New QSAR system to predict acute inhalation toxicity

Dr Roustem Saiakhov
MultiCASE Inc

PETA October 31st, 2018

Acute inhalation toxicity is the adverse effect caused by a substance after a single uninterrupted exposure by inhalation over a period of 24 hours or less. Environmental pollution and workplace/household exposure require special attention to the inhalation toxicity of chemicals that can cause damage to the respiratory organs and organism as a whole. While regulatory bodies, e.g. European Union, require inhalation toxicity evaluation to be performed for high volume production chemicals, the use of animals is not permitted in such evaluations.

In place of animal testing, it is often recommended to use search and retrieval of existing data, identification of structural alerts to judge toxicity, grouping of chemicals for read-across analysis, and the development and use of quantitative structure-activity models. However, the volume of experimental data and number of reliable QSAR models, concerning the acute inhalation toxicity is limited which hinders all in-silico approaches.

In this study, we developed a set of QSAR models built using publicly available acute inhalation toxicity data in rat. The set consists of 4 categorical models, designed to predict one of the four GHS categories for inhalation toxicity, and a model to predict LC50 values directly.

All models demonstrated leave group out cross-validation performance in the range of 52-71% sensitivity, 71-74 % specificity, 64-73 % positive predictivity, and 61-68 % negative predictivity. The inhalation toxicity set has assign the correct GHS category to 75% of the external validation compounds. Coverage was maintained in the range of 71-84 %. Identified alerts are discussed along with the advantages and disadvantages of using the categorical versus the direct model. Overall, this approach provides a reliable QSAR method for predicting acute inhalation toxicity of chemical substances, only using their structures.

Regulatory Context

- Acute inhalation toxicity data are used to satisfy hazard classification and labeling requirements, to estimate the toxicity of mixtures, and to assess human health and environmental risk.
- The derivation of either a point estimate of the LC50 value (using TG 403) or a range estimate of the LC50 (using TG 436) generally meets the acute inhalation toxicity regulatory requirements for classification and labeling of industrial chemicals, consumer products, and many pesticide applications.
- (Q)SAR models for Inhalation toxicity endpoints may be used as early screening tools during drug development or to support regulatory safety decisions when experimental data are limited or unavailable.
- Predictions may be used to contribute to the weight-of-evidence supporting a regulatory safety decision.

Objectives

Build optimized InhTOX (Q)SAR models for

- GHS categorical prediction (range estimate of LC50)
- Point estimate of LC50 predictions

Software and Training Sets for (Q)SAR Models

- All training sets used to construct (Q)SAR models were generated from an existing non-proprietary database of Inhalation Toxicity assays sourced from publicly available databases and the published literature[1-2].

  [1] RTECS, Registry Of Toxic Effects of Chemical Substances. [Atlanta, Ga.] :U.S. Dept. of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 1997

