ehp ENVIRONMENTAL HEALTH PERSPECTIVES

ehponline.org

A Novel Two-step Hierarchical Quantitative Structure Activity Relationship Modeling Workflow for Predicting Acute Toxicity of Chemicals in Rodents

Hao Zhu, Lin Ye, Ann Richard, Alexander Golbraikh, Fred A. Wright, Ivan Rusyn, and Alexander Tropsha

doi: 10.1289/ehp.0800471 (available at http://dx.doi.org/)
Online 9 April 2009

NIEHS
National Institute of Environmental Health Sciences

National Institutes of Health
U.S. Department of Health and Human Services

A Novel Two-step Hierarchical Quantitative Structure Activity Relationship Modeling Workflow for Predicting Acute Toxicity of Chemicals in Rodents

Hao Zhu[1], Lin Ye[1], Ann Richard[2], Alexander Golbraikh[1], Fred A. Wright[3], Ivan Rusyn[4,&], and Alexander Tropsha[1,*,&]

[1]Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599;

[2]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711;

[3]Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina 27599;

[4]Department of Environmental Sciences and Engineering, School of Public Health, University of North Carolina, Chapel Hill, North Carolina 27599;

*, Corresponding author:

Alexander Tropsha

327 Beard Hall, University of North Carolina

Chapel Hill, NC 27599-7360

Telephone: (919) 966-2955, FAX: (919) 966-0204

Email: alex_tropsha@unc.edu

&, Equally contributing senior authors.

Running title: Two-step QSAR for Acute Rodent Toxicity Prediction

Key words: acute toxicity, computational toxicology, IC_{50}, LD_{50}, LOAEL, NOAEL, QSAR

Acknowledgements

We thank Dr. Todd Martin (U.S. EPA), and Drs. Judy Strickland and Marcus Jackson (ILS, Inc) for providing some of the data used in this study. We also thank Dr. Woodrow Setzer (U.S. EPA) for his interest in this study and valuable comments on the moving M-regression method. This work was supported, in part, by grants from NIH (GM076059 and ES005948) and U.S. EPA (RD832720 and RD833825). The research described in this article has not been subjected to each funding agency's peer review and policy review and therefore does not necessarily reflect their views and no official endorsement should be inferred. This manuscript has been reviewed by the U.S. EPA National Center for Computational Toxicology and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. The authors declare no competing financial interests.

List of abbreviations:

CV, cross validation

HTS, high throughput screening

ICCVAM, Interagency Coordinating Committee on the Validation of Alternative Methods

LOAEL, Lowest Observed Adverse Effect Level

LOO, leave-one-out (LOO)

*k*NN, *k* Nearest Neighbors

MAE, Mean Absolute Error

MultiCASE: Multiple Computer-Automated Structure Evaluation

NIH, National Institutes of Health

NIEHS, National Institute of Environmental Health Sciences

NIOSH, National Institute for Occupational Safety and Health

NOAEL, No Observed Adverse Effect Level

NTP, National Toxicology Program

QSAR, Quantitative Structure-Activity Relationships

RTECS: Registry of Toxic Effects of Chemical Substances

TOPKAT, Toxicity Prediction by Komputer Assisted Technology

U. S. EPA, The United State Environmental Protection Agency

ZEBET, German Center for the Documentation and Validation of Alternative Methods

Outline of section headers:

Abstract

Introduction

Methods

- Datasets
- QSAR Modeling Approaches
- Moving M-Regression for Data Partitioning
- Model Validation

Results

- Failure of Conventional QSAR Modeling of Rodent Acute Toxicity
- Data Partitioning Using the Moving M-Regression Approach
- Hierarchical QSAR Modeling of the Partitioned Rodent Toxicity Data
- Stability of the *in vitro* and *in vivo* Moving M-Regression Parameters
- Comparison between the two-step hierarchical LD_{50} QSAR model and TOPKAT

Discussion

Conclusions

References

Tables

Figure Legends

Figures

Abstract

Background: Accurate prediction of *in vivo* toxicity from *in vitro* testing is a challenging problem. Large public-private consortia have been formed with the goal of improving chemical safety assessment by the means of high-throughput screening. **Objective:** A wealth of available biological data requires new computational approaches to link chemical structure, *in vitro* data, and potential adverse health effects. **Methods and Results:** A database containing experimental *in vitro* IC_{50} cytotoxicity values and *in vivo* rodent LD_{50} values for more than 300 chemicals was compiled by the German Center for the Documentation and Validation of Alternative Methods (ZEBET). The application of conventional Quantitative Structure-Activity Relationship (QSAR) modeling approaches to predict mouse or rat acute LD_{50} from chemical descriptors of ZEBET compounds yielded no statistically significant models. The analysis of these data showed no general significant correlation between IC_{50} and LD_{50}. However, a linear IC_{50} vs. LD_{50} correlation could be established for a fraction of compounds. To capitalize on this observation, a novel two-step modeling approach was developed as follows. First, all chemicals are partitioned into two groups based on the relationship between IC_{50} and LD_{50} values: one group is formed by compounds with linear IC_{50} vs. LD_{50} relationship, and another group consists of the remaining compounds. Second, conventional binary classification QSAR models are built to predict the group affiliation based on chemical descriptors only. Third, *k*-Nearest Neighbor continuous QSAR models are developed for each sub-class to predict LD_{50} from chemical descriptors. All models have been extensively validated using special protocols. **Conclusions:** The novelty of this modeling approach is that it uses the relationships between *in vivo* and *in vitro* data only to inform the initial construction of the hierarchical two-step QSAR models. Models resulting from this approach employ chemical descriptors only for external prediction of acute rodents toxicity.

Introduction

Development of accurate and predictive i*n vitro* toxicity testing methods that could be used as alternatives for lengthy and costly *in vivo* experiments has long been an elusive goal for both industry and regulatory agencies (National Research Council 2007). New bold research programs were recently established at the National Toxicology Program (NTP) (Xia et al. 2008), the U.S. Environmental Protection Agency (U.S. EPA) (Dix et al. 2007) and coordinated at the inter-agency level by the US Government (Collins et al. 2008) to address this important challenge in a systematic way. The overall goal of these initiatives is to explore a diverse array of *in vitro* toxicity assays, such as cell-based and cell-free high-throughput screening (HTS) techniques, as well as toxicogenomic technologies, to evaluate the toxic potential of chemicals and prioritize candidates for animal testing. However, the utility of *in vitro* data as indicators of *in vivo* effects will be fully realized only if rigorous correlation between the toxicity of chemicals *in vitro* and *in vivo* can be established (National Research Council 2007; Rabinowitz et al. 2008).

Many previous studies have indicated that the correlation between the *in vitro* toxicity results and animal toxicity test data (such as acute, sub-acute, sub-chronic and chronic rodent toxicity test results) is generally poor. Most notably, in 2001, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) hosted a workshop to assess the relationship between cytotoxicity and rodent acute toxicity for 300+ diverse compounds; the data were compiled by the German Center for the Documentation and Validation of Alternative Methods (ZEBET) (ICCVAM and NICEATM 2001). It was concluded that there is no clear correlation between cytotoxicity (IC_{50}) and acute toxicity (LD_{50}) data in rodents. Similarly, poor correlation was found between *in vitro* cytotoxicity and *in vivo* rodent carcinogenicity, even

when a diverse set of *in vitro* endpoints from high throughput screening was used (Xia et al. 2008; Zhu et al. 2008).

