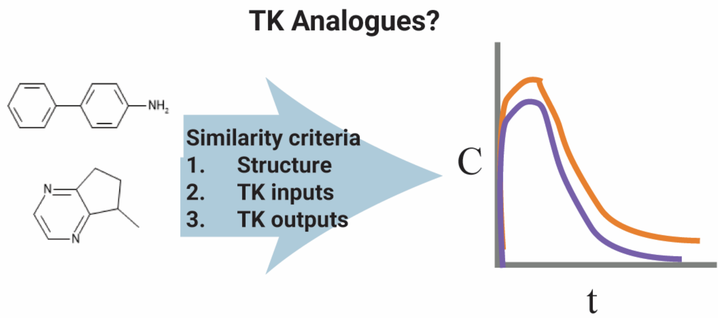
Clustering of environmental compounds based on structure and toxicokinetic properties

Michael Lawless, Rafał A. Bachorz, and Priyata Kalra
Simulations Plus, Inc. Research Triangle Park NC
michael.lawless@simulations-plus.com

4070/K569

INTRODUCTION



Traditional toxicokinetic (TK) models rely heavily on *in vivo* data, necessitating animal testing. At the same time, the scientific toolbox is expanding with new approach methodologies (NAMs) that do not rely on TK studies. One promising strategy (functional similarity[1] , TK-equivalence, or read-across[2]) involves leveraging known *in vivo* data of a *reference* compound to develop and validate a TK model for a *target* compound without additional animal data. Of course, this assumes the absence of activity cliffs. Here, we compare three methods (structure, TK inputs, and TK outputs) to find a reference compound (nearest neighbor (NN)) for the target compound. The similarity of the Cp-time curves of each pair, defined as a distance, is used to judge success.

METHODOLOGY

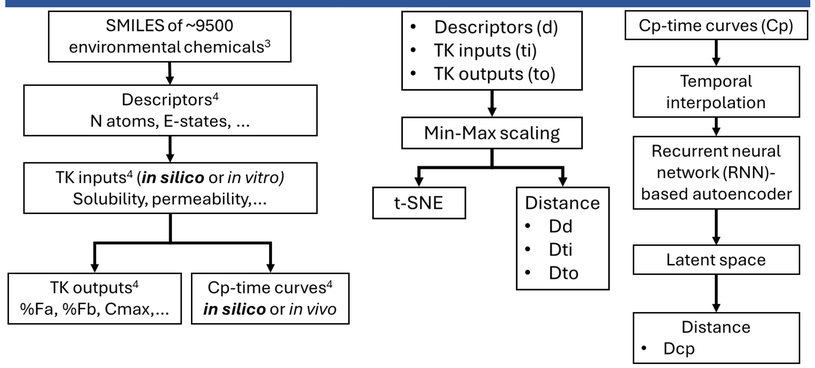


Figure 1 – Overview of workflow. Descriptors, TK inputs, and TK outputs are calculated with ADMET Predictor. There are scaled from 0 to 1 and distance is computed as the square root of the sum of squared scaled differences. The time scale of the Cp-time curves is standardized and input to an RNN-based autoencoder. The distance between two curves is calculated in the latent space.

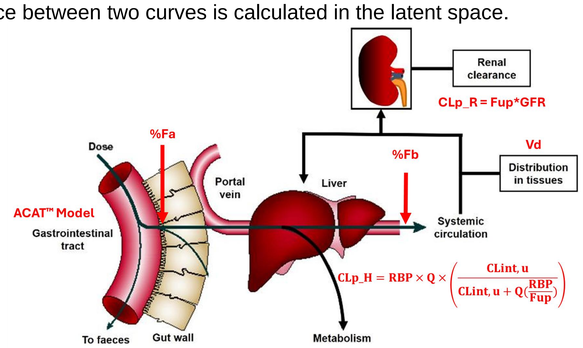


Figure 2 – ADMET Predictor HTPK simulations. A 70 mg immediate release tablet is dosed to a 70 kg human subject (1 mg/kg). The ACAT model is used to simulate the fraction absorbed (%Fa). First pass extraction (FPE) is the percent of compound metabolized during its initial passage through the liver. The oral bioavailability (%Fb) is the percent of the dose reaching systemic circulation. The portion that enters systemic circulation distributes into organs and tissues. Systemic clearance combines renal and liver clearance.

RESULTS

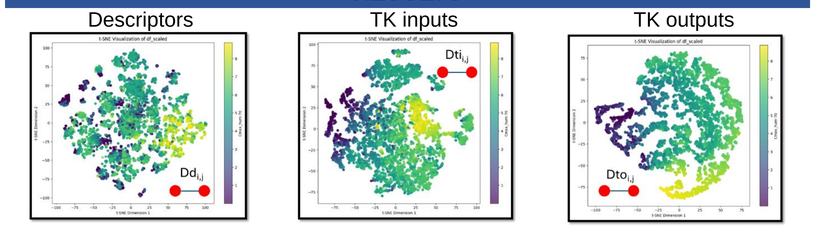


Figure 3 – t-SNE plots of various parameter spaces. The points are colored by the log of Cmax. TK outputs nicely clusters Cmax, the clusters are a little more scattered for TK inputs, and even worse for descriptors.

