EPA
United States
Environmental Protection
Agency

Practical Considerations for Estimation of Oral Rat LD50

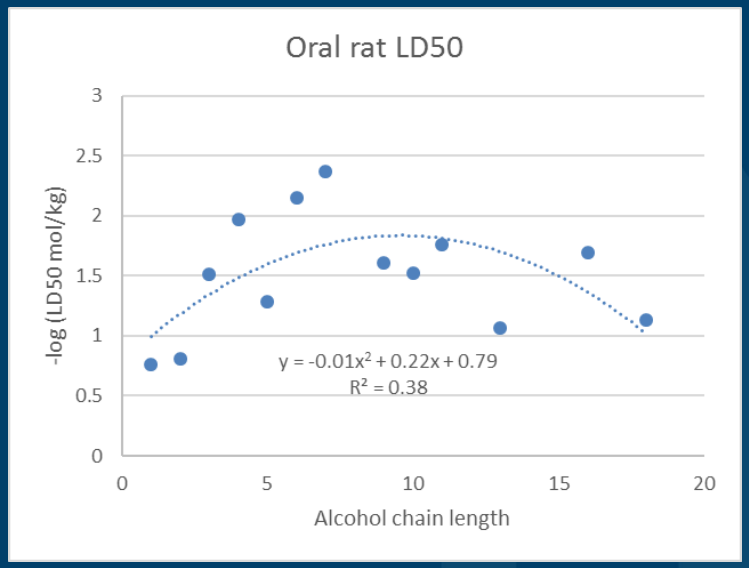
Todd Martin, US EPA

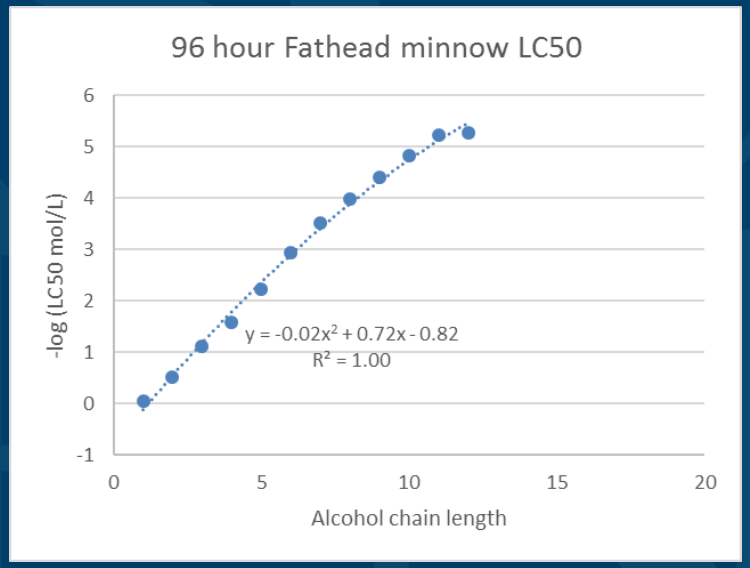
Office of Research and Development
National Risk Management Research Laboratory / Sustainable Technology Division / Clean Processes Branch