Cheminformatics approaches such as Quantitative Structure-Activity Relationship (QSAR) modeling have been widely used in toxicology (Dearden 2003; Johnson et al. 2004). Several software packages, such as Toxicity Prediction by Komputer Assisted Technology (TOPKAT) (Venkatapathy et al. 2004) and Multiple Computer-Automated Structure Evaluation (MultiCASE) (Matthews et al. 2006), have been developed and actively used by both industry and regulatory agencies. However, existing modeling tools generally do not achieve good external accuracy of prediction for compounds not used in model development, and few successful QSAR models have been successful in predicting *in vivo* toxicity endpoints for diverse sets of environmental compounds (Benigni et al. 2007; Stouch et al. 2003).

There are several possible reasons as to why previous attempts to establish relationships between *in vitro* and *in vivo* toxicity data were largely ineffective. These include, among other factors, inadequate attention paid to the chemical diversity of the compounds used for screening and modeling and consequently, unjustified confidence in the ability of models to extrapolate significantly outside the chemistry space of the training set. Furthermore, the conventional QSAR modeling efforts have been disconnected from the growing efforts to employ *in vitro* screening (i.e., HTS data) to predict *in vivo* outcomes. Recently, we have proposed the use of hybrid chemical-biological descriptors, i.e., a combination of conventional chemical descriptors with HTS profile data regarded as biological descriptors. We have demonstrated that these hybrid descriptors afford QSAR models with significantly higher accuracy of prediction of rodent carcinogenicity *vs*. models using chemical descriptors alone, and much higher accuracy *vs*. models that used biological *in vitro* data alone (Zhu et al. 2008).

These recent studies suggest that the explicit consideration of chemical structure (in the form of chemical descriptors) along with *in vitro* assay data could potentially account for discrepancies between *in vitro* and *in vivo* results and produce more accurate predictive models of *in vivo* toxicity. To validate this hypothesis further, in this study we have used the ZEBET dataset (ICCVAM and NICEATM 2001) for which previous attempts to establish the direct *in vitro/in vivo* correlation proved largely unsuccessful (Freidig et al. 2007). We have observed that chemicals can be partitioned into two classes based on comparison between cytotoxicity and acute toxicity data: (1) those for which the linear *in vitro* (IC_{50}) – *in vivo* (LD_{50}) correlation could be demonstrated; and (2) those that correlate poorly. Furthermore, and of central importance for applying our models to the external set of chemicals for which no *in vitro* data exist, we have built binary QSAR models that could discriminate between compounds in these two classes with reasonable accuracy based on their chemical features alone. Finally, we have established rigorous and externally predictive class-specific QSAR models of rodent acute toxicity measured by LD_{50} values. We show that a two step hierarchical QSAR modeling workflow where compounds are first assigned to a class using binary QSAR models and then their LD_{50} value is predicted using class-specific continuous QSAR models affords accurate prediction of LD_{50} values for compounds not included in the training set. In addition, we show that this two-step model statistical prediction accuracy compares favorably to currently available commercial toxicity predictors. Our studies suggest that the two-step QSAR modeling workflow can improve performance of predictive acute toxicity models for diverse organic compounds and aid in prioritizing compounds for rodent toxicity testing.

Methods

Datasets

The ZEBET database consists of data for 361 chemicals compiled from literature studies and published in a consolidated ICCVAM report (ICCVAM and NICEATM 2001). Every compound in this dataset has at least one cytotoxicity result (IC_{50}) and at least one type of rodent acute toxicity value (rat or mouse LD_{50}). ZEBET criteria to select cytotoxicity data for this dataset were defined as follows: 1) at least two different IC_{50} values were available, either from different cell types or from different cytotoxicity endpoints; 2) only cytotoxicity data obtained with mammalian cells were accepted; 3) cytotoxicity data obtained with hepatocytes were not acceptable; 4) chemical exposure time in the cytotoxicity tests was at least 16 hours. Furthermore, only the results obtained from the following cytotoxicity tests were accepted: 1) cell proliferation measured by cell number, protein, DNA content, DNA synthesis, or colony formation; 2) cell viability and metabolic indicators including MIT-24, MTT, MTS, XTTC; 3) cell viability and membrane indicators, including Neutral Red Uptake, Trypan blue exclusion, cell attachment, and cell detachment; 4) differentiation indicators.

For the purpose of this work, the dataset was curated to select the subset of organic compounds, whereas inorganic and organometallic compounds, as well as compound mixtures, were excluded since conventional chemical descriptors used in QSAR studies could not be computed in these cases. There were 254 and 235 compounds that had rat or mouse LD_{50} (mmol/kg-bodyweight/day) values, respectively. Only LD_{50} values published in the Registry of Toxic Effects of Chemical Substances (RTECS) (Norager et al. 1978; Ruden and Hansson 2003) were used. The distribution of $\log(1/LD_{50})$ values of ZEBET compounds, with the exception of a single outlier, were from -2.61 to 2.30 for the rat and from -2.50 to 2.19 for the mouse. One

compound, 2,3,7,8-tetrachloro-dibenzo-p-dioxin (CAS 1746-01-6), was considered an activity outlier since its $\log(1/LD_{50})$ value was -4.21 for the rat, which deviated significantly from the activity range of the dataset. After excluding this single outlier, the datasets used for modeling were comprised of 253 compounds for the rat and 235 compounds for the mouse. An additional set of 115 compounds with complete data (both LD_{50} and IC_{50}) for the rat was recently released by ICCVAM and used in these studies (referred to as the "ICCVAM dataset"). See Supplemental Material, Table 1 for raw data.

The rat chronic Lowest Observed Adverse Effect Level (LOAEL) and rat chronic No Observed Adverse Effect Level (NOAEL) data were compiled from an internal low-dose toxicity dataset established in our laboratory (See Supplemental Material, Table 2). These data include a combination of multiple toxicity phenotypes, such as liver toxicity, kidney toxicity, etc. In comparison with the ZEBET dataset, there are 42 unique compounds that have both rat LOAEL and *in vitro* IC_{50} values and 41 compounds that have both NOAEL and *in vitro* IC_{50} values. Due to limited availability of LOAEL and NOAEL data, we only used these two datasets to illustrate the data partitioning algorithm but did not build any QSAR models for them.

QSAR Modeling Approaches

We used the *k* Nearest Neighbor (*k*NN) QSAR modeling approach that has been developed in our group (Zheng and Tropsha 2000). In brief, the method is based on the *k* nearest neighbors principle and the variable selection procedure. It employs the leave-one-out (LOO) cross-validation (CV) procedure and a simulated-annealing algorithm for the variable selection. The procedure starts with the random selection of a predefined number of descriptors from all

descriptors. If the number of nearest neighbors k is higher than one, the estimated activities $\hat{y}_i$ of compounds excluded by the LOO procedure are calculated using the following formula:

$$\hat{y}_i = \frac{\sum_{j=1}^{k} y_j w_{ij}}{\sum_{j=1}^{k} w_{ij}} = \frac{\sum_{j=1}^{k} y_j w_{ij}}{k-1} \quad [1]$$

where y_j is the activity of the j-th compound. Weights w_{ij} are defined as:

$$w_{ij} = 1 - \frac{d_{ij}}{\sum_{j'=1}^{k} d_{ij'}} \quad [2]$$

and d_{ij} is Euclidean distances between compound i and its j-th nearest neighbor. The further details of the algorithms and workflow are described elsewhere (Medina-Franco et al. 2005; Ng et al. 2004; Shen et al. 2002; Zheng and Tropsha 2000).