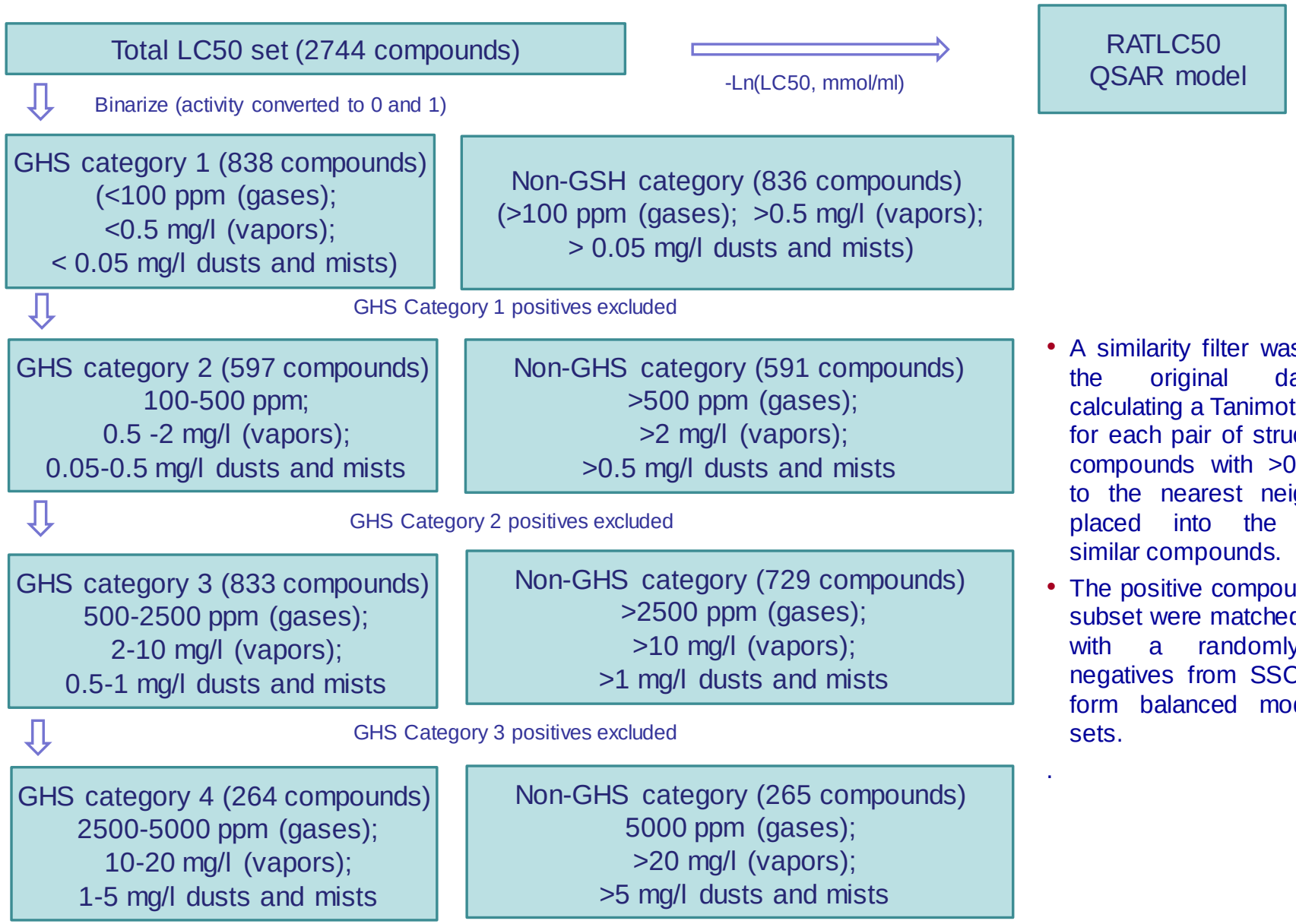
  [2] O. A. Raevsky, E. A. Modina, O. E. Raevskaya SAR models of the inhalation toxicity of organic compounds Pharmaceutical Chemistry Journal 45(3):165-169, 2011

- CASE Ultra (CU) v1.6.2.3 (MultiCASE, Inc.) software was used to build statistical QSAR models.
- A binary scoring system was used to classify assay results as negative (0) or positive (1) for GHS model bundle.
- The endpoints of acute rat inhalation toxicity (4 hours exposure) was chosen for this project. The training sets consisted of 2744 findings, where 30% were Category 1, 21% were Category 2, 30% were Category 3 and 10% were Category 4.

Data preprocessing

- A ***binary scoring system*** was used to classify assay results as negative (0) or positive (1) for GHS model bundle.
- The endpoints of ***acute rat inhalation toxicity (4 hours exposure)*** was chosen for this project. The training sets consisted of 2744 findings, where 30% were Category 1, 21% were Category 2, 30% were Category 3 and 10% were Category 4.
- After ***downsampling***, newly constructed model training sets were created containing an equal number of positive and negative findings with the total size of models ranging from 529 to 1674 records. The remaining compounds were used to create supporting datasets to decrease “out of domain” calls produced by the final model.

Similarity-Based Downsampling



- A similarity filter was applied to the original dataset by calculating a Tanimoto coefficient for each pair of structures. Only compounds with >0.3 similarity to the nearest neighbor were placed into the structurally similar compounds.
- The positive compounds of GHS subset were matched in 1:1 ratio with a randomly selected negatives from SSC subset, to form balanced model training sets.

.