April 16, 2018

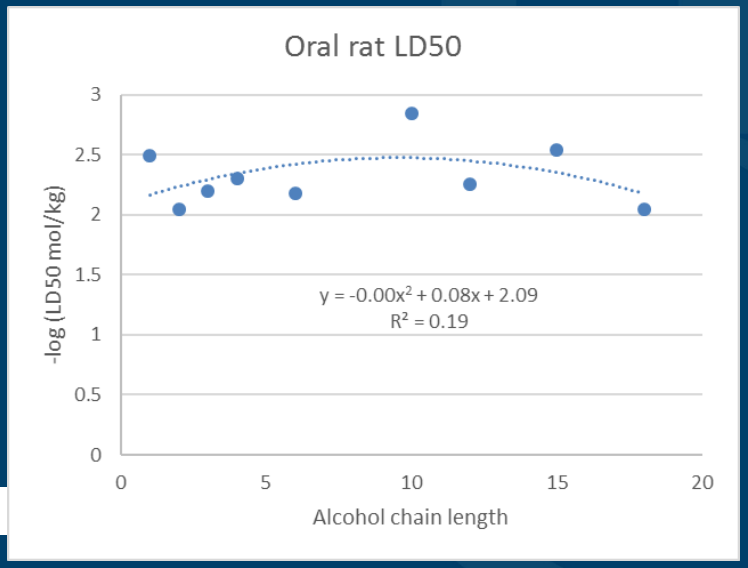
Difficulty with oral rat LD50 endpoint

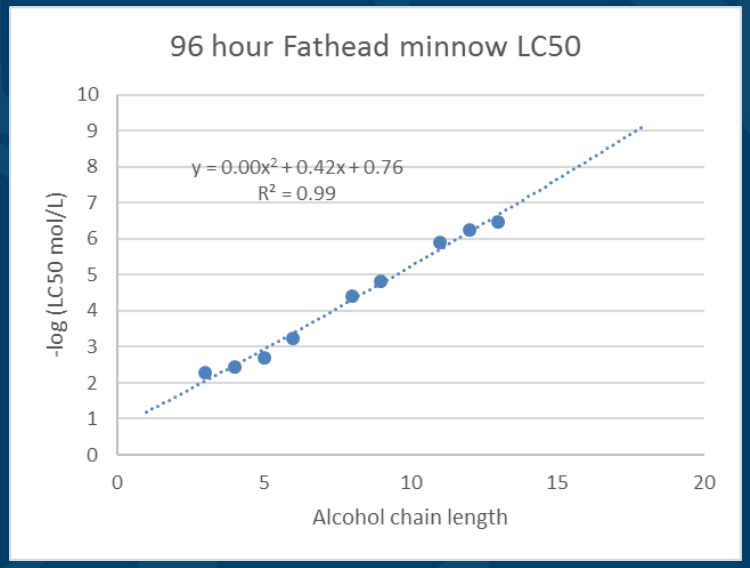
United States Environmental Protection Agency





n-alcohols





n-amines

Additional data to improve results

- ChemIDplus lists effects for some chemicals for oral rat LD50

rat	LD50	oral	930mg/kg (930mg/kg)	BEHAVIORAL: TREMOR BEHAVIORAL: CONVULSIONS OR EFFECT ON SEIZURE THRESHOLD	Toxicology and Applied Pharmacology. Vol. 7, Pg. 767, 1965.

- There is pest target species database:
  - C.L. Russom, *Pesticide Acute MOA Database: Overview of procedures used in compiling the database and summary of results*, US EPA, Duluth, MN, 2013, p. 26.
  - T. M. Martin, C. R. Lilavois & M. G. Barron (2017) Prediction of pesticide acute toxicity using two-dimensional chemical descriptors and target species classification, SAR and QSAR in Environmental Research, 28:6, 525-539, DOI: 10.1080/1062936X.2017.1343204
- Both of these can help bin chemicals but may not improve external prediction accuracy since not available for new chemicals (have to predict them before can bin a chemical)

Practical Considerations

- What were the major practical considerations identified when modelling – are there recommendations you could propose that would have helped resolve some/all of these?
  - Should we fit models to continuous LD50 values or to binary/category endpoints? Or both depending on final goal?
- Were there limits to appropriate methods for the different endpoints?
  - It's very difficult to fit a single regression model to the entire LD50 training set so need clustering or neighbor based method

QSAR Method Selection

- What were the rationale to applying particular methods
  - Needed a method that could be applied to very large dataset with lots of experimental error
- What are advantages/disadvantages of making specific method choices?
  - Hierarchical clustering
    - Slow for model building and prediction
    - Can correlate differences in toxicity for subsets of the data
    - Hard to make “external” predictions for training compounds
  - Nearest neighbor
    - Very fast and easy to understand
    - Can’t quantify differences in toxicity between test compound and the neighbors
    - Can always make “external” predictions for training compounds

