Quantitative Structure–In vitro–In vivo Correlations: Data Partitioning and QSAR Modeling of Acute Toxicity in Rats

Hao Zhu[1], Lin Ye[1], Ann Richard[2], Ivan Rusyn[3], and Alexander Tropsha[1]

[1]Laboratory for Molecular Modeling, School of Pharmacy, UNC-Chapel Hill; [2]National Center for Computational Toxicology, Office of Research & Development, US EPA;
[3]Department of Environmental Sciences and Engineering, School of Public Health, UNC-Chapel Hill

INTRODUCTION

- The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) hosted a workshop to study the relationship between cytotoxicity and acute rodent toxicity for more than 361 diverse compounds compiled by the German Center for the Documentation and Validation of Alternative Methods (ZEBET) [1]. This workshop showed that there is no clear relationship between these two types of endpoints. For example, the correlation between the cytotoxicity (IC50s) and the rat acute toxicity (LD50s) for 253 compounds in this dataset is poor (equation 1):

$$Log(1/LD50) = -0.64 + 0.44 \times Log(1/IC50)$$
$$R^2 = 0.45, \quad SE = 0.71, \quad N = 253 \qquad (1)$$

- The primary goal of this study is to develop robust and externally validated predictors for acute rodent toxicity. However, all previous attempts to develop *in vivo* rat LD50 QSAR models based on ZEBET dataset failed.

- In our recent study, we found that using High Throughput Screening (HTS) data as additional biological descriptors significantly improved the QSAR models for toxicity endpoints *in vivo* [2]. In this study, we have employed a similar hybrid modeling approach using cytotoxicity IC50 data to develop the *in vivo* LD50 QSAR models (Figure 1).

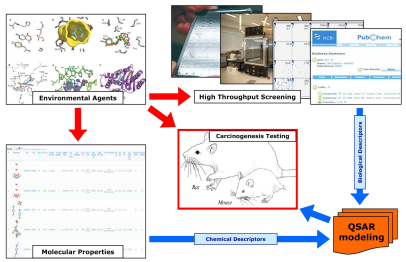

Figure 1. The use of hybrid chemical and biological descriptors for developing *in vivo* LD_{50} QSAR models.

This poster does not necessarily reflect EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

DATA

- For the purposes of this work, the dataset was curated as follows. Inorganic and organometallic compounds, as well as compound mixtures were excluded since these do not have conventional chemical descriptors used in QSAR studies. The curated subset from the original ZEBET dataset used in this work is comprised of 291 organic compounds.
- For 253 out of these 291 compounds, rat LD50 values were available. The following criteria were used to select the LD50 values: 1) Only LD50 values published in Registry of Toxic Effects of Chemical Substances (RTECS) [3] were used; 2) If different issues of RTECS reported different LD50 values, then the largest LD50 value was used.

METHODOLOGY

- Visual inspection of the plot comparing cytotoxicity and rat acute toxicity data (Figure 2) leads to the following observations:
  1) cytotoxicity is directly correlated with the acute toxicity only for some compounds in the dataset;
  2) most of the remaining compounds have higher acute toxicity in rats vs *in vitro*;
  3) a small fraction of compounds have lower acute toxicity in rats vs *in vitro*.
- Based on these observations, we could partition compounds in the modeling set into three subsets: **Class 1**, in which compounds' acute rat toxicity linearly correlates with their cytotoxicity; **Class 2**, in which the acute rat toxicity does not correlate with cytotoxicity with points positioned above the regression line for Class 1; and a small set of **outliers** with points located below the regression line for Class 1.
- Based on this approach to data partitioning, we have designed the analytical workflow as shown in Figure 3.

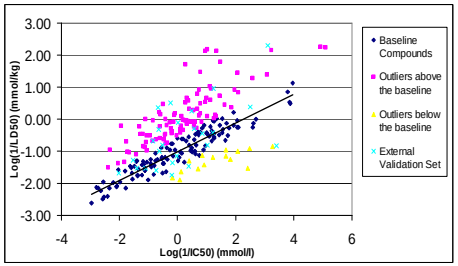

Figure 2. The identification of the baseline correlation between IC_{50} and LD_{50} values for the modeling set.

