or inactive prediction call is termed as the test coverage.

Defining the Applicability Domain for a Positive Alert. Building the applicability domain for a positive alert essentially means defining a chemistry space using the structural aspects of the training set chemicals that contain the alert. The definition of the chemical space and how the distances are measured within it is very important. A chemistry space built using all the fragments obtained from the whole structures of the member chemicals irrespective of their relation with the end point is not suitable. This is due to the fact that often the training set member chemicals of a positive toxicity alert are very diverse (e.g., an aromatic nitro group from a mutagenicity model). As a result, two chemicals may seem totally different based on their overall structural similarity, but they are mutagenic just because of the presence of the nitro group. Therefore, we used only those fragments to build the applicability domains that were found to be statistically related to the end point in question as described in the main methodology. This created a use for all those potential positive alerts which were identified during the model building. The majority of these alerts did not make into the final list of positive alerts but were statistically related to the end point nonetheless. The domain of a particular positive alert is thus simply a list of statistically relevant fragments coupled with their frequency of occurrence in the member chemicals of the alert that are active (multiple occurrence of a fragment in the same chemical is counted as one occurrence).

As an example, a positive alert (ID: A7E_197) is shown in Table 1 which is from the model A7E (fungal composite mutagenicity model), and it is essentially an amide group. This

Table 1. Partial List of Fragments That Form the Applicability Domain of a Positive Alert (ID: A7E_197), Taken from the Fungal Composite Mutagenicity Model A7E



Alert ID: A7E_197[a] (T/A/I = 25/18/7)

no.	T/A/I[b]	Shannon's entropy	frequency[c]	domain fragments[d]
1	46/44/2	0.258	0.833	N2═O
2	44/43/1	0.156	0.833	N3—N2═O
3	35/35/0	0.000	0.833	C2—N3—N2═O
4	32/32/0	0.000	0.833	O═C2—N3—N2═O
5	42/38/4	0.454	0.833	C2—N3—N2
6	14/11/3	0.750	0.278	c-N3H—C2═O
7	6/5/1	0.650	0.222	cH:c:cH:cH:c-Cl
8	4/4/0	0.000	0.167	O—C2(═O)—N3—N2═O
9	5/5/0	0.000	0.167	cH:c(—N3H—C2═O):cH:cH:c-Cl
10	2/2/0	0.000	0.111	cH:cH:cH:c—O—C2(═O)—N3(—C3H$_3$)—N2═O
11	12/11/1	0.414	0.111	c:cH:c-N3H
...	...	...	...	...

[a]The atoms shown with a "*" stand for any non-H atom. [b]Each listed fragment's distribution in the chemicals of the *whole* training set. T/A/I stands for number of total, active, and inactive training set chemicals. [c]Frequency of occurrence for each listed fragment in the subset of the 18 active training set chemicals that contain the alert A7E_197. [d]This is only a partial list of the fragments that constitutes the applicability domain for alert A7E_197. They are sorted in descending value of their frequency. A number next to a heavy atom denotes its hybridization, e.g. C2 is an sp2 carbon. A lowercase heavy atom symbol (e.g., "c") is for aromatic atoms.

alert is present in 25 training set chemicals (18 active and 7 inactive), and Table 1 lists some of the fragments that are part of the domain of applicability of this alert in descending order of their frequency of occurrence in the 18 active chemicals. It is evident that about 83% of the active chemicals that contain this alert also have the *N*-Nitroso fragment which is well-known for being related to mutagenicity.[23]

Since a CASE Ultra model contains several positive alerts, separate fragment domains are created for each alert and stored with the model.

Domain Adherence of an Alert Identified in a Test Molecule. The objective of having an applicability domain for a positive alert is to determine its validity (in quantitative terms) when identified in a test chemical. In other words, to check if the test chemical belongs to the "activity relevant" chemical space originally defined by the training set member chemicals of the positive alert in question.

If a positive alert is identified in a test chemical, the domain adherence of this alert is calculated by checking all the fragments that constitute its applicability domain for their presence in the test chemical. The adherence value is the ratio of the sum of the squared frequency of occurrence values of the subset of the fragments that are present in the test chemical and sum of the squared frequency of occurrence of all the fragments that constitute the domain of the positive alert in question.

$$\text{domain adherence of a positive alert in a test chemical} = \frac{\sum_{i=1}^{N}(T_i \mathrm{Fr\_tr}_i^2)}{\sum_{i=1}^{N}(\mathrm{Fr\_tr}_i^2)} \quad (2)$$

where $\mathrm{Fr\_tr}_i$ is the frequency of occurrence of the ith fragment as listed in the positive alert's applicability domain in the training set. N is the total number of fragments listed in the applicability domain of the particular positive alert in the training set. T_i is a Boolean variable, and it assumes a value of 1 if the ith domain fragment is found in the test chemical and 0, otherwise. Equation 2 can be derived from the Tanimoto similarity measure used by Sutherland et al.[34] for measuring similarity between two protein targets using the fragment frequencies of small molecules that bind to them and has the following form:

$$\mathrm{FragSim}_{1,2} = \frac{\sum_{i=1}^{N}\mathrm{Fr}_{1,i}\mathrm{Fr}_{2,i}}{\sum_{i=1}^{N}(\mathrm{Fr}_{1,i}^2 + \mathrm{Fr}_{2,i}^2 - \mathrm{Fr}_{1,i}\mathrm{Fr}_{2,i})} \quad (3)$$

where $\mathrm{Fr}_{1,i}$ is the frequency of the ith fragment among binders of protein 1. In our case we have used the frequency of occurrence of a fragment as listed in the applicability domain of a positive alert. When the ith domain fragment is not found in the test chemical, then the numerator in eq 3 becomes zero and in the denominator the only term remaining is the square of its frequency; if the ith domain fragment is found in the test chemical, then the square of its frequency is the term in both the numerator and denominator. The result is identical to eq 2.