Composition of InhTOX Models

Model	Endpoint	Balanced model size
INHTOX_CAT1	Inhalation Toxicity GHS category 1 (<100 ppm)	1674 (P:838/N:836)
INHTOX_CAT2	Inhalation Toxicity GHS category 2 (100-500 ppm)	1188 (P:597/N:591)
INHTOX_CAT3	Inhalation Toxicity GHS category 3 (500-2500 ppm)	1562 (P:833/N:729)
INHTOX_CAT4	Inhalation Toxicity GHS category 4 (2500-5000 ppm)	529 (P:264/N:265)
INHTOX_RATLC50	Inhalation Toxicity, Acute, Rat 4 hours LC50	2744 (P:1345/M:43/N:1356)

P=Number of positives; N=Number of negatives; M= Number of marginals

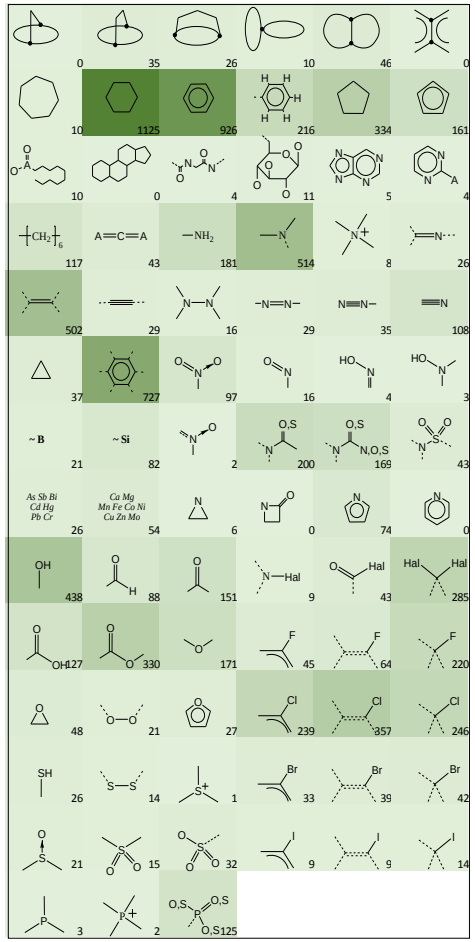
Predictive Performance of InhTOX Balanced Models

Model	Leave group out mode % Off/Times	Balanced model					External validation, correct classification
		Sensitivity	Specificity	Positive Accuracy	Negative Accuracy	Coverage	
INHTOX_CAT1	10/10	52.1	71.7	63.8	61.5	64.4	75.1
INHTOX_CAT2	10/10	50.8	72.7	62.3	62.6	62.3	80.5
INHTOX_CAT3	10/10	70.5	71.0	74.6	66.6	64.7	100.0
INHTOX_CAT4	10/10	60.4	71.8	66.9	65.5	61.7	78.0
INHTOX_RATLC50	10/10	69.5	75.1	72.3	72.6	66.8	n/a

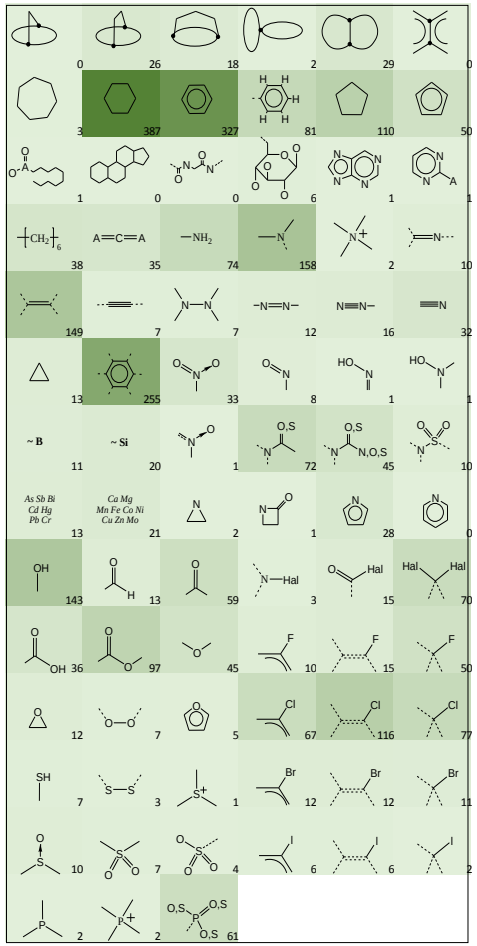
- If a query chemical raises an out of domain call in a GHS balanced model but has no unknown fragments in a corresponding INHTOX_RATLC50 model, the out of domain call is considered as negative.