Applicability Domain

- How will I know if my chemistry is within the domain of applicability of your model?
  - Hierarchical clustering
    - Model ellipsoid, Rmax, and Fragment constraints

$$h_{00} = X_o^T \left(X^T X \right)^{-1} X_0$$

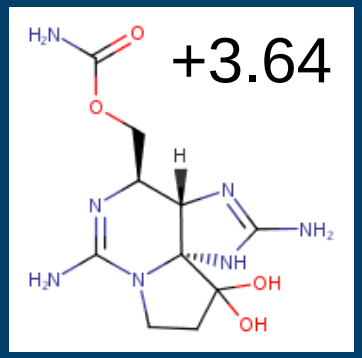
$$distance_i = \sum_{j=1}^{d} \left(x_{ij} - C_j \right)^2$$

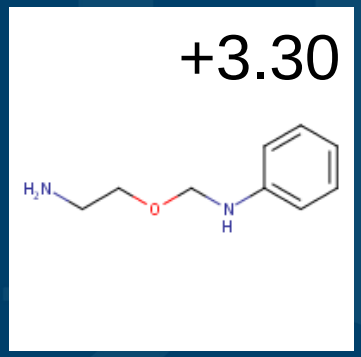
  - Nearest neighbor
    - Three most similar chemicals must exceed a minimum cosine similarity coefficient of 0.5

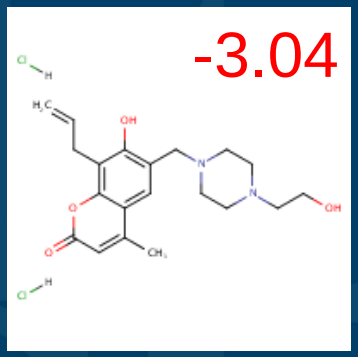
$$SC_{i,k} = \frac{\sum_{j=1}^{\#descriptors} x_{ij} x_{kj}}{\sqrt{\sum_{j=1}^{\#descriptors} x_{ij}^2 \cdot \sum_{j=1}^{\#descriptors} x_{kj}^2}}$$

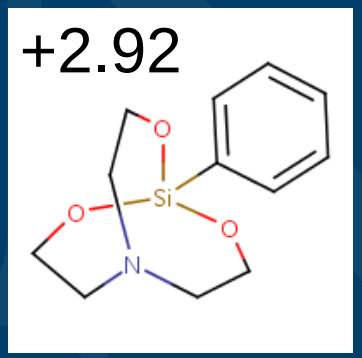
Difficult to predict chemicals

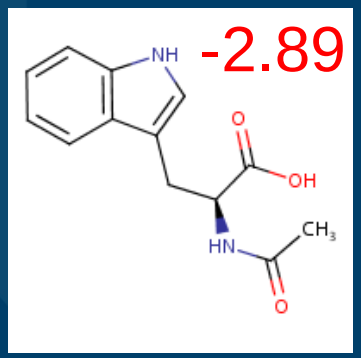
- What elements do “difficult to predict” chemistries have in common?
  - 16 of the top 20 worst predicted compounds contain **Nitrogen**
    - The nitrogens are usually in an aromatic ring or attached to one
  - Usually it’s hard for machine-based methods to detect reactive substructures purely from descriptor values