In the case when a test chemical contains all the fragments of the domain of the positive alert along with the positive alert itself, eq 2 produces an adherence value of 1.0, and for fewer fragments, the adherence value is less than 1.0. However, the value never becomes 0 because the positive alert itself is part of its own domain. This is done keeping in view that some alerts are very strongly related to activity (e.g., nitrogen mustard group >N−CH2−CH2−Cl to mutagenicity) and their applicability domains usually contain very few fragments. In such cases, if the alert is not included in its own domain, and none of its few other domain fragments appear in the test chemical, the calculated domain adherence value will be zero and the prediction will be wrong in many cases.

Tuning the Predictive Ability of a Positive Alert Using Its Domain Adherence in Chemicals. Before explaining the concept of applying domain adherence cutoffs for a multiple alert model, it is imperative to explain it just for a single alert. Since we can now calculate the domain adherence of a positive alert identified in a test chemical, a cutoff can be enforced on the domain adherence value, below which a positive alert will be disqualified. In other words, the ability of a positive alert to differentiate between active and inactive chemicals can be altered by varying the domain adherence cutoff value. In Figure 1, an alert (ID: A7B_7) was used as an example from the

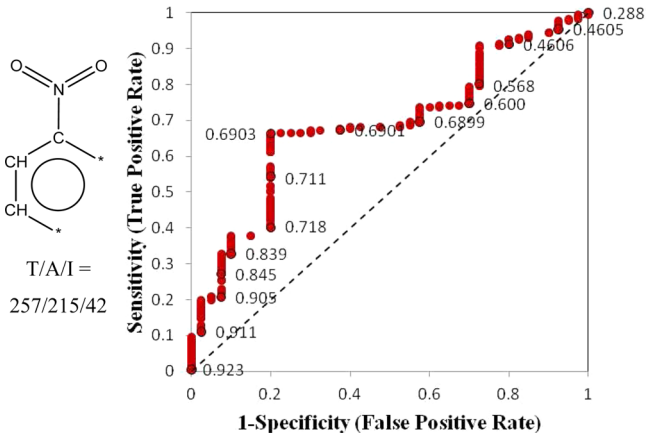


Figure 1. ROC profile of an alert (ID: A7B_7) from the *Salmonella* mutagenicity model A7B. The aromatic ring is left incomplete to include both 5 and 6 member rings, and the ring atoms that are not explicitly shown can be any atom with or without substitution. The domain adherence cutoff values are shown next to the ROC points. Only a few values are shown for clarity.

Salmonella mutagenicity model A7B to demonstrate this. This alert primarily represents the aromatic nitro group and is a well-known mutagenic alert.[23] It was found in a subset of 257 training set chemicals of A7B and consists of 215 active chemicals and 42 inactive chemicals and is highly statistically significant.

The domain adherence of this alert can be calculated for every training set chemical that contains this alert using the method described in the preceding section. The adherence values of this alert in its 257 member chemicals vary from 0.288 to 0.923. If we predict mutagenic potential of these 257 chemicals only using this alert, we can construct a ROC curve by varying the acceptable domain adherence (cutoff) between 0.288 and 0.923. For any cutoff value, the chemicals in which the alert's domain adherence is below the cutoff value are labeled as inactive and chemicals in which the adherence is on or above the cutoff value are labeled as active. Comparing these labels with their experimental activity gives a series of sensitivity and specificity measures which are plotted as the ROC curve shown in Figure 1 (if some marginally active "M" training chemicals also contain this alert, they are excluded from the sensitivity and specificity computations).

Using a Domain Adherence Cutoff for the Whole Model. As mentioned in the Methods section, a CASE Ultra model usually contains multiple positive and deactivating alerts, and identification or absence of these alerts in a test chemical dictate the final prediction call. It is quite normal to find more than one positive alert in a single test chemical, and separate domain adherence values are calculated for each identified positive alert using eq 2. The domain adherence value of an alert in a test chemical indicates the alert's suitability in predicting the activity of the test chemical. As explained earlier, this value depends on the match between the alert's fragment based applicability domain and the structural features of the test molecule. Although each positive alert has a separate domain originally defined by the training set, the magnitude of their domain adherence value in a test chemical has a simple meaning, i.e. a domain adherence value close to zero means the positive alert is not suitable for predicting activity of that chemical and an alert with a value close to one can be used for prediction. Therefore, a single domain adherence value can be associated with a model that will determine the model's predictive behavior. A high cutoff (closer to 1) will produce an increased number of inactive calls as a result of the disqualification of positive alerts identified in many of the test chemicals, whereas a cutoff value closer to 0 will allow the positive alerts to remain effective in many of the test chemicals and reduce the number of inactive calls (i.e., the cutoff can be varied to produce ROC curves). The optimal value for such a cutoff can be determined from the ROC curves generated either during model validation or from external validation as will be explained in the following sections.

