An open source in silico prediction model of hepatic intrinsic clearance for in vitro to in vivo extrapolation

Jeremy Fitzpatrick, Pankajini Mallick, Kevin Bronson, Patrick McMullen
ScitoVation, LLC, RTP, NC

ORCID: 0000-0002-5401-9706
Email Us: science@scitovation.com

ScitoVation
PIONEERS IN CHEMICAL SAFETY ASSESSMENT

ABSTRACT

Chemical risk assessment is moving away from animal testing replacing it with computational and in vitro approaches in risk assessment. Dosimetry, by determining the amounts and key rates for distribution of a chemical in the body, is essential for the use of new approach methodologies in risk-based decision making. Use of models such as high-throughput in vitro to in vivo extrapolation (HT-IVIVE) has allowed dosimetry considerations to be more readily incorporated earlier in safety decision making processes. To keep pace with de novo computational approaches for estimating hazard, new tools are needed that can provide reasonable estimates of chemical clearance without requiring wet-lab experimentation and analytical chemistry. To this end we have developed a tool—using published metabolism information and physicochemical properties—to estimate a key determinant of compound dosimetry: intrinsic clearance. This tool allows a user to input a novel structure by hand or a series of structures using a chemical structure file such as a SMILES or SDF file. A prediction is then generated using a nearest neighbors model. The prediction is reported to be in or out of the model's domain based on its structural similarity to its nearest neighbors. The domain was initially determined by using a leave-one-out validation of a training set of 426 chemicals. The optimized model was then used to generate predictions for a test of 107 chemicals.

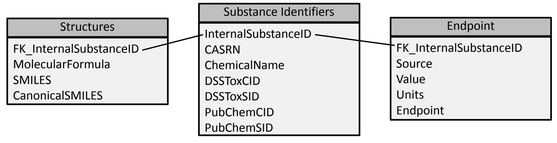
DATA

Data Sources

Source	Location	Data	References
Comptox Dashboard	https://comptox.epa.gov/dashboard	Structures, Chemical Identifiers	Williams et al. 2017
HTTK package	https://cran.r-project.org/web/packages/httr/index.html	Hepatic Intrinsic Clearance Values	Pearce et al. 2017, Wetmore et al. 2012, 2014

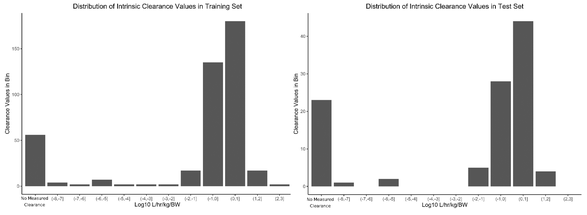
Data Source Table: Data sources, locations, and which data was taken from which source. Details on the values incorporated from each source are reported in the SQLite Database section.

SQLite Database



SQLite Database: This database incorporates data taken from EPA's Comptox Dashboard into the structures table. The canonical SMILES are generated using the ChemmineR package in R (Cao Y, et al. 2008). The substance identifiers table is populated with information from the EPA's Comptox Dashboard. The internal substance ID is an autoincrement key which allows the substance identifiers table to like to the structures table and endpoint table. The end point is populated with hepatic intrinsic clearance data from the HTTK package. (Pearce, R.G., et al. 2017)

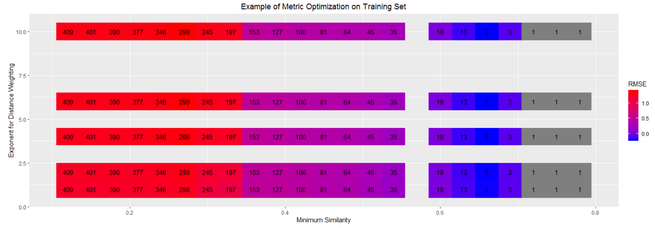
Intrinsic Clearance Data Distribution



Distribution of Hepatic Intrinsic Clearance Data: Before modeling the data was separated into a training set containing 426 chemicals and a test set containing 107 chemicals. The distribution of the measured intrinsic clearance values are given above. The bulk of the data falls between -2 and 2 log10(L/hr/kg/BW) with a significant portion also have no measured clearance.

METHODS

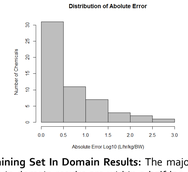
Several nearest neighbors' model were created to test the use of different metrics and to optimize the following: number of neighbors selected, minimum number of neighbors required, metrics for optimizing the inverse distance weighted average, distance to neighbors. The similarity of neighbors was based on the atom pair descriptors generated by the ChemmineR package (Cao Y, et al. 2008). We choose a minimum similarity based on the number of atom pairs that chemicals had in common over the total atom pairs which were needed to fully describe both molecules. We then set a minimum number of neighbors needed to make a prediction as well as the maximum number of chemicals used. The minimum number of neighbors needed to make a prediction is important to exclude outliers from having and outsized impact. Capping the maximum number of neighbors allows the clearance values for the nearest chemicals to be used for the prediction improving accuracy. Once the appropriate neighbors were selected, we used Shepard's method to calculate the inverse distance weighted average. This was used as our predicted intrinsic clearance value.