Distribution of Main Structural Features in InhTox Models

RATLC50



Cat1

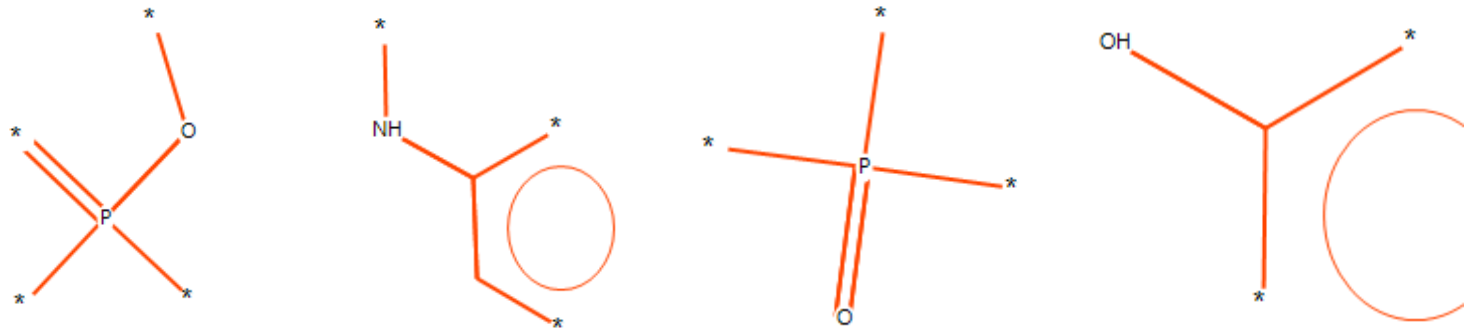


Structural feature representation is maintained between the categorical and LC50 models

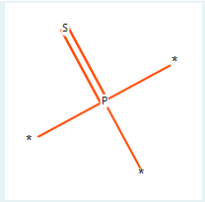
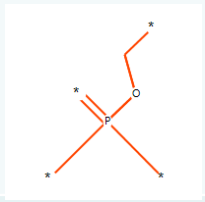
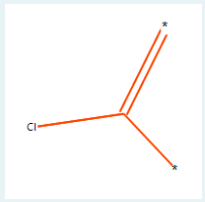
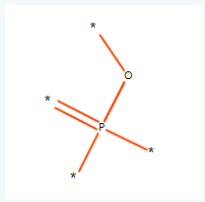
Analysis of Structural Alerts in Final InhTox Models

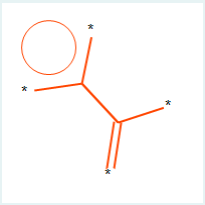
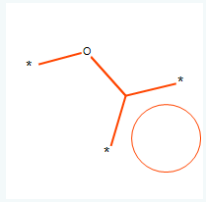
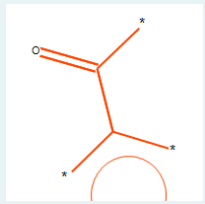
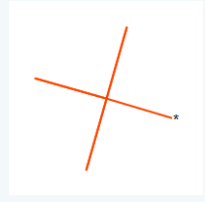
- There is a clear distinction between alerts, identified for different GHS categories. The distinction is based on the mechanistic mode of action
- LC50 model identifies more modes of actions, compare to the GHS models

Top alerts LC50 model



Top alerts for GHS models

Model	Alert	Positives	Negatives
InhTOX_CAT1		37	12
InhTOX_CAT1		35	12
InhTOX_CAT2		41	14
InhTOX_CAT2		30	12

Model	Alert	Positives	Negatives
InhTOX_CAT3		93	27
InhTOX_CAT3		92	27
InhTOX_CAT4		13	6
InhTOX_CAT4		10	3

Conclusions

- Combining a similarity-based downsampled (Q)SAR model and a supporting database of additional chemicals provides more sensitive predictions of inhalation toxicity, without a significant decrease in structural coverage.
- The sensitivity and negative predictivity of the new Inhalation Toxicity models demonstrate a reliable QSAR method for predicting acute inhalation toxicity of chemical substances, only using their structures

Questions?

Please forward your inquiries to:

saiakhov@multicase.com or sales@multicase.com