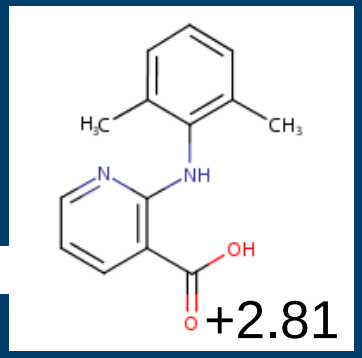


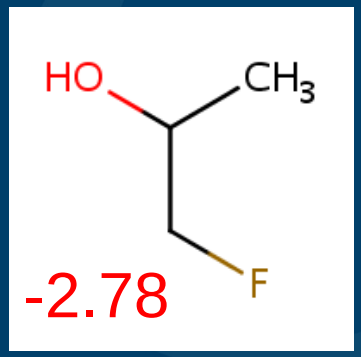


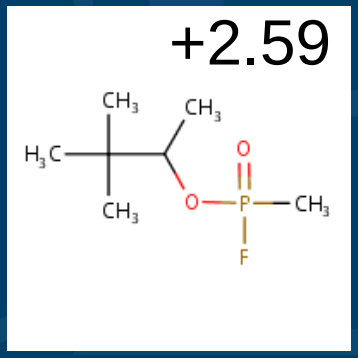


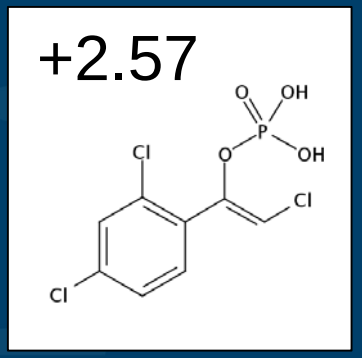


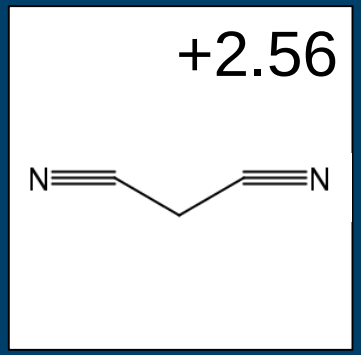






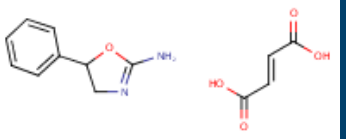




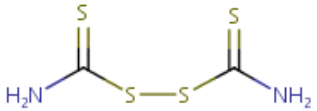


Difficult to predict, cont.

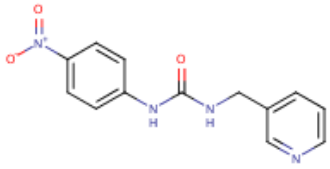
+2.53



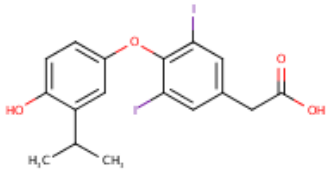
-2.50



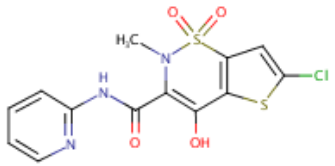
2.40



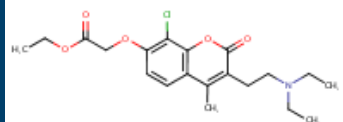
+2.38



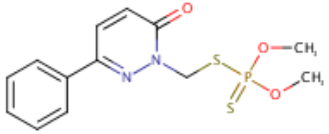
+2.33



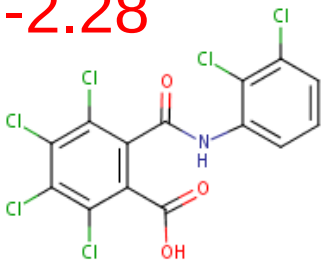
-2.30

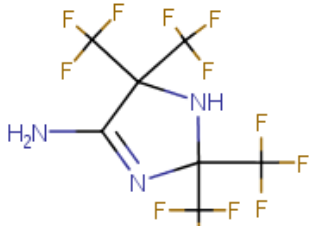


-2.29



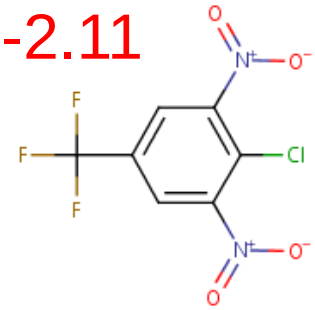
-2.28





+2.27

-2.11



* Numbers represent the prediction error in log10 (LD50 mg/kg)

Metabolism / bioavailability

- Does metabolism (or bioavailability) play a role in the ability to predict acute toxicity?
  - Usually QSAR models take that into account from the different functional groups in the molecule if the training set is big enough
  - May not be able to detect reactive substructures
    - e.g. OH near C=C for an allylic alcohol