Rat and mouse LD_{50} QSAR models for ZEBET compounds were developed using Dragon chemical descriptors (DRAGON for Windows, v.5.4, 2006, Teleste s.r.l., Milan, Italy). Prior to model construction, 23 compounds with rat LD_{50} and 24 compounds with mouse LD_{50} results were selected at random to serve as external validation sets. The remaining 230 (rat) and 211 (mouse) compounds were used as modeling sets, and each was divided multiple times into training/test sets using the Sphere Exclusion approach (Golbraikh et al. 2003). The statistical significance of the models was characterized with the standard leave-one-out cross-validated (LOO-CV) R^2 (q^2) for the training sets and the conventional R^2 for the test sets when modeling real values data (i.e., continuous QSAR). For classification modeling, we used correct classification rate expressed as a fractional value between zero and one. The model acceptability cutoff values of the LOO-CV accuracy of the training sets and the prediction accuracy for test sets were both set to 0.65 for classification models. For continuous models, the acceptability

thresholds for LOO-CV regression q^2 for the training sets and R^2 values for the test set were both set at 0.5. Models that did not meet both training and test set cutoff criteria were discarded.

Moving M-Regression for Data Partitioning

A novel approach related to a class of M-regression methods (Andersen 2007), that we termed "moving M-regression," was used to select compounds for which there is a strong correlation between IC_{50} and LD_{50} values (Class 1). The approach is a variant of the least squares regression that takes into account only those data points contained within a band around the regression line $y^{regr}=ax+b$. For each y, only the points within the interval $[y-d_i,y+d_i]$ are candidates for Class 1, whereas points outside of this band are excluded from Class 1. If the line $y=ax+b$ is moved, some new points will enter the band, whereas some other points will leave it, which may result in a higher regression R^2; this also explains why we use the term "moving M-regression". For each point, $\{x_i,y_i\}$, i=1,…,n, the moving M-regression inclusion function is defined as

$$\eta(x_i, y_i) = \begin{cases} 1, \text{if } y_i \in [ax_i + b - d_1, ax_i + b + d_2] \\ 0, \text{otherwise} \end{cases} \quad [3]$$

Thus, the moving M-regression line can be found by minimization of the following expression

$$F(a,b) = \sum_{i=1}^{n} \eta(x_i, y_i)(y_i - ax_i - b)^2 . \quad [4]$$

Function F is not differentiable at all points (x_i,y_i) such that $y_i=ax_i+b-d_1$ and $y_i=ax_i+b+d_2$. For practical purposes, $\eta(x_i, y_i)$ is approximated by sums of two sigmoid functions:

$$\eta(x_i, y_i) \sim \frac{1}{2}\left\{\frac{1}{1+\exp[-P_1(y_i - ax_i - b + d_1)]} + \frac{1}{1+\exp[P_2(y_i - ax_i - b - d_2)]}\right\} \quad [5]$$

where P_1 and P_2 are large (~100) positive parameters. Indeed, if P_1 and P_2 are approaching infinity, the expression on the right side of equation [5] is approaching the right side of equation [3]. Small approximation errors in the vicinity of points $\{ax_i+b-d_1,y_i\}$ and $\{ax_i+b+d_2,y_i\}$ are approaching zero with both P_1 and P_2 approaching infinity. It is as if the data points are gradually included within, or excluded from, the band when the regression line is moving. Finally, replacing $\eta(x_i, y_i)$ by expression [3], we obtain

$$F(a,b) = \sum_{i=1}^{n} \frac{1}{2} \left\{ \frac{1}{1+\exp[-P_1(y_i - ax_i - b + d_1)]} + \frac{1}{1+\exp[P_2(y_i - ax_i - b - d_2)]} \right\} (y_i - ax_i - b)^2 . \qquad [6]$$

To optimize $F(a\ b)$, the following system of equations is to be solved:

$$\begin{cases} \dfrac{\partial F}{\partial a} = 0 \\ \dfrac{\partial F}{\partial b} = 0 \end{cases} . \qquad [7]$$

Equations [7] are nonlinear, so depending on the dataset and parameters P_1, P_2, d_1 and d_2, there can be multiple solutions (a,b).

In these studies, x_i and y_i were the *in vitro* $\log(1/IC_{50})$ and *in vivo* $\log(1/LD_{50})$ values, respectively, for a dataset of compounds under study. Instead of using equation [6], the compounds that belong to Class 1 were determined by maximizing the number of data points within the band. With this correction, our target function takes the form:

$$F^*(a,b) = \sum_{i=1}^{n} \left\{ \frac{1}{1+\exp[-P_1(y_i - ax_i - b + d_1)]} + \frac{1}{1+\exp[P_2(y_i - ax_i - b - d_2)]} \right\} \qquad [8]$$

To obtain the baseline toxicity regression (see the Results section), we opted to minimize the number of outliers below the regression line. For this purpose, we added additional terms for the lower border of the band weighted by an arbitrary parameter α. Thus, we minimized the following target function:

$$F^{**}(a,b) = \sum_{i=1}^{n} \left\{ \frac{1+\alpha}{1+\exp[-P_1(y_i - ax_i - b + d_1)]} + \frac{1}{1+\exp[P_2(y_i - ax_i - b - d_2)]} \right\} \quad [9]$$

The initial point (*a,b*) for minimization of F^{**} was selected manually. P_1 and P_2 were equal to 100, d_1 and d_2 were equal to 0.4, and α was equal to 1. To optimize function [9], the system of equations [7] in which F is replaced by F^{**} should be solved. Figure 1 summarizes the data analytical workflow that was employed in this study for rodent acute toxicity modeling.

Model Validation

Training set models were validated by evaluating their external predictive power on the test sets as described above. Furthermore, a 5-fold external cross validation analysis was performed for the original ZEBET dataset: the dataset was randomly split into five equal size subsets of compounds and five independent sets of calculations were conducted each time using 80% of the whole dataset as a modeling set and the remaining 20% compounds as a test set. In addition, robustness of QSAR models was verified using a Y-randomization (randomization of response) approach as follows. The modeling set compounds were randomly divided into Class 1 and Class 2 subsets and *k*NN QSAR LD_{50} models were developed for each subset using the same protocol and the same cutoff criteria (q^2 and R^2>0.5) as for compounds in Classes 1 and 2 that were generated by means of the moving regression. The purpose of this study was to see if statistically significant QSAR models could be obtained for any random division of the original data into two classes. Independently, the test was applied to compounds in unique Classes 1 and Class 2 by randomizing their LD_{50} values and re-developing training set models. Both Y-randomization tests were repeated 10 times.

Results

Failure of Conventional QSAR Modeling of Rodent Acute Toxicity

The modeling set including 230 compounds with known rat LD_{50} data was partitioned into 32 training and test sets and the conventional *k*NN QSAR modeling approach was applied to all training sets as detailed in Materials and Methods. Each training set model was characterized by its q^2 value; for the five best training set models these values ranged between 0.5 and 0.57. These five models were used for predicting LD_{50} values for the respective external validation set (23 compounds). However, for each of these models the R^2 value for this external set was less than 0.5. When other types of in-house or commercial QSAR methods (e.g., Support Vector Machine, Partial Least Square, etc.) and other types of descriptors (e.g., MolConnZ descriptors, Molecular Operating Environment descriptors, etc.) were used, no statistically significant predictive QSAR models could be obtained (data not shown). Likewise, modeling of the mouse dataset (211 modeling compounds and 24 external validation compounds) was unsuccessful (data not shown). This negative result corroborates the well known inability of conventional QSAR modeling approaches to arrive at statistically significant and externally predictive models of *in vivo* toxicity.