4. RESULTS AND DISCUSSIONS

A methodology was developed for defining fragment based applicability domains for individual positive alerts of CASE Ultra models, and for a model as such, a straightforward way was devised for measuring the validity of a positive alert identified in a test chemical during prediction by computing its domain adherence. The following are some immediate advantages:

1. Optimization of the Predictive Performance of a CASE Ultra Model. The cutoff values imposed on the domain adherence of alerts in a set of test chemicals can also serve as an effective way to obtain a ROC (receiver operating characteristic[35]) profile for any CASE Ultra model and can help optimize its predictive performance. For each cutoff value, a check is made through the entire test chemical set and if in any test chemical an identified alert has a domain adherence value that is lower than the cutoff, the alert is excluded from contributing to the activity prediction of that chemical. Thus a series of sensitivity and specificity values are obtained for the test set, which ranges from high sensitivity and low specificity to low sensitivity and high specificity.

It is important to mention here that increasing the domain adherence cutoff value does not result in an increase in the number of "out of domain" calls during prediction, it only results in more positive alerts being disqualified in the test chemicals and thus increase in inactive prediction calls. The prediction coverage (percentage of test chemicals with either active or inactive calls) does not fall increasingly while building the ROC curve. However, as mentioned earlier, the presence of more than three unknown structural fragments in a test chemical results in an out of domain call.

The optimization of model performance can be achieved using one of the following two approaches:

i. Performance Optimization Using Model Validation Results. Figure 2 shows the ROC plot for the fungal composite mutagenicity model A7E. A 10 times 10% out validation procedure was performed on the model, and the test results for the excluded chemicals on each validation cycle was used for the ROC plot (average test coverage = 82%, marginally active

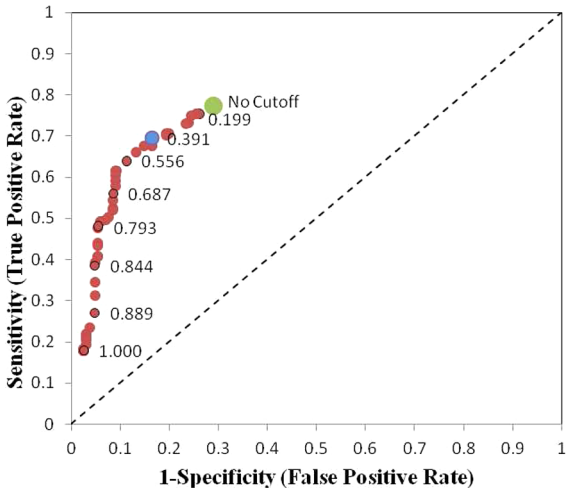


Figure 2. ROC profile of the fungal composite mutagenicity model A7E estimated from the 10 times 10% out validation procedure. The green circle represents sensitivity and specificity when no domain adherence cutoff was applied. The blue circle represents the ROC point that is closest to the perfect classification (0,1). The domain adherence cutoff values are shown next to the ROC points. Only a few values are shown for clarity.

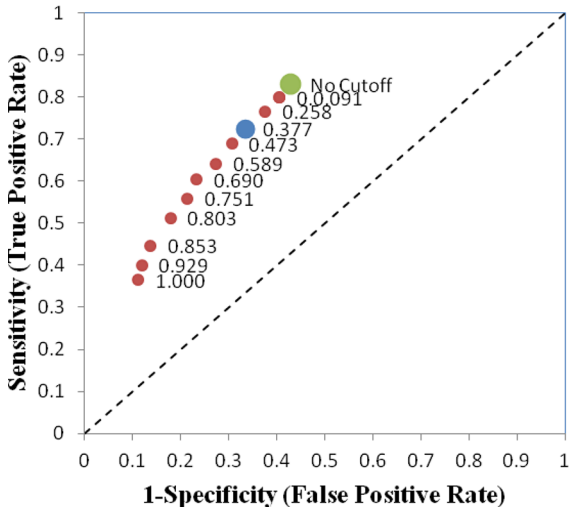


Figure 3. ROC plot obtained from the external test of 4072 chemicals against the *Salmonella* mutagenicity model A7B. The green circle corresponds to the situation when no domain adherence condition was applied. The blue circle represents the ROC point that is closest to the perfect classification point of (0,1). The domain adherence cutoff values are shown next to the ROC points.

chemicals were not included). In this plot, the false positive rate (1 − specificity) is plotted against the true positive rate (sensitivity). Each point is an average of the sensitivity and "1 − specificity" values from the 10 cycles of the validation process. In the plot, the point (0,1) represents a perfect classifier, at (0,0) all test chemicals are predicted as inactive and at (1,1) all chemicals are predicted as active. ROC profiles of CASE Ultra models usually do not cover the higher end of sensitivity values because while we can exclude positive alerts by increasing the domain adherence cutoff values, new positive alerts are not created by relaxing the cutoff.

The ROC profile shown in Figure 2 can be used to choose the domain adherence cut off value that gives the optimum sensitivity and specificity measures. For the purpose of this study we have used the distance of the ROC point from the perfect classification (0,1) point as the criterion to decide which ROC point represents the best sensitivity and specificity. The closer the ROC point is to the (0,1) point, the better it is with regard to classification. In Figure 2, the blue circle was found to be the closest to (0,1) and corresponds to sensitivity = 69.5% and specificity = 83.6%. The domain cutoff value for the blue point is 0.391. For comparison, if we completely disregard applicability domain adherences of the positive alerts in the test chemicals, we get the sensitivity = 77.3% and specificity = 71.1% which it is shown as the green circle in Figure 2. Since the blue circle is closer to the perfect classification point as compared to the green circle, it represents better classification power for the model. The domain adherence cutoff value corresponding to this point can be used for future toxicity prediction of external chemicals.

ii. Performance Optimization Using External Test Sets. Predictive performance of a CASE Ultra model can also be optimized in a similar way if an external test set is available. As an example, we tested the Hansen's external test set of 4072 chemicals against the *Salmonella* mutagenicity model A7B. The ROC plot is shown in Figure 3 (average test coverage = 85%, marginally active chemicals were not included). Again, the green circle corresponds to disregarding applicability domain adherences and gave sensitivity and specificity values of 83.04% and 57.01%, respectively (high sensitivity but rather low specificity). Among all the ROC points, the point shown as the blue circle is closest to the perfect classification point (0,1) and gives sensitivity and specificity values of 72.5% and 66.4%, respectively. The cutoff value corresponding to this point is 0.377.

Results of the predictive performance optimization using these two methods are summarized in Table 2. It is important to note here that we have used the distance of the ROC point from the perfect classification point (0,1) as the criterion to select the best point as an example. However, there is no perfect way of describing the results of a binary classification by a single number and therefore one can use some other suitable criteria to decide which cut off value is best depending on one's particular requirements.

2. Reducing False Positive Predictions Caused by *Generic Alerts*. The applicability domain of the positive alerts can also be used to minimize false positive predictions caused by alerts that have high statistical significance but are not mechanistically related to the end point being modeled. In the case when a training set contains different distinct groups of chemicals and each group has a different reason for activity, the learning algorithm sometimes finds a fragment as a positive alert that is usually generic, small, and common to all such chemicals instead of identifying distinct positive alerts for each group. Unfortunately, this situation is rather unavoidable due to our reliance on the statistical measures used in the learning algorithm. This type of alerts can significantly reduce the performance of a model by producing false positives during prediction. This situation is evident in the CASE Ultra model A7H for the mutagenicity end point of *Drosophila* sex-linked recessive lethal assay. In this model, CASE Ultra identified a total of 51 positive alerts explaining the reasons of mutagenicity in the training set chemicals. The topmost positive alert >N—CH_2— (alert ID: A7H_1) covers 114 training set chemicals (87 active and 27 inactive member chemicals). The alert is shown in Table 3. It is evident that it is a generic fragment, and

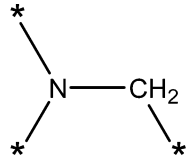
dx.doi.org/10.1021/ci300111r | J. Chem. Inf. Model. 2012, 52, 2609−2618

Table 2. Results of Performance Optimization of CASE Ultra Models A7E and A7B Utilizing Applicability Domain of Positive Alerts

	test statistics without using alert domain adherence[a]				test statistics optimized using alert domain adherence[a]			
method	sens	spec	MCC	χ^2	sens	spec	MCC	χ^2
10 times 10% out validations on model A7E	0.773	0.711	0.479	10.526	0.695	0.836	0.529	13.028
test of 4072 external chemicals with model A7B	0.830	0.570	0.418	568.870	0.725	0.664	0.388	507.443

[a]Sens = sensitivity; spec = specificity; MCC = Matthews correlation coefficient; χ^2 = Pearson's χ-squared statistics.

Table 3. Partial List of Fragments That Form the Domain of Applicability of a Generic Positive Alert (ID: A7H_1), Taken from the *Drosophila* Sex-Linked Recessive Lethal Mutagenicity Assay Model A7H



Alert ID: A7H_1[a] (T/A/I = 114/87/27)

no.	T/A/I[b]	Shannon's entropy	frequency[c]	domain fragments[d]
1	45/41/4	0.433	0.402	N3—N2═O
2	50/43/7	0.584	0.402	N2═O
3	60/49/11	0.687	0.391	$C3H_2$—Cl
4	35/34/1	0.187	0.391	N3—$C3H_2$—$C3H_2$—Cl
5	41/30/11	0.839	0.207	N3—C2═O
6	22/21/1	0.267	0.195	C2—N3—N2
7	21/17/4	0.702	0.195	$C3H_2$—$C3H_2$—N3-c
8	32/22/10	0.896	0.172	$C3H_2$—N3H
9	18/18/0	0.000	0.161	C2—N3—N2═O
10	16/16/0	0.000	0.149	O═C2—N3—N2═O
11	13/13/0	0.000	0.138	N3H—C2—N3—N2═O
12	11/11/0	0.000	0.126	N3H—C2(═O)—N3(—$C3H_2$—$C3H_2$—Cl)—N2═O
13	15/13/2	0.567	0.103	C2—$C3H_2$—O
14	11/11/0	0.000	0.103	$C3H_2$—C2H═C2(—C3H)—$C3H_2$—O—C2═O
...	...	...	...	...

[a]The atoms shown with a "*" stand for any non-H atom. [b]Each listed fragment's distribution in the chemicals of the *whole* training set, T/A/I stands for number of total, active, and inactive training set chemicals. [c]Frequency of occurrence for each listed fragment in the subset of the 87 active training set chemicals that contain the alert A7H_1. [d]This is only a partial list of the fragments that constitutes the applicability domain for alert A7H_1. They are sorted in descending value of their *frequency*. A number next to a heavy atom denotes its hybridization, e.g. C2 is an sp2 carbon. A lowercase heavy atom symbol (e.g., "c") is for aromatic atoms.

if we examine its member active chemicals, then it becomes clear that they can be subdivided into different groups based on their predominant structural feature for which there is plenty of evidence for being directly related to mutagenicity. The 87 member active chemicals of this alert can be divided mainly into two distinct categories: (1) chemicals with >N—CH_2—CH_2—Cl group (nitrogen mustard alkylating agents[23]) and (2) chemicals with >N—N═O (*N*-nitroso group[23]).

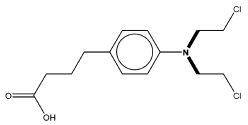
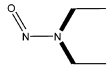
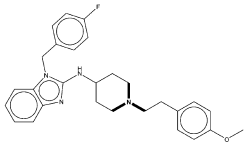
If we do not apply the applicability domain concept to the alert A7H_1, the model will predict any tertiary amine with at least one −CH_2− group connected to the nitrogen to be mutagenic and as a result the number of false positive prediction will increase significantly if the model is being used for testing large sets of diverse chemicals. Although there is evidence that some tertiary amines undergo chemical transformation in the acidic environment of stomach to form carcinogenic dialkylnitrosamines,[36] it happens only in the presence of diets containing nitrites (preserved meat or fish, or crops grown on mineral-deficient soil).

The majority of the training set chemicals that contain the alert A7H_1 are active either because of a nitrogen mustard group or a *N*-nitroso group. If we examine the fragments that constitutes the applicability domain of the alert A7H_1, we can see that the domain accurately captured the importance of nitrogen mustard and *N*-nitroso groups. It is important to note here that the domain of this alert contains a total of 107 fragments but only 14 fragments occur in 10% or more active member chemicals of this alert. These 14 fragments are shown in Table 3 along with their frequency of occurrence.

It is evident from Table 3 that the *N*-nitroso group (no. 1) and the nitrogen mustard group (no. 4) are present in 40.2% and 39.1% of active member chemicals, respectively, and they dominate the chemistry space of the alert A7H_1. If alert A7H_1 is identified in a test chemical, then the domain adherence score of this alert will depend largely on the presence or absence of the nitrogen mustard group or *N*-nitroso group in the test chemical. Alert A7H_1 will receive a relatively lower domain adherence score in a test chemical that contains neither a nitrogen mustard nor an *N*-nitroso group. In Table 4, we have listed three chemicals that are training set member chemicals of alert A7H_1 (with known activity) to illustrate this issue.

dx.doi.org/10.1021/ci300111r | J. Chem. Inf. Model. 2012, 52, 2609−2618

Table 4. Some Chemicals That Contain Alert A7H_1 along with the Domain Adherence Values and Experimental Activity

Name	Mutagenicity	Structure[a]	Domain Adherence score
Chlorambucil	Active		0.609
Diethylnitrosamine	Active		0.660
Astemizole	Inactive		0.442

[a]Only the bond between the two nonstarred atoms of the alert A7H_1 is shown as thick lines, and multiple occurrences of the alert are overlaid. The alert occurs twice in chlorambucil and diethylnitrosamine and three times in astemizole.

3. Minimizing the Negative Impact of a Training Set's Active/Inactive Chemical Ratio on the Model's Predictive Performance. As mentioned earlier, each CASE Ultra model is built using a training set containing examples of active and inactive chemicals. The ratio of the training set's active to inactive chemicals (A/I ratio) has significant effect on the model's predictive performance. If the A/I ratio is small, then the model gets biased toward higher specificity, and in the case of higher A/I ratios, the model usually produces higher sensitivity. However, using the domain adherence cutoff adjustments, we can significantly reduce the severity of this effect. To demonstrate this, 300 active and 600 inactive chemicals (total of 900 chemicals) were randomly picked from the training set of the model A7B to be used for making three new smaller learning sets. The remaining 2635 chemicals were set aside to be used as a test set. Three models with varying A/I ratio were built from the set of 900 chemicals: A7B_SMALL_1_2 (A/I = 300/600), A7B_SMALL_1_1 (A/I = 300/300), and A7B_SMALL_2_1 (A/I = 300/150). The active chemical portions of the three models were kept identical. The mutagenicity of the 2635 chemicals of the test set were predicted independently by the three models. ROC curves were plotted using the test results and is shown in Figure 4 (marginally active chemicals in the test set were not included in the sensitivity/specificity calculations).

The solid green circles in Figure 4 represent sensitivity and specificity of the test results from the three models when domain adherences of alerts in the test chemicals were disregarded. It is evident that in the absence of the applicability domains, the sensitivity/specificity of the models are very sensitive to the A/I ratio and the predictive performance of A7B_SMALL_1_2 and A7B_SMALL_2_1 are significantly different from A7B_SMALL_1_1 which has a balanced A/I ratio. Models with higher A/I ratio, gives higher sensitivity and lower specificity. However, the use of the domain adherence cutoffs offer a possibility to reduce the severity of this effect to some extent. The three black circles on the ROC curves represent the optimal points (from the standpoint of closeness to the ideal classification point) for each model. For the models A7B_SMALL_1_2 and A7B_SMALL_1_1, the optimal points are very close to the ROC points that correspond to the absence of domain adherence (green circles). However, for the model A7B_SMALL_2_1 (A/I = 2.00), the optimal point is the sixth point on the ROC curve and much closer to the optimal point of the model A7B_SMALL_1_1 (A/I = 1.00). This is an example of how the domain adherence cutoff can be changed to adjust the predictive performance of a model built from a training set that has much more active chemicals than inactive chemicals and bring it closer to the performance of a model with balanced A/I ratio. Although the performance of the A7B_SMALL_1_2 (A/I = 0.5) cannot be adjusted to match that of A7B_SMALL_1_1 using domain adherence cutoff, it is not a significant bottleneck because the model can be easily rebuilt after removing half of the inactive chemicals from its training set to give it a balanced A/I ratio. On the other hand, removal of active chemicals from A7B_SMALL_2_1 is not an option because that would result in a model with diminished ability to identify reasons for activity.

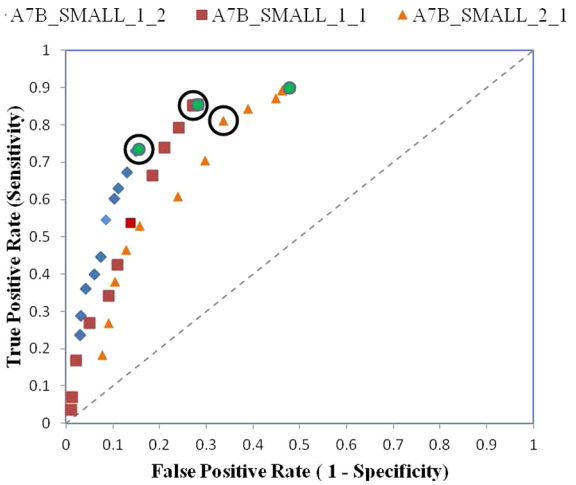


Figure 4. ROC curves from the test results of mutagenicity (*Salmonella*) predictions in 2635 test chemicals using three CASE Ultra models with significantly different active/inactive ratios (as described in the text). The green circles represent prediction statistics when no domain adherence check was performed. The optimal ROC points are highlighted using black circles.

4. Comparison with Classical Chemical Similarity Measure for Building Applicability Domain of Positive Alerts. Classical fragment based chemical similarity measure using fragment based bit-strings or fingerprints[37] can also be used to build the applicability domains of the positive alerts. Fingerprints are bit vectors, and each element of the vector is set to 1 or 0 to indicate the presence of absence of structural fragments in the chemicals being compared. We implemented a fragment based similarity measurement tool in CASE Ultra to compare it with the current approach. This utility measures the Tanimoto similarity between test chemicals and the training set chemicals. Two to four sized linear fragments were used to calculate the similarity between two chemicals A and B using the following formula for Tanimoto similarity for bit vectors:

$$\text{Tanimoto similarity} = nAB/(nA + nB - nAB) \quad (4)$$

where, nA is number of fragments in chemical A, nB is the number of fragments in chemical B, and nAB is the number of fragments that are common between A and B. If a positive alert is identified in a test molecule, this formula was used to compute an array of similarity values. The array consists of similarity measurements computed between the test molecule

and the training set chemicals from which the positive alert was originally derived. If the minimum similarity value in this array is lower than the allowed similarity cut off threshold, the positive alert is disqualified. Figure 5 shows the ROC curves

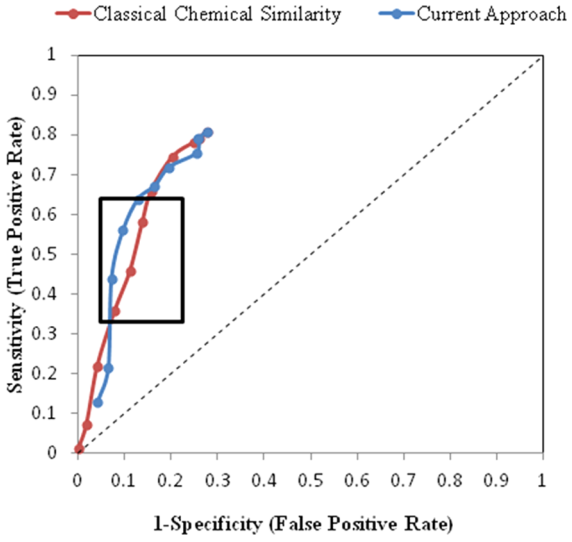


Figure 5. ROC profile of the fungal composite mutagenicity model A7E estimated from the 10 times 10% out validation procedure using the current approach (blue) and classical similarity (red) for estimating the domain adherence. The cutoff value for each approach was varied between 0 and 1 with 0.1 intervals. The black rectangle shows the region where the current approach achieves higher specificity as compared to the classical similarity while achieving same sensitivity.

generated by a 10 times 10% out cross validation runs for the fungal composite mutagenicity model A7E using the current approach (blue line) and the classical similarity measure (red line). For each approach, the cutoff threshold was varied between 0 to 1 with the intervals of 0.1 resulting in 11 points in each curve (marginally active chemicals were not included in the sensitivity/specificity calculations). It is apparent from Figure 5 that within a limited region of the ROC curve, the current approach can achieve higher specificity for the same level of sensitivity than the classical similarity approach. The corresponding region is highlighted using a black rectangle in Figure 5. The similarity based approach is better near the top of the ROC curves, and there is no clear winner from a overall performance point of view.

However, the most important difference between the approach adopted in this work and the classical fingerprint based chemical similarity measure is that of the mechanistic interpretation of the predictions. In the current work, a fragment domain is being constructed for each positive alert using fragments that are statistically related to activity. These domains are stored with the model and used for computing domain adherence of positive alerts when they are identified in test chemicals. If there are 80 positive alerts in a particular CASE Ultra model, there will be 80 such domains available. These domains are valuable source of information as shown in the examples in Tables 1 and 3. Each fragment is paired with its frequency of occurrence and its relationship with activity (Shannon's entropy). A domain gives information about the activity related structural features of the training set chemicals and sometimes provides valuable insight into the mechanistic details. In case of a positive alert being disqualified in a test molecule for poor domain adherence, it becomes immediately apparent which domain fragments are absent in the test chemical and as a result decreasing the domain adherence. In contrast, classical similarity based approach does not require building a domain for individual alerts since the method is based on a series of pairwise similarity measurements between the test molecule and the training set chemicals that contains the positive alert. If a positive alert is disqualified because of lack of acceptable similarity with any of the training set chemicals, it becomes difficult to pinpoint the reason for the disqualification and the methodology becomes more of a "black box".

5. CONCLUSIONS

In this study, a method was developed to construct the applicability domains of individual toxicity alerts of the CASE Ultra expert system models. These domains quantitatively define the necessary structural environment of the toxicity alerts within which the alerts can be considered as valid. By imposing cutoffs on the acceptable domain adherence of positive alerts identified in the test chemicals we can explore the trade-off between true positive and false positive prediction rates of a CASE Ultra model (ROC profile) and optimize the performance of the model. This was demonstrated by applying domain adherence cutoffs on a 10 times 10% out validation results on the fungal composite mutagenicity model and also with the aid of an external test set of 4072 chemicals for prediction of *Salmonella* mutagenicity using the model A7B. The ability of the defined domains to reduce false positive predictions caused by generic toxicity alerts was also shown. The applicability domains can be further utilized to reduce drastic effects of the active/inactive ratios of the training sets on the predictive performance of the models. This study has already resulted in significant enhancement in the predictive abilities of the CASE Ultra expert system and may potentially benefit in silico toxicity assessment of chemicals and drugs in general.

AUTHOR INFORMATION

Corresponding Author

*E-mail: chakravarti@multicase.com. Phone: 216-831-3740. Fax: 216-831-3742.

Notes

The authors are employed by MultiCASE Inc.

REFERENCES

(1) DSSTox. http://www.epa.gov/ncct/dsstox/ (accessed June 22, 2012).

(2) ChemIDplus Advanced. http://chem.sis.nlm.nih.gov/chemidplus/ (accessed June 22, 2012).

(3) REACH. http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm (accessed June 22, 2012).

(4) Kruhlak, N. L.; Benz, R. D.; Zhou, H.; Colatsky, T. J. (Q)SAR modeling and safety assessment in regulatory review. *Clin. Pharmacol. Ther.* **2012**, *91*, 529−534.

(5) Greene, N. Computer systems for the prediction of toxicity: an update. *Adv. Drug Delivery Rev.* **2002**, *54*, 417−431.

(6) Dearden, J. C. In silico prediction of drug toxicity. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 119−127.

(7) Mahadevan, B.; Snyder, R. D.; Waters, M. D.; Daniel Benz, R.; Kemper, R. A.; Tice, R. R.; Richard, A. M. Genetic toxicology in the 21st century: reflections and future directions. *Environ. Mol. Mutagen.* **2011**, *52*, 339−354.

(8) Saiakhov, R. D.; Klopman, G. Benchmark performance of MultiCASE Inc. software in Ames mutagenicity set. *J. Chem. Inf. Model.* **2010**, *50*, 1521−1521.
(9) Grant, S. G.; Zhang, Y. P.; Klopman, G.; Rosenkranz, H. S. Modeling the mouse lymphoma forward mutational assay: the GeneTox program database. *Mutat. Res.* **2000**, *465*, 201−229.
(10) Matthews, E. J.; Kruhlak, N. L.; Daniel Benz, R.; Contrera, J. F. A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. *Regul. Toxicol. Pharmacol.* **2007**, *47*, 115−135.
(11) Matthews, E. J.; Kruhlak, N. L.; Daniel Benz, R.; Ivanov, J.; Klopman, G.; Contrera, J. F. A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. *Regul. Toxicol. Pharmacol.* **2007**, *47*, 136−155.
(12) Marchant, C. A.; Fisk, L.; Note, R. R.; Patel, M. L.; Suárez, D. An expert system approach to the assessment of hepatotoxic potential. *Chem. Biodiversity* **2009**, *6*, 2107−2114.
(13) Dearden, J. C.; Barratt, M. D.; Benigni, R.; Bristol, D. W.; Combes, R. D.; Cronin, M. T. D.; Judson, P. N.; Payne, M. P.; Richard, A. M.; Tichý, M.; Worth, A. P.; Yourick, J. J. The development and validation of expert systems for predicting toxicity. *ATLA, Altern. Lab. Anim.* **1997**, *25*, 223−252.
(14) Japertas, P.; Didziapetris, R.; Petrauskas, A. Fragmental methods in the design of new compounds. applications of the advanced algorithm builder. *Quant. Struct.−Act. Relat.* **2002**, *21*, 23−37.
(15) Klopman, G. MULTICASE 1. A hierarchical computer automated structure evaluation program. *Quant. Struct.−Act. Relat.* **1992**, *11*, 176−184.
(16) *DEREK for Windows*, version 10.0.2 Service Pack 3; Lhasa Ltd.: Leeds, UK, 2007.
(17) Cross, K. P.; Myatt, G.; Yang, C.; Fligner, M. F.; Verducci, J. S.; Blower, P. E. Finding discriminating structural features by reassembling common building blocks. *J. Med. Chem.* **2003**, *46*, 4770−4775.
(18) Klopman, G. Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315−7321.
(19) Klopman, G.; Chakravarti, S. K.; Zhu, H.; Ivanov, J. M.; Saiakhov, R. D. ESP: A method to predict toxicity and pharmacological properties of chemicals using multiple MCASE databases. *J. Chem. Inf. Comput. Sci.* **2004**, *41*, 671−678.
(20) *CaseTox*, version 2.4.0.8; Multicase Inc.: Beachwood, OH, 2012.
(21) Klopman, G.; Chakravarti, S. K. Structure-activity relationship study of a diverse set of estrogen receptor ligands (I) using MultiCASE expert system. *Chemosphere* **2003**, *51*, 445−449.
(22) Klopman, G.; Chakravarti, S. K. Screening of high production volume chemicals for estrogen receptor binding activity (II) by the MultiCASE expert system. *Chemosphere* **2003**, *51*, 461−468.
(23) Kazius, J.; McGuire, R.; Bursi, R. Derivation and validation of toxicophores for mutagenicity prediction. *J. Med. Chem.* **2005**, *48*, 312−320.
(24) Ellison, C. M.; Sherhod, R.; Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Judson, P. N. Assessment of methods to define the applicability domain of structural alert models. *J. Chem. Inf. Model.* **2011**, *51*, 975−985.
(25) Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benigni, R.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D. W.; Schultz, T. W.; Stanton, D. T.; van de Sandt, J. J. M.; Tong, W. D.; Veith, G.; Yang, C. H. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships - The report and recommendations of ECVAM Workshop 52. *ATLA, Altern. Lab. Anim.* **2005**, *33*, 155−173.
(26) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694−701.
(27) Nikolova-Jeliazkova, N.; Jaworska, J. An approach to determining applicability domains for QSAR group contribution models: An analysis of SRC KOWWIN. *ATLA, Altern. Lab. Anim.* **2005**, *33*, 461−470.
(28) Helma., C. Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. *Mol. Diversity* **2006**, *10*, 147−158.
(29) Hansen, K.; Mika, S.; Schroeter, T.; Sutter, A.; Laak, A.; Steger-Hartmann, T.; Heinrich, N.; Muller, K. Benchmark data set for in silico prediction of Ames mutagenicity. *J. Chem. Inf. Model.* **2009**, *49*, 2077−2081.
(30) Matthews, E. J.; Kruhlak, N. L.; Cimino, M. C.; Daniel Benz, R.; Contrera, J. F. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. *Regul. Toxicol. Pharmacol.* **2006**, *44*, 83−96.
(31) Matthews, E. J.; Kruhlak, N. L.; Cimino, M. C.; Benz, R. D.; Contrera, J. F. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. *Regul. Toxicol. Pharmacol.* **2006**, *44*, 97−110.
(32) CDER RCA. http://www.multicase.com/news/cder_rca.htm (accessed June 22, 2012).
(33) Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379−423.
(34) Sutherland, J. J.; Higgs, R. E.; Watson, I.; Vieth, M. Chemical fragments as foundations for understanding target space and activity prediction. *J. Med. Chem.* **2008**, *51*, 2689−2700.
(35) Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861−874.
(36) Lijinsky, W. Reaction of drugs with nitrous acid as a source of carcinogenic nitrosamines. *Cancer Res.* **1974**, *34*, 255−258.
(37) Nikolova, N.; Jaworska, J. Approaches to measure chemical similarity - a review. *QSAR Comb. Sci.* **2003**, *22*, 1006−1026.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on September 18, 2012, with an incorrect Conflict of Interest statement. The correct version was published ASAP on October 2, 2012.