Choice of Dependent Variable For Predicting Category Endpoints

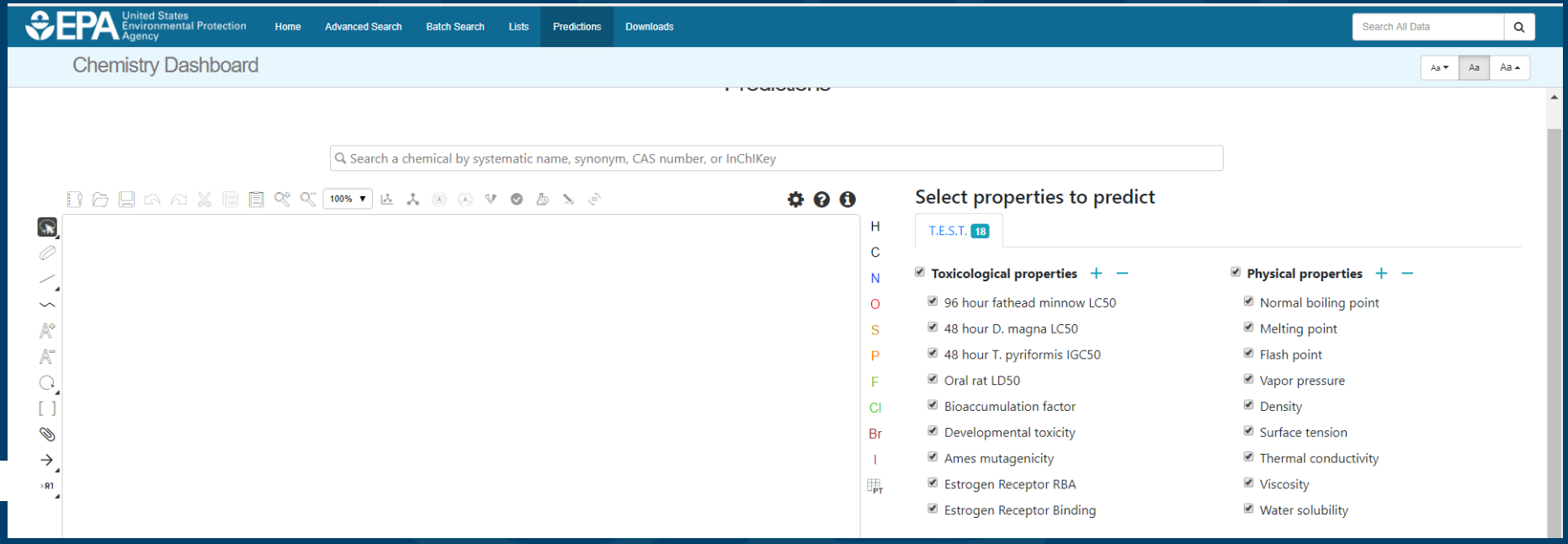
- Additional data points (8953 vs 6706 chemicals for “very toxic” endpoint)
- Continuous endpoint (-log(LD50 mol/kg) has more concentration dependence information

Method	Use method AD: AA*	Use method AD: Cov.	Predictions in common: AA	Predictions in common: Cov.
HC Cat. Input	0.515	0.794	0.527	0.659
HC LD_{50} input	0.467	0.821	0.487	0.659
NN Cat. Input	0.498	0.878	0.498	0.875
NN LD_{50} input	0.426	0.993	0.445	0.875

*AA = average accuracy over five categories,
Cov. = fraction of chemicals predicted

Availability

- How would end-users access your model and run it on new chemicals?
  - T.E.S.T. downloadable Java program
    - https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test
  - **WebTest:**
    - https://comptox.epa.gov/dashboard/predictions/index
    - Via graphical user interface or API call
  - Training set is probably slightly different



Mechanistic Considerations

- What was there coverage/evaluation of different mechanisms of action?
  - Look at pesticide data or effect data from ChemIDplus?
- What was the approach taken to determine mechanistic interpretations? What were the strengths and weaknesses of that approach? How does that compare/contrast with other approaches?
  - Didn't investigate that since prior work didn't show much promise

Mechanistic Considerations, cont.

- Is "structure-basis" (i.e., SAR or QSAR) sufficient for mechanistic interpretation?
  - In my opinion there is so much experimental scatter that it's hard to validate any sort of mechanistic interpretation
- Should models be linked to adverse outcome pathways? How can we facilitate that?
  - Are there AOPs assigned to enough chemicals?
  - I am not sure that will increase external prediction accuracy aside from increasing confidence by toxicologists in the predicted values

EPA
United States
Environmental Protection
Agency

Questions???

The views expressed in this presentation are those of the author and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency