The Use of *in vitro* Dose-Response Profiles Enhances QSAR Modeling of *in vivo* Toxicity

Alexander Sedykh[1], Hao Zhu[1], Hao Tang[1], Liying Zhang[1], Ann Richard[2], Ivan Rusyn[3] and Alexander Tropsha[1]

[1]Laboratory for Molecular Modeling, School of Pharmacy, University of North Carolina at Chapel Hill; [2]National Center for Computational Toxicology, Research Triangle Park, United States Environmental Protection Agency; [3]Department of Environmental Sciences & Engineering, School of Public Health, University of North Carolina at Chapel Hill

INTRODUCTION

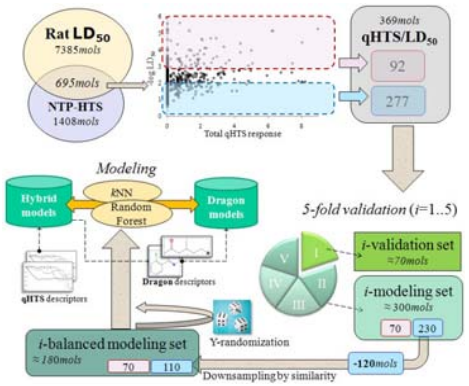
Background: Toxicity evaluation of chemicals *in vivo* is time-consuming and expensive. To develop possible alternatives, EPA's ToxCastTM program and the Tox21 program (a collaboration between EPA, the National Toxicology Program, and NIH), are developing approaches that rely on *in vitro* high-throughput screening (HTS) toxicity assays. Thus far, *in vitro/in vivo* concordance has been generally poor. Quantitative Structure-Activity Relationship (QSAR) models can predict toxicity directly from the chemical structure with limited accuracy that can be further improved by incorporating *in vitro* profiles as additional biological descriptors of chemicals (Zhu et al, 2008).

Objective: To translate dose-response information from quantitative HTS (qHTS) assays into novel biological descriptors of chemicals and employ these new descriptors to increase the accuracy of QSAR models in predicting *in vivo* toxicity of environmental chemicals

DATA

- **NTP-qHTS Dataset.** Dose-response profiles of 1,408 substances screened for their effect on cell viability are available through PubChem for 13 cell lines. Each substance was assessed at 14 different concentrations ranging from 0.6 nM to 92 μM.
- **Rat LD_{50} Dataset.** 7,385 unique organic compounds with LD_{50} values expressed as a negative logarithm of mol/kg units (see also http://www.epa.gov/nrmrl/std/cppb/qsar/DataSets.zip).
- **qHTS/LD_{50} Dataset.** 369 compounds with qHTS profiles and LD_{50} values ≤ 10^{-3} mol/kg ("toxic") and ≥ 10^{-2} mol/kg ("non-toxic").

Modeling workflow for exploring chemical and biological dose-response descriptors for predicting *in vivo* toxicity.

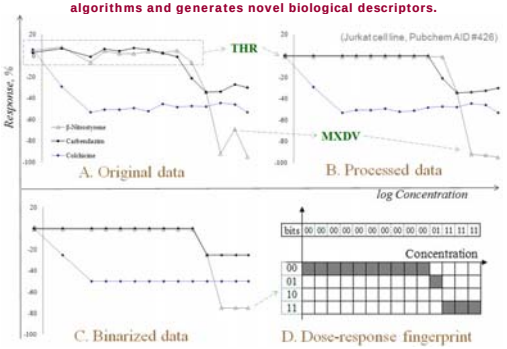


This poster does not necessarily reflect EPA policy. Mentioning of trade names or commercial products does not constitute endorsement or recommendation for use.

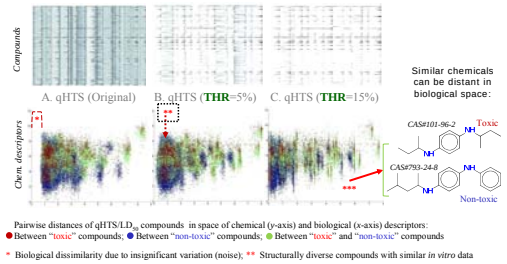
METHODOLOGY

- **Chemical descriptors.** 382 2D descriptors from Dragon software were used: constitutional, functional group counts, atom-centered fragments, molecular properties, 2D binary fingerprints and 2D frequency fingerprints.
- **qHTS descriptors and data treatment.** Dose-response profiles were processed to suppress noise by adjusting two parameters: **THR (threshold)**, which controls for the variation near baseline, and **MXDV** – maximum deviation from a monotonous behavior. Processed qHTS measurements at each concentration were used for modeling (13 cell-lines×14 concentrations=182 qHTS "biological dose-response" descriptors).

Processing of the qHTS dose-response data employs noise-reduction algorithms and generates novel biological descriptors.



Impact of noise-reduction thresholds on qHTS data content.



Pairwise distances of qHTS/LD_{50} compounds in space of chemical (*y*-axis) and biological (*x*-axis) descriptors:
● Between "toxic" compounds; ● Between "non-toxic" compounds; ● Between "toxic" and "non-toxic" compounds

* Biological dissimilarity due to insignificant variation (noise); ** Structurally diverse compounds with similar *in vitro* data

- The *in vitro* data noise (*) is suppressed after treatments.
- Structurally diverse (y-axis) compounds can be close in biological space (x-axis, **) and vice versa (***). Therefore hybrid (qHTS + chemical) descriptors may have higher explanatory power towards *in vivo* toxicity.

MODELING DETAILS

- **Classification parameters:**
  Sensitivity = (Compounds predicted toxic) / (Experimentally toxic)
  Specificity = (Predicted non-toxic) / (Experimentally non-toxic)
  Correct Classification Rate, ***CCR*** = (Sensitivity + Specificity) /2
- **Random Forest (RF)**: "Decision tree" algorithm. A group of *n* models, each based on *m* descriptors (*n*=500, *m*=13).
- ***k*NN:** The activity of a data-point is calculated from its *k* nearest neighbors (as a distance-weighted average). A group of models with *k*=1,...,9 and 5,...,40 subset of descriptors to define neighbors. Models with Leave-One-Out (LOO) Correct Classification Rate (CCR) < 0.65 were removed.
- **Applicability Domain (AD):** A maximum distance of a test-compound from modeling dataset, beyond which no prediction will be made. AD=$d+0.5\sigma$, where d and σ are mean and st.dev of the modeling set distances.
- **Model evaluation:**
  5-fold validation: 20% of data as external set, 80% as modeling set, repeated 5 times and prediction results averaged.
  Y-randomization: Activity labels are randomly shuffled and models re-derived (5 times). One-tail t-test is then applied to estimate difference from actual models.

RESULTS & DISCUSSION

Application of noise-reduction threshold to dose-response data improves predictions of models based on qHTS descriptors.

%	Chem. only	Hybrid (Original)	Hybrid (THR=5%)	Hybrid (THR=15%)	Hybrid (THR=25%)
			***k*NN**		
Sensitivity	68±8	63±9	76±9	76±5	74±9
Specificity	85±4	86±4	87±3	87±2	88±2
CCR	**76±5 ***	**74±5**	**82±6**	**82±3**	**81±6**
			Random Forest (RF)		
Sensitivity	74±9	66±8	76±12	77±10	75±12
Specificity	82±7	87±4	85±3	86±3	85±4
CCR	**78±4 ***	**77±5**	**80±6**	**82±5**	**80±6**

Shown are averaged results of five-fold external validation. *Chem.only models were significantly different ($p < 0.05$) from all other models of the corresponding group by the permutation test (10,000 times).

Hybrid QSAR models have higher predictive power than commercial TOPKAT software

%	TOP-KAT	Chem. only		Hybrid (Original)		Hybrid (THR=5%)		Hybrid (THR=15%)		Hybrid (THR=25%)	
		*k*NN	RF	*k*NN	RF	*k*NN	RF	*k*NN	RF	*k*NN	RF
Sens.	0.45	0.73	0.73	0.55	0.82	0.91	0.91	0.91	0.91	0.82	0.91
Spec.	0.93	0.78	0.80	0.85	0.78	0.85	0.83	0.85	0.83	0.88	0.83
CCR	**0.69 ***	**0.75**	**0.77**	**0.70**	**0.80**	**0.88**	**0.87**	**0.88**	**0.87**	**0.85**	**0.87**

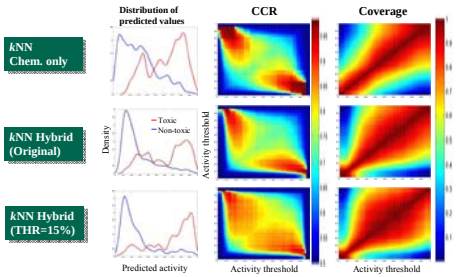
Results are shown for 52 compounds in our external validation sets, which were also absent in the TOPKAT training set. *TOPKAT model was significantly different ($p < 0.05$) from all other models by the permutation test (10,000 times).

- Conventional QSAR models (chemical descriptors only) were better than models based on the original, unprocessed *in vitro* data (Hybrid-Original), but worse than Hybrid models built on *in vitro* data after the noise-reduction.

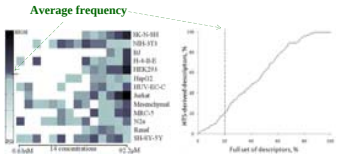
Activity thresholds:

Continuous (between 0 and 1) *k*NN predictions can be classified to **1** (toxic) or **0** (non-toxic) using different thresholds (default is 0.5) for each class, e.g. "≥0.8" for toxic and "<0.5" for non-toxic. Model accuracy can be improved by excluding borderline predictions (thereby, reducing model's coverage).

The use of qHTS dose-response data improves prediction accuracy and coverage of QSAR models



Relative contribution of qHTS dose-response descriptors to QSAR models varies between cell-lines.



A. Relative occurrence B. Frequency-ranked list in Hybrid *k*NN (THR=15%) model

CONCLUSIONS

- qHTS *in vitro* data is not sufficient to predict *in vivo* toxicity in this study.
- When employed as "biological descriptors" in combination with conventional chemical descriptors, qHTS dose-response profiles can improve the outcome of QSAR modeling of *in vivo* endpoints.
- The improvement is only enabled if *in vitro* data noise is suppressed using specially developed algorithms

REFERENCES

- Zhu H, Martin TM, Ye L, Sedykh A, Young DM, Tropsha A. *QSAR Modeling of Rat Acute Toxicity by Oral Exposure*. Chem.Res.Toxicol. (2009), 22[12], 1913-1921.
- Zhu H, Rusyn I, Richard A, Tropsha A.* Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. Environ Health Perspect 2008; (116): 506-513

This work was supported, in part, by grants from NIH (GM076059 and ES005948), EPA (RD832720 and RD833825) and The Johns Hopkins Center for Alternatives to Animal Testing (2009-13).