Data Partitioning Using the Moving M-Regression Approach

It is well known that *in vitro* cytotoxicity correlates poorly with *in vivo* toxicity endpoints when any relatively large set of compounds is considered. The ZEBET dataset is no exception; the data show that cytotoxicity (IC_{50}) correlates with acute toxicity (LD_{50}) for only a fraction of the compounds in either the rat (Figure 2A) or mouse (Figure 2B) datasets. Most of the compounds are more toxic *in vivo* than *in vitro*. Similar patterns could be found between

cytotoxicity and other *in vivo* toxicity endpoints, e.g., rat chronic LOAEL and rat chronic NOAEL (Figures 2C and 2D).

To devise a mathematical means for identifying compounds with strong *in vitro/in vivo* correlation, we extended concepts that have been previously employed in calculating the "baseline regression" that correlated the aquatic toxicity of (some) chemicals with the logarithm of the n-octanol water partition coefficient (logP) (Klopman et al. 1999; Klopman et al. 2000); this approach is reviewed elsewhere (Mayer and Reichenberg 2006). Here, we have developed a novel approach, termed "moving M-regression," to identify a subset of compounds with strong IC_{50} *vs*. LD_{50} correlation. Using this method, we have partitioned compounds in the modeling set into two classes: Class 1, in which compound's acute toxicity linearly correlates with cytotoxicity; and Class 2, in which acute toxicity does not correlate well with cytotoxicity, with these points positioned above the regression line.

This analysis for the rat ZEBET dataset resulted in 122 compounds assigned to Class 1, i.e. within the linear regression correlation band between LD_{50} and IC_{50} values. The points corresponding to the 93 out of 108 remaining compounds are located above the regression line band and are classified as Class 2 whereas 15 compounds fall below the regression line (cf. Figure 2A). Although these compounds are likely to be activity outliers, in the absence of an objective rationale for their outlier status, we merged them into Class 1 to obtain the highest coverage of the resulting models and to provide a more realistic measure of external predictivity. The correlation between the LD_{50} and IC_{50} values of the resulting 137 Class 1 compounds is shown in equation [10] and in Figure 2A.

$$Log(1/LD50) = -1.1 + 0.4 \times Log(1/IC50) \quad [10]$$

$$R^2 = 0.74, \qquad SE = 0.36, \qquad N = 137$$

This approach was also applied to analyze the relationship between the *in vitro* IC_{50} and other *in vivo* toxicity data, including mouse LD_{50}, rat chronic LOAEL, and rat chronic NOAEL (Table 1). The same trend was found for all datasets, i.e., in all cases the data were partitioned into two classes: (1) points on the baseline; and (2) points off the baseline (Table 1 and Figure 2). The ratio of Class 1 to Class 2 compounds was found to be similar for each of the four *in vitro* - *in vivo* toxicity datasets. This result further supports the generality of the "moving M-regression" approach.

Hierarchical QSAR Modeling of the Partitioned Rodent Toxicity Data

Using class assignments from the data partitioning described above, we employed a two-step QSAR approach (Figure 1) for: (i) classification modeling (i.e., establishing that compounds assigned to Classes 1 and 2 based on their biological activity data could be sub-divided into the same Classes based on their chemical structure), and (ii) predictive continuous modeling for all compounds in each class (i.e., estimation of the LD_{50} based on chemical structure, not IC_{50} data). For ZEBET rat data, three modeling sets were generated: (I) 230 compounds (137 Class 1 *vs*. 93 Class 2) for classification modeling; (II) 137 Class 1 compounds; and (III) 93 Class 2 compounds for developing two continuous rat LD_{50} models. The analysis of these three datasets resulted in 252 classification models, as well as 1,207 continuous LD_{50} models for Class 1 compounds, and 40 continuous LD_{50} models for Class 2 compounds that satisfied the statistical significance threshold criteria. The statistical figures of merit for the best *k*NN models obtained for these three modeling sets are listed in Table 2.

In order to demonstrate that these QSAR models have significant external prediction accuracy, we have employed several concurrent approaches for model validation. First,

following our general model validation workflow (Tropsha and Golbraikh 2007), 23 compounds excluded randomly from the entire dataset were used as an external validation set. The following two-step prediction protocol for external compounds was used: (i) *k*NN classification models were used to assign compounds to Class 1 or Class 2; and (ii) depending on the outcome, the respective class-specific continuous QSAR models was employed to predict the LD_{50} values for each compound. The results demonstrate that the overall accuracy of prediction for this external set is reasonably good. In the first step, the classification model had 65% prediction accuracy (the fraction of correctly identified Class 1 and Class 2 compounds). In the second step, we obtained R^2=0.70, Mean Absolute Error (*MAE)*=0.39 for the prediction coverage (i.e., the fraction of the external set compounds within the applicability domains of the models) of 74% for the external test set when combining the predictions for Class 1 and Class 2 compounds.

Second, a 5-fold external cross validation analysis was performed to test the robustness of the modeling outcome using 253 rat ZEBET compounds. The dataset was randomly split into five equal size subsets of compounds and the modeling procedure was repeated five times using each subset as a test set and the remaining four subsets as training set as detailed in the Methods section. The statistical results of this exercise were as follows: $Slope_{reg}$=0.45±0.01, R^2_{reg}=0.71±0.04, R^2_{ext}=0.55±0.05, *MAE*=0.44±0.04, Coverage=73%±3%.

Third, Y-randomization tests were performed to establish whether our models are statistically robust. Random partitioning of the compounds into two classes (10 times) produced only three (for Class 1) and 28 (for Class 2) models that satisfied the criteria of $q^2/R^2 > 0.5$ compared to 1,207 and 40, respectively, models for "moving M-regression"-assisted partitioning (see above). When LD_{50} data were randomized, no model with $q^2/R^2 > 0.5$ could be generated for Class 1 and Class 2 compounds.

Fourth, we performed additional Y-randomization analyses by randomly moving or rotating the correlation line (including negative correlation) and redefining compounds into Class 1 and 2. The randomly assigned Class 1 and Class 2 sets were used to develop QSAR LD_{50} models individually and the procedure was repeated 10 times. We found that at most very small number (<7) of acceptable (Q^2>0.5, R^2>0.5) models could be developed.

Similar modeling results were obtained using the ZEBET mouse LD_{50} data. After partitioning 211 modeling set compounds into 119 Class 1 compounds and 92 Class 2 compounds, we developed 843 classification models for Class 1 *vs*. Class 2, 236 continuous LD_{50} models for Class 1, and 356 models for Class 2 compounds. A two-step prediction protocol for evaluation of the 24 external compounds resulted in similarly good external prediction accuracy: R^2=0.69, *MAE*=0.42 and prediction coverage of 54%.

As a true external validation challenge we have used our model to make predictions for the 115 compounds with rat LD_{50} data in the new ICCVAM dataset. This dataset was compiled after we finished the development of the above-described QSAR LD_{50} models, so it could be viewed as a true "blind" validation test. The statistical parameters of the prediction results for these compounds were: R^2=0.57, *MAE*=0.48 and prediction coverage of 70%. Although somewhat less accurate than the results of the previous external prediction, this validation reinforces the statistical significance and utility of the model. Y-randomization tests were also performed for the mouse LD_{50} dataset. Similar to the rat data, after 10 random assignments of compounds into the two classes, at most 4 (for Class 1) and 38 (for Class 2) models ($q^2/R^2 > 0.5$) were developed as compared to 843 and 236 models, respectively, when a classification model was used. Randomization of LD_{50} values produced no significant models.

Stability of the in vitro and in vivo Moving M-Regression Parameters

Since the regression correlation between *in vitro* (IC_{50}) and *in vivo* (LD_{50}) data is required to classify the modeling set compounds and, subsequently, to create the *k*NN classification models, this linear correlation is an essential factor to determine the robustness of our final models. Hence, the slope of the correlation and associated correlation coefficient (R^2) should remain stable when new compounds are added into the modeling set. To validate this supposition, we compiled all available ZEBET and ICCVAM compounds with rat LD_{50} data to create a new modeling set including the original modeling set (230 compounds), the external validation set (23 compounds) and additional data (115 compounds). We also included the compounds previously not used for modeling (inorganic, organometallic and mixtures) since no chemical descriptors are used in this validation. Using the moving M-regression approach for all 425 compounds with IC_{50} and LD_{50} values, the resulting *in vitro/in vivo* correlation parameters are similar to those obtained from our original modeling set in equation [10]:

$$Log(1/LD50) = -1.1 + 0.36 \times Log(1/IC50) \qquad [11]$$

$$R^2 = 0.71, \qquad SE = 0.37, \qquad N = 258$$

The proportion of Class 1, Class 2 compounds and outliers among these 425 compounds were also comparable to that of the original modeling set of 230 compounds (Table 1). We conclude that adding new compounds into the modeling set, which should be important to improve the final model by enriching its chemical and biological diversity, does not affect the *in vitro- in vivo* regression statistics.

Comparison between the two-step hierarchical LD_{50} QSAR model and TOPKAT

We compared the performance of our modeling approach with that of the Toxicity Prediction by Komputer Assisted Technology (TOPKAT, v.6.1) software [accelrys.com/products/discovery-studio/toxicology/; (Enslein 1988)]. Two types of comparison were considered. First, we have analyzed 27 out of the 115 ICCVAM compounds that have been used neither for building our model, nor in the TOPKAT LD_{50} training set. Figure 3 shows the correlation between the experimental and predicted LD_{50} values obtained from our model vs. TOPKAT. The R^2 and *MAE* of TOPKAT were 0.16 and 0.78, respectively, for all 27 compounds, which is significantly worse that the same statistical parameters for prediction of the same dataset using our model, i.e., R^2 and *MAE* of 0.64 and 0.38, respectively. For seven compounds that were outside of the applicability domain for our model, the R^2 and *MAE* using TOPKAT were 0.60 and 0.50, respectively, while our model produced the values of 0.86 and 0.29, respectively (Table 3).

Secondly, we have used our models to predict acute toxicity compounds in the RTECS (Norager et al. 1978) dataset (data were kindly provided by Dr. Todd Martin from U.S. EPA), which contains approximately 7,000 compounds with rat LD_{50} data. We removed compounds which were found within the ZEBET dataset, as well as inorganic compounds and mixtures. This procedure produced a library of 4,003 compounds spanning a diverse chemical space of organic molecules for which experimental rat LD_{50} data are available.

Since the size of the RTECS library is much larger than that of our original modeling set we drew from our experience in using QSAR models for virtual screening (Oloff et al. 2005) and narrowed the model applicability domain. Consequently, predictions were made only for compounds that had greater than 70% confidence level in assigning them to either class 1 or 2 in step 1 of our workflow (i.e., we required that greater than 70% of all QSAR models meeting our

acceptability domain criteria would predict a compound in the same class). We determined that there were 1,562 compounds (out of 4,003) that were not included in the training set of TOPKAT rat LD_{50} model and for which predictions could be made based on the aforementioned criteria. The TOPKAT model predicted LD_{50} values for these compounds with an R^2=0.16 and *MAE*=0.78 (Figure 4, Table 3). The same parameters for the two-step QSAR model were 0.26 and 0.65, respectively. After implementing the applicability domain filter, we made predictions for 965 RTECS chemicals. TOPKAT model had the following parameters: R^2=0.22 and *MAE* =0.65. The same parameters for the two-step model were 0.33 and 0.54, respectively (Table 3), which is better than or comparable to prediction accuracy of various commercial QSAR modeling packages (Moore et al. 2003), albeit there is room for improvement.

It should be noted that the prediction accuracy of the two-step model can be improved by applying more strict criteria in the classification step. For instance, a 90% cut-off for correct class prediction results in prediction model statistics of R^2=0.62 and *MAE*=0.42, but the coverage of the model diminishes considerably to include 101 compounds (Table 3). The performance of TOPKAT for the same 101 compounds is poor: R^2=0.26 and *MAE*=0.66. Considering that the TOPKAT LD_{50} training set contains many more compounds (~6,000) than the training set used to develop the two-step model (~200), it is noteworthy that higher prediction accuracy can be achieved using our modeling approach for a much larger dataset. Furthermore, our approach outperforms TOPKAT consistently over a range of error thresholds either for 965 RTECS compounds or for 101 RTECS compounds (Figure 4). In addition, we used Wilcoxon test to calculate the p-values for the differences in MAEs obtained using two methods. Both for the whole set (965 compounds), and the reduced set (101 compounds) the improvement achieved by our method, as compared to TOPKAT, is statistically significant (p-value<0.005).

One obvious reason that the prediction accuracy of our models for RTECS compounds is lower than that obtained from the external validation set of ICCVAM data is due to the different "activity" ranges of compounds in these two datasets. For example, the activity range (log 1/Activity, in mMol units) of ZEBET compounds is from -2.61 to 2.30, whereas the activity range of RTECS compounds is considerably larger, i.e. from -3.34 to 4.21. It should be stressed that the *k*NN method used in our study can not extrapolate in the activity space since external compound activity is predicted by averaging the activities of nearest neighbor compounds in the training set (see Methods). The *MAE* for the prediction of RTECS compounds that have experimental activity above 2 or below -2, is 1.14 log units. On the other hand, the *MAE* for the prediction for RTECS compounds that have experimental activity between -2 and 2 is considerably lower, i.e., 0.52 log units. The likely explanation for the better performance of our models in the latter range is that more than 90% of our modeling set compounds have rat LD_{50} activity in the same range, i.e. between -2 and 2. Increasing the diversity and activity range of compounds in the modeling set should significantly improve the prediction accuracy of our models.

Discussion

The conventional wisdom in mechanistic and regulatory toxicology is that predictions of the *in vivo* toxicity endpoints from *in vitro* measures, even within the same species, are difficult. However, an approximate linear correlation between *in vitro* IC_{50} and rodent LD_{50}, two of the widely acceptable benchmark parameters used for regulatory purposes, can be established for a significant fraction of the compounds. Indeed, we confirmed this notion by quantitative analysis of the IC_{50}/LD_{50} relationships and devised an objective, computational means to partition compounds into two groups, i.e., those having good linear fit within a defined band, or those falling outside the band and exhibiting, for the most part, higher *in vivo* than *in vitro* toxicity. Our hypothesis to explain this observation is that whereas cytotoxicity assays can be reflective of some of the toxicity mechanisms resulting in adverse health effects on the whole animal level, the *in vitro* tests can not fully reproduce the complex mechanisms of the *in vivo* toxicity. For example, it is well known that many compounds are not toxicants themselves but their metabolites are toxic. We argue that the two-step prediction model based on chemical descriptors only that was developed in our studies, also assists in identification of the compound subset that may act directly (i.e., without being biotransformed) and through the mechanisms likely to be predictive of the potential *in vivo* effects. A similar argument was presented previously in ecotoxicity research where logP was found to be a mechanistically-relevant predictor (Verhaar et al. 2000).

To further substantiate this argument, we considered the top 10 chemical fragment descriptors that were used most frequently in statistically significant QSAR models, i.e., descriptors with the highest discriminatory power (See Supplemental Material, Table 3). It is noteworthy that the “primary aromatic amine,” “hydrazine” and “sulfonamide” moieties, found

within compounds that are known to be toxic both *in vitro* and *in vivo* (Alaejos et al. 2008; Carr et al. 1993; Toth 1988), were found predominantly in compounds of Class 1. On the other hand, "pyrrolidine" and "tertiary amine" moieties which require biotransformation (Domagala 1994) were predictors for Class 2. We have also demonstrated that this objective division of the dataset into two major groups affords robust hierarchical QSAR models, an assertion further supported by successive challenges to the models with external datasets, cross validation and randomization of data.

The approach advocated in this study for biologically-informed SAR data partitioning differs from conventional cheminformatics clustering approaches. Traditional methods partition compounds into multiple sub-groups using their chemical structure-based properties only (i.e., chemical descriptors). The underlying reasoning for chemically-based clustering is that similar structures are expected to have similar biological properties and mechanisms of activity. However, it is a well known limitation of SAR that the absence or presence of a functional group or other minor change of the chemical structure may result in a large change of biological activity (Maggiora 2006). In our studies, the conventional chemical structure-based clustering method did not yield any statistically meaningful models, either global or local. The distribution of pairwise chemical similarities for all compounds within the modeling sets (Class 1 vs. Class 2) of rat LD_{50}s using Dragon descriptors is very similar (data not shown). This observation reconfirms that chemical clustering would not have partitioned compounds in a way similar to the biological data-based partitioning.

Conclusions

Although the cytotoxicity data generally show weak correlation with rodent acute toxicity, we have demonstrated that these data can be used to inform and improve QSAR modeling of *in vivo* acute toxicity. We have developed a novel two-step kNN QSAR modeling approach that affords a successful prediction of acute toxicity (LD_{50}) values from chemical structure for both rats and mice. Furthermore, LD_{50} values were predicted for external compounds with accuracy, exceeding that of previously published QSAR models developed with the commercial (TOPKAT) software. It should be stressed that although *in vitro* cytotoxicity data have been used to establish the rules for partitioning the majority of compounds into two classes, the ultimate models, both classification and continuous, employ chemical descriptors only. This vital feature of our approach makes it possible to achieve accurate predictions of rodent acute toxicity directly from chemical structure alone, even bypassing the need for *in vitro* studies of new compounds. We believe that this biological data-based partitioning approach using *in vitro* toxicity data for the modeling set only, coupled with subsequent chemical structure-based classification and continuous QSAR modeling techniques, holds a promise for modeling other complex *in vivo* toxicity endpoints. This approach charts a future course for combining *in vitro* screening methods and QSAR modeling to prioritize chemicals for *in vivo* animal toxicity testing.

References

Alaejos MS, Pino V, Afonso AM. 2008. Metabolism and toxicology of heterocyclic aromatic amines when consumed in diet: Influence of the genetic susceptibility to develop human cancer. A review. Food Res Internat 41:327-340.

Andersen M. 2007. Modern methods for robust regression. Thousand Oaks, CA:Corwin Press.

Benigni R, Netzeva TI, Benfenati E, Bossa C, Franke R, Helma C et al. 2007. The expanding role of predictive toxicology: an update on the (Q)SAR models for mutagens and carcinogens. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 25:53-97.

Carr A, Tindall B, Penny R, Cooper DA. 1993. In vitro cytotoxicity as a marker of hypersensitivity to sulphamethoxazole in patients with HIV. Clin Exp Immunol 94:21-25.

Collins FS, Gray GM, Bucher JR. 2008. Toxicology. Transforming environmental health protection. Science 319:906-907.

Dearden JC. 2003. In silico prediction of drug toxicity. J Comput Aided Mol Des 17:119-127.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95:5-12.

Domagala JM. 1994. Structure-activity and structure-side-effect relationships for the quinolone antibacterials. J Antimicrob Chemother 33:685-706.

Enslein K. 1988. An overview of structure-activity relationships as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation, and acute oral toxicity. Toxicol Ind Health 4:479-498.

Freidig AP, Dekkers S, Verwei M, Zvinavashe E, Bessems JG, van de Sandt JJ. 2007. Development of a QSAR for worst case estimates of acute toxicity of chemically reactive compounds. Toxicol Lett 170:214-222.

Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. 2003. Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 17:241-253.

ICCVAM, NICEATM. Report of the International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity. 01-4499.NIH,2001.

Johnson DE, Smith DA, Park BK. 2004. Linking toxicity and chemistry: think globally, but act locally? Curr Opin Drug Discov Devel 7:33-35.

Klopman G, Saiakhov R, Rosenkranz HS. 2000. Multiple computer-automated structure evaluation study of aquatic toxicity II. Fathead minnow. Environ Toxicol Chem 19:441-447.

Klopman G, Saiakhov R, Rosenkranz HS, Hermens JLM. 1999. Multiple Computer-Automated Structure Evaluation program study of aquatic toxicity 1: Guppy. Environ Toxicol Chem 18:2497-2505.

Maggiora GM. 2006. On outliers and activity cliffs--why QSAR often disappoints. J Chem Inf Model 46:1535.

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF. 2006. 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 44:97-110.

Mayer P, Reichenberg F. 2006. Can highly hydrophobic organic substances cause aquatic baseline toxicity and can they contribute to mixture toxicity? Environ Toxicol Chem-2639.

Medina-Franco JL, Golbraikh A, Oloff S, Castillo R, Tropsha A. 2005. Quantitative structure-activity relationship analysis of pyridinone HIV-1 reverse transcriptase inhibitors using the k

nearest neighbor method and QSAR-based database mining. J Comput Aided Mol Des 19:229-242.

Moore DR, Breton RL, MacDonald DB. 2003. A comparison of model performance for six quantitative structure-activity relationship packages that predict acute toxicity to fish. Environ Toxicol Chem 22:1799-1809.

National Research Council. 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: National Academies Press.

Ng C, Xiao Y, Putnam W, Lum B, Tropsha A. 2004. Quantitative structure-pharmacokinetic parameters relationships (QSPKR) analysis of antimicrobial agents in humans using simulated annealing k-nearest-neighbor and partial least-square analysis methods. J Pharm Sci 93:2535-2544.

Norager O, Town WG, Petrie JH. 1978. Analysis of the registry of toxic effects of chemical substances (RTECS) files and conversion of the data in these files for input to the environmental chemicals data and information network (ECDIN). J Chem Inf Comput Sci 18:134-140.

Oloff S, Mailman RB, Tropsha A. 2005. Application of validated QSAR models of D1 dopaminergic antagonists for database mining. J Med Chem 48:7322-7332.

Rabinowitz JR, Goldsmith MR, Little SB, Pasquinelli MA. 2008. Computational molecular modeling for evaluating the toxicity of environmental chemicals: prioritizing bioassay requirements. Environ Health Perspect 116:573-577.

Ruden C, Hansson SO. 2003. How accurate are the European Union's classifications of chemical substances. Toxicol Lett 144:159-172.

Shen M, LeTiran A, Xiao Y, Golbraikh A, Kohn H, Tropsha A. 2002. Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. J Med Chem 45:2811-2823.

Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweyko A, Li Y. 2003. In silico ADME/Tox: why models fail. J Comput Aided Mol Des 17:83-92.

Toth B. 1988. Toxicities of hydrazines: a review. In Vivo 2:209-242.

Tropsha A, Golbraikh A. 2007. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. Curr Pharm Des 13:3494-3504.

Venkatapathy R, Moudgal CJ, Bruce RM. 2004. Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. J Chem Inf Comput Sci 44:1623-1629.

Verhaar HJ, Solbe J, Speksnijder J, van Leeuwen CJ, Hermens JL. 2000. Classifying environmental pollutants: Part 3. External validation of the classification system. Chemosphere 40:875-883.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH et al. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116:284-291.

Zheng W, Tropsha A. 2000. Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci 40:185-194.

Zhu H, Rusyn I, Richard A, Tropsha A. 2008. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. Environ Health Perspect 116:506-513.

Table 1. The results of data partitioning for the compounds with rat LD_{50}, mouse LD_{50}, rat chronic LOAEL and rat chronic NOAEL data in ZEBET dataset using cytotoxicity IC_{50} values.

Models	Number of C1	C1 ratio	Number of C2	C2 ratio
Rat LD_{50} (original set)	137	60%	93	40%
Mouse LD_{50}	119	56%	92	44%
Rat LOAEL	21	49%	21	51%
Rat NOAEL	19	46%	22	54%
Rat LD_{50} (full dataset)	258	61%	167	39%

Table 2. Statistical information for the five most statistically significant *k*NN QSAR models based on three modeling sets.

Nm	N-training	Pred.-training	N-test	Pred.- test	NNN
The best kNN classification model for 137 C1 vs. 93 C2 compounds					
1	173	0.84	55	0.73	1
2	147	0.86	74	0.70	1
3	193	0.83	37	0.73	1
4	165	0.86	59	0.70	1
5	173	0.81	55	0.75	1
The best kNN continuous model for 137 C1 compounds					
1	103	0.66	34	0.81	3
2	103	0.73	34	0.71	2
3	111	0.71	26	0.74	3
4	115	0.65	22	0.79	5
5	77	0.73	60	0.71	2
The best kNN continuous model for 93 C2 compounds					
1	80	0.61	13	0.84	2
2	77	0.67	16	0.77	1
3	80	0.69	13	0.74	1
4	80	0.65	13	0.76	2
5	79	0.63	14	0.78	2

N-training: number of compounds in the training set; Pred.-training: the overall predictivity of the training set (correct classification rate for classification models, q^2 for continuous models);

N-test: number of compounds in the test set; Pred.-test: the overall predictivity of the test set (correct classification rate for classification models, r^2 for continuous models); NNN; number of the nearest neighbors used for prediction.

Table 3. Comparison between TOPKAT and the two-step model prediction of the external compounds.

	Two-step model		TOPKAT	
	No applicability domain	With applicability domain	No applicability domain	With applicability domain
Prediction of 27 new ZEBET compounds				
R^2	0.64	0.86	0.16	0.60
MAE	0.38	0.29	0.78	0.50
Coverage	10%	67%	100%	67%
Prediction of 1,562 RTECS compounds with 70% confidence level				
R^2	0.26	0.33	0.19	0.22
MAE	0.65	0.54	0.76	0.65
Coverage	100%	62%	100%	62%
Prediction of 1,562 RTECS compounds with 90% confidence level				
R^2	0.42	0.62	0.19	0.26
MAE	0.60	0.42	0.84	0.66
Coverage	12%	6%	12%	6%

Figure Legends

Figure 1. The workflow of the two-step kNN QSAR LD_{50} modeling.

Figure 2. The identification of the baseline correlation between cytotoxicity (IC_{50}) and various types of *in vivo* toxicity testing results. (A) rat LD_{50}; (B) mouse LD_{50}; (C) rat LOAEL; (D) rat NOAEL.

Figure 3. The correlation between experimental and predicted LD_{50} values for 27 external compounds within applicability domain: (A) using TOPKAT and (B) using the two-step model developed in this study.

Figure 4. Percentage of compounds *vs*. the prediction errors obtained by the two-step rat LD_{50} model, TOPKAT and random sampling for 965 and 101 RTECS compounds.

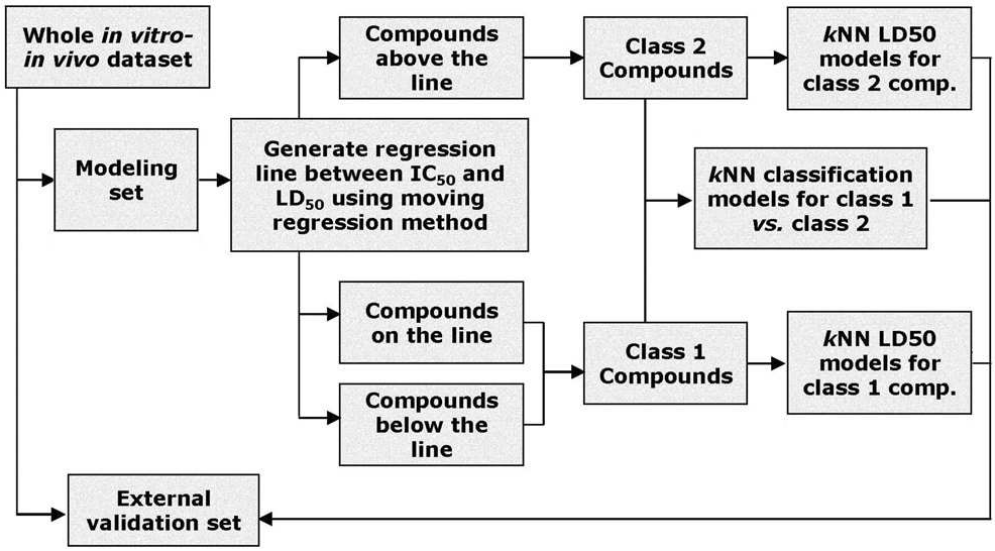

Figure 1. The workflow of the two-step kNN QSAR LD50 modeling.
83x46mm (300 x 300 DPI)

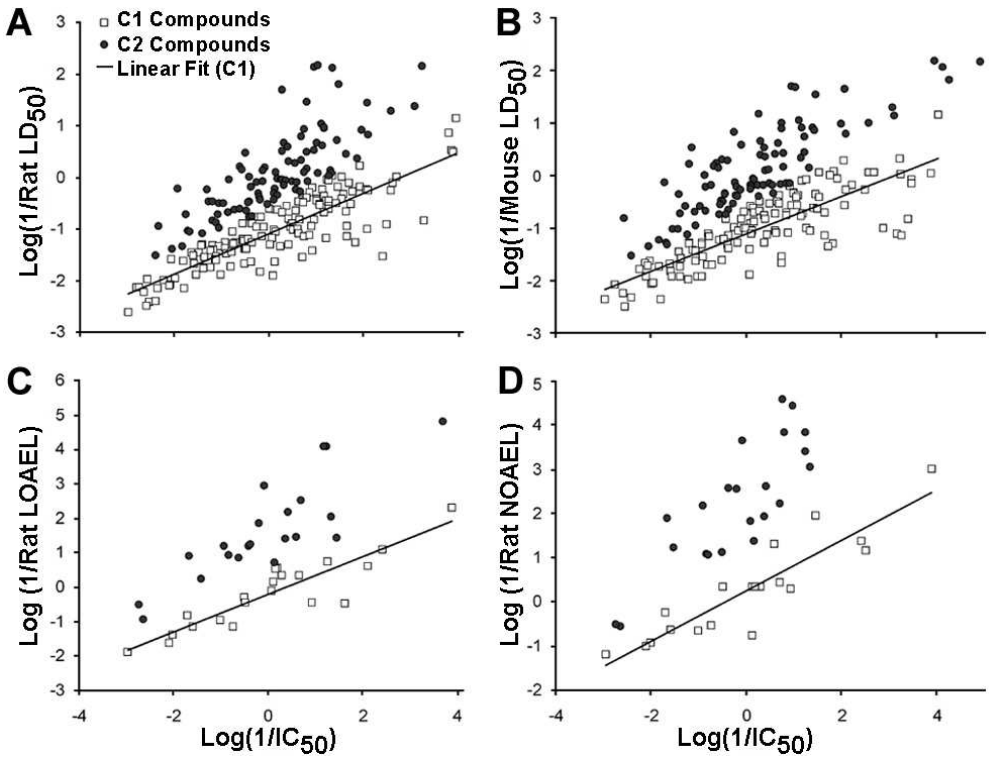

Figure 2. The identification of the baseline correlation between cytotoxicity (IC50) and various types of in vivo toxicity testing results. (A) rat LD50; (B) mouse LD50; (C) rat LOAEL; (D) rat NOAEL.
83x63mm (300 x 300 DPI)