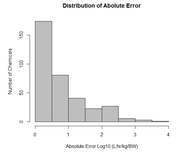
Optimization on Training Set: Visualization of selected data from optimizing the predictions based on the training set. The maximum and minimum chemicals selected in this case were set to 3. The y-axis is the exponent used for calculating the inverse distance a higher exponent gives a greater weight to closer chemicals. The x-axis is the minimum similarity required for chemicals to be used in a prediction. There must be enough chemicals to satisfy the minimum chemicals requirement and the similarity requirement. The number in each block in the amount of the 426 chemicals which could be predicted with the given requirements of our leave one out validation. The RMSE is given as the log10(L/hr/kg/BW) value.

RESULTS

Training Set Results

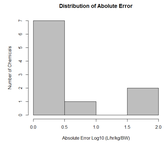


Training Set In Domain Results: The majority of the in domain results are within a half log unit of their actual intrinsic clearance value.

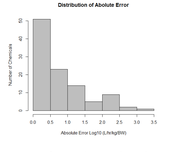


Training Set Out of Domain Results: About half of the out of domain results are within a half log unit of their actual intrinsic clearance value.

Test Set Results



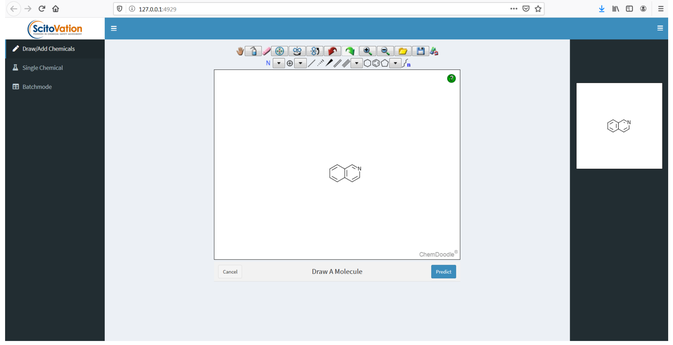
Test Set In Domain Results: The majority of results are predicted within a half log unit.



Test Set Out of Domain Results: The results are much more widely distributed for out of domain predictions.

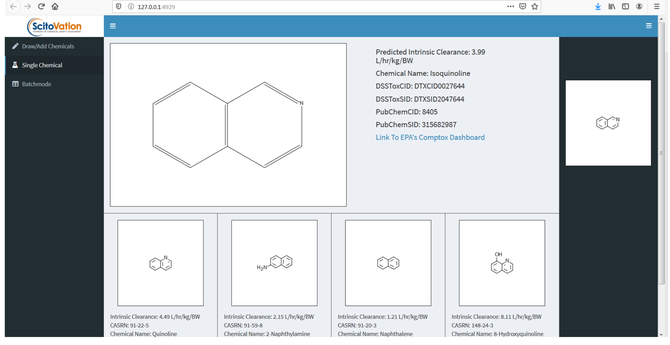
USING OUR MODEL

Single Chemical Input Interface



Single Chemical Input: The integrated ChemDoodle® interface allows the user to draw a structure using a basic of chemical tools. The input is the converted to a canonical SMILES which allows database to be queried for information on the chemical. The structure is also used to make a nearest neighbors' prediction based on criteria discussed in the methods section.

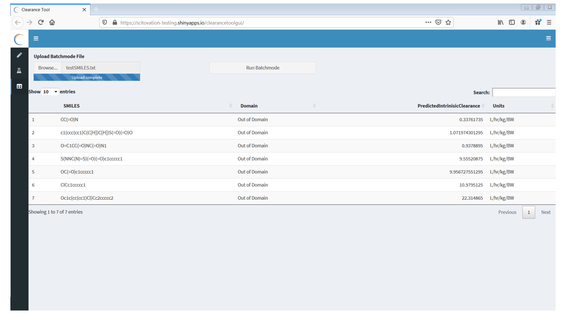
Single Chemical Output Interface



Single Chemical Output: The output gives the structure of the chemical for which the hepatic intrinsic clearance is being predicted. It also gives chemical identifiers if there are in our database. The four nearest neighbors are also reported below the test compound with their intrinsic clearance values on additional chemical identifiers. We also post a link when it is available to EPA's Comptox Dashboard.

USING OUR MODEL

Multichemical Input and Out Interface



Multiple Chemical Input and Results: The multiple chemical input file is loaded using the upload batch mode file box. Currently input files can be either sdf files or SMILES files. The output is then given as table. The table is searchable by the small box in the upper left-hand corner. It is currently being updated to include known identifier values and other information from our database on chemicals.

FUTURE IMPROVEMENTS

There are several major improvements which we are working on making to our model. The first is that we are currently working to add as much addition hepatic intrinsic clearance data to the database as possible. The second is that we are working to add additional chemical descriptor information such as physiochemical properties and other *in vitro* assay data to the database. The third is continuing to optimize the prediction methods used in the modeling, after which we will work to improve the user interface.

REFERENCES

Cao Y, et al. ChemmineR: a compound mining framework for R. Bioinformatics, 2008, 24(15), 1733–1734.

Pearce, R.G., et al., httk: R Package for High-Throughput Toxicokinetics. Journal of Statistical Software, 2017, 79(4): 1-26.

Wetmore, B.A., et al., Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicological Sciences, 2012, 125(1): 157-74.

Wetmore, B.A., et al., Incorporating population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicological Sciences, 2014, 142(1): 210-24.

Williams, A.J., et al., The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics, 2017, 9(61): 1-27

Acknowledgement

This work was made possible by funding from the American Chemistry Council – Long-research Initiative.

Try Out Our Tool

https://scitovation-testing.shinyapps.io/clearancetoolgui/