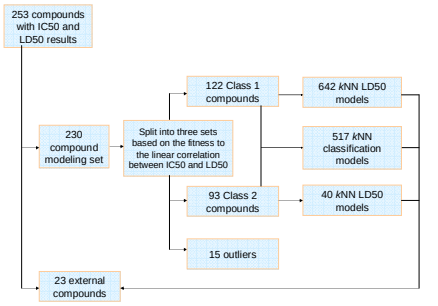

Figure 3. Flowchart of data modeling for *in vivo* LD_{50} prediction using chemical structure and IC_{50} data.

RESULTS & DISCUSSION

- There is no clear difference in chemical similarity distribution between compounds in Class 1 and Class 2 (Figure 4). The predictions of LD50 values for the 23 external compounds are shown in Table 1.

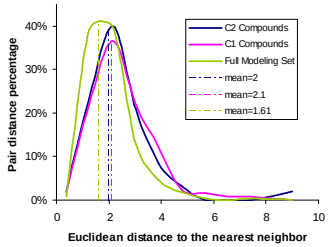

Figure 4. The comparison of chemical similarity between the original modeling set and two new subsets.

Table 1. Consensus prediction of 23 compounds in the external validation set using the two step kNN QSAR approach.

CAS	Chemical Name	Exp. $Log(1/IC_{50})$	Exp. $Log(1/LD_{50})$	Real Class	Pred. Class	Pred. $Log(1/LD_{50})$
72571	Trypan blue	1.02	-0.81	C1	no pred.	no pred.
7327879	Dihydralazine sulfate	0.85	-0.45	C1	no pred.	no pred.
132605	Cinchophen	0.57	-0.30	C2	C1	-0.44
65452	Salicylamide	-0.03	-1.14	C1	C1	-1.02
54115	Nicotine	-0.25	0.51	C2	C2	0.05
84662	Diethyl phthalate	-0.74	-1.59	C1	C1	-0.98
20624253	Diethyldithiocarbamate sodium 3H2O	3.41	-0.82	Outli.	C1	-1.39
51183	Triethylene melamine	3.11	2.30	C2	no pred.	no pred.
77474	Hexachlorocyclopentadiene	2.51	0.39	C1	no pred.	no pred.
76448	Heptachlor	1.23	0.96	C2	C2	0.88
86544	Hydralazine	0.48	0.25	C2	C2	-0.26
136607	n-Butyl benzoate	0.39	-1.46	Outli.	C1	-0.96
108610782	3-Cyano-2-morpholino-5-(pyrid-4-yl)-pyridine	0.02	-0.11	C2	no pred.	no pred.
5435643	Isononylaldehyde	-0.18	-1.36	C1	C1	-1.19
110407	Diethyl sebacate	-0.21	-1.75	Outli.	no pred.	no pred.
69727	Salicylic acid	-0.53	-0.81	C2	C1	-1.03
59427	Phenylephrine	-0.65	-0.32	C2	C1	-0.96
78415722	Milrinone	-0.68	0.37	C2	C2	0.08
62533	Aniline	-0.84	-0.67	C2	C2	-0.47
123728	n-Butanal	-1.11	-1.54	C1	C1	-1.54
71410	1-Pentanol	-1.40	-1.54	C1	C1	-1.46
75092	Dichloromethane	-1.54	-1.27	C2	C1	-1.89
78933	Ethyl methyl ketone	-2.02	-1.67	C1	C1	-1.81

CONCLUSIONS

The cytotoxicity data show weak direct correlation with *in vivo* acute toxicity. Nevertheless, these data could be used to assist QSAR modeling of *in vivo* acute toxicity. We have developed a novel two-phase approach that leads to successful *k*NN QSAR rat LD50 models. LD50 values were predicted for 23 external compounds with high accuracy (R^2 = 0.80, SE = 0.34, Coverage = 74%). We believe that this activity-based partitioning approach using the *in vitro* toxicity data can be successfully applied to other complex *in vivo* toxicity endpoints. This approach makes it feasible to combine *in vitro* screening methods and QSAR modeling to prioritize chemicals for *in vivo* animal toxicity testing.

REFERENCES

- ICCVAM: http://iccvam.niehs.nih.gov/
- Zhu, H., Rusyn, I., Richard, A., Tropsha, A. *Environ. Health Persp.* 2008, 116, in press.
- Ruden, C., Hansson. S. *Toxicol. Lett.* 2003, 144, 159-172.

This study was supported, in part, by the NIH RoadMap grant GM076059 and by EPA STAR grant RD832720