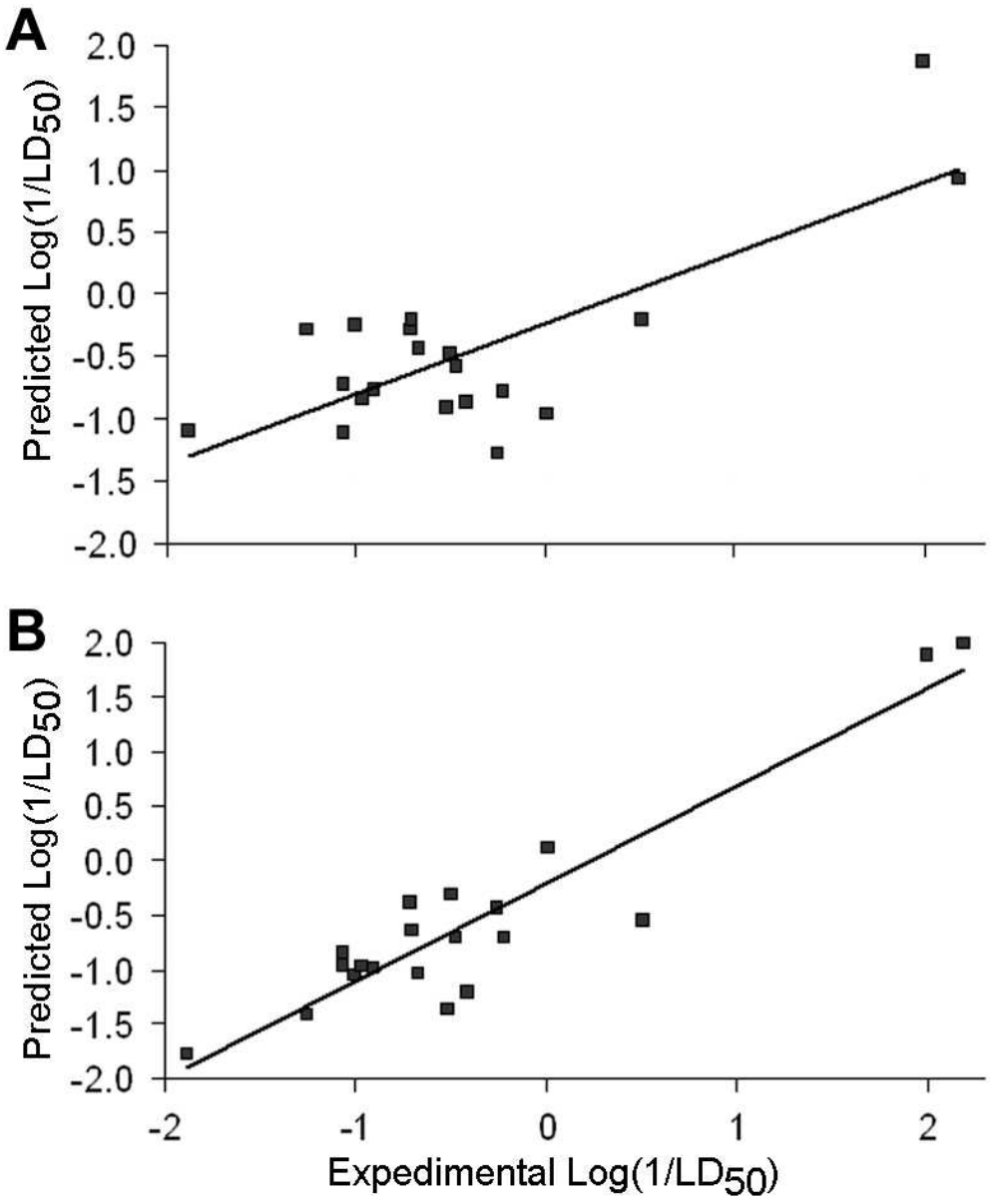

Figure 3. The correlation between experimental and predicted LD50 values for 27 external compounds within applicability domain: (A) using TOPKAT and (B) using the two-step model developed in this study.
66x79mm (300 x 300 DPI)

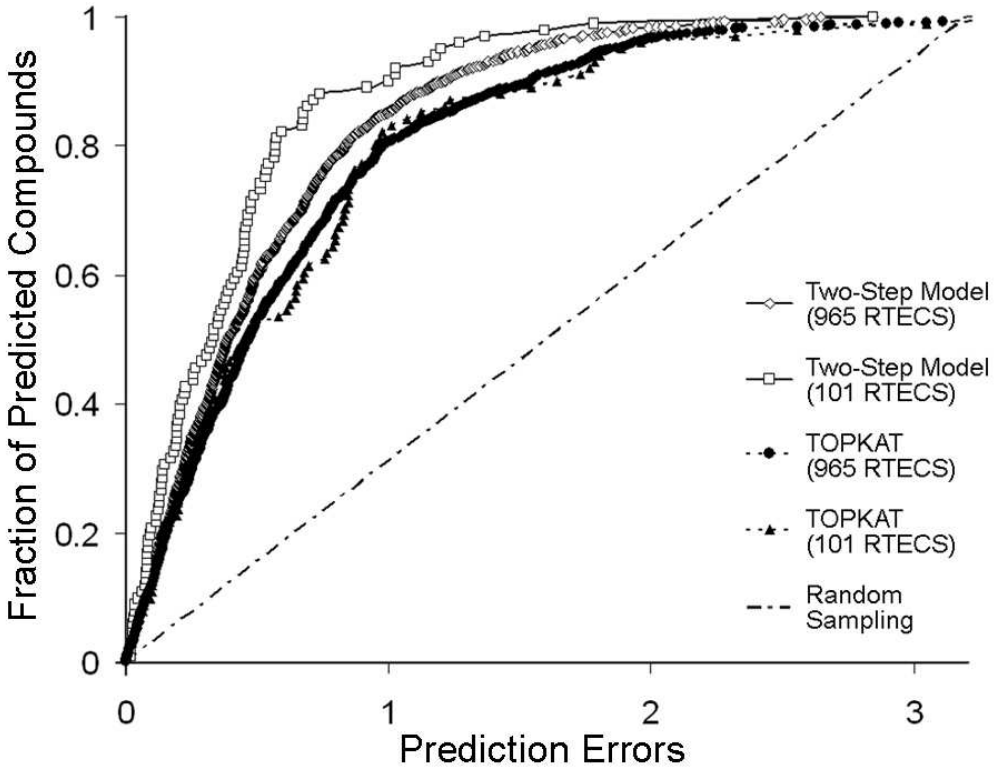

Figure 4. Percentage of compounds vs. the prediction errors obtained by the two-step rat LD50 model, TOPKAT and random sampling for 965 and 101 RTECS compounds.
76x58mm (300 x 300 DPI)