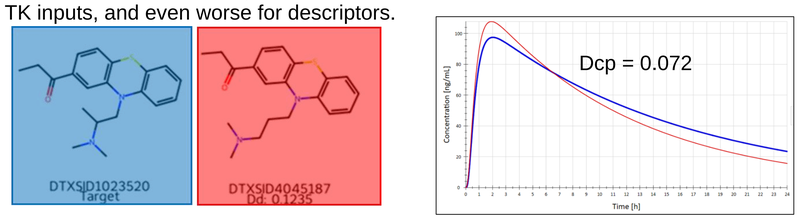


Figure 4 – The target's NN in **descriptor space** is 0.123. The Cp-time curves distance is 0.072. The %Fb, Cmax, and AUC are slightly different for the molecules.

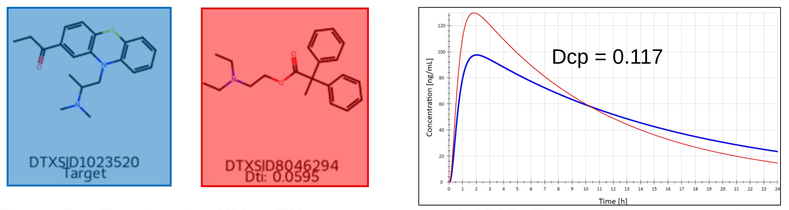


Figure 5 – The target's NN in **TK input space** is 0.0595. The native solubility pH and first pass extraction are the two largest contributors to the distance. The Cp-time curve distance is 0.117, so it is less similar to the target than the descriptor NN.

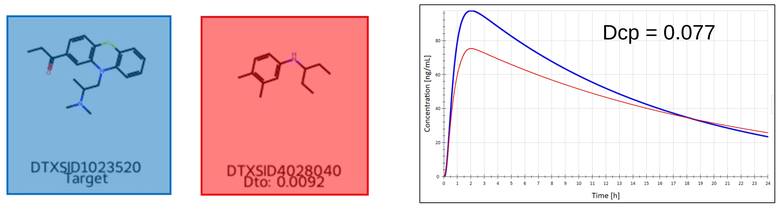


Figure 6 – The target's NN in **TK output space** is 0.0092. The Cp-time curve distance is 0.077. This is slightly higher than descriptor NN's Dcp.

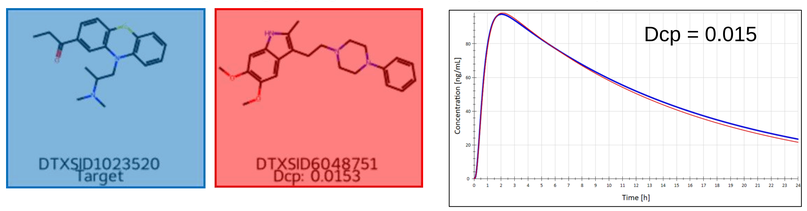


Figure 7 – The target's NN in **Cp-time space** is 0.015.

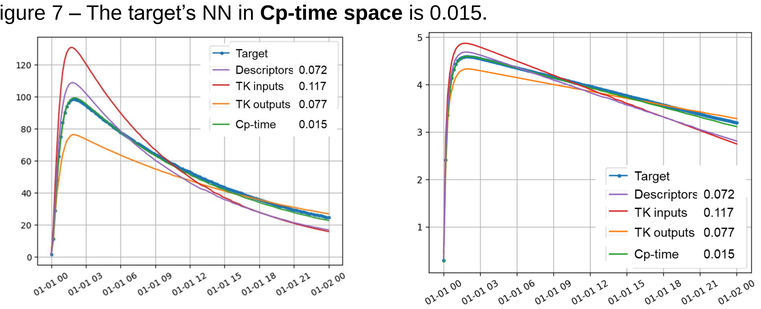


Figure 8 – Cp-time curves for the target and various neighbors. The y-axis is linear in the plot on the left and natural log on the right. As expected, the Cp-time NN (green line) overlays with the target curve the best. The descriptor NN is the next best match.

RESULTS

Structure	Identifier	NN-from	NN-distance	Cp-distance
	DTXSID00201257	Target	0.000	0.000
	DTXSID0021129	Descriptors	0.542	0.463
	DTXSID9041362	TK inputs	0.096	0.494
	DTXSID60927527	TK outputs	0.033	0.100
	DTXSID3046405	Cp-time	0.057	0.057

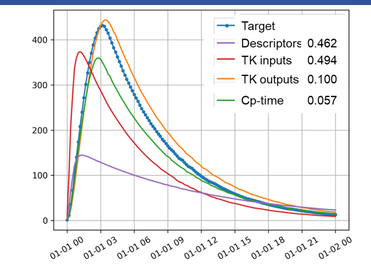


Figure 9 – The descriptor NN only differs from the target by a carbonyl group, but the Cp-time curves differ considerably. The predicted Vss is higher without the carbonyl which results in a lower Cmax. In linear space, the TK output curve (orange) looks like it matches better. However, in log space the green curve (Cp-time) matches better. This is due to the elimination phase where the green curve overlays on top of the blue curve.

Structure	Identifier	NN-from	NN-distance	Cp-distance
	DTXSID3023471	Target		
	DTXSID1046639	Descriptors	0.697	0.123
	DTXSID7048908	TK inputs	0.127	0.112
	DTXSID8022117	TK outputs	0.012	0.061
	DTXSID6048886	Cp-time	0.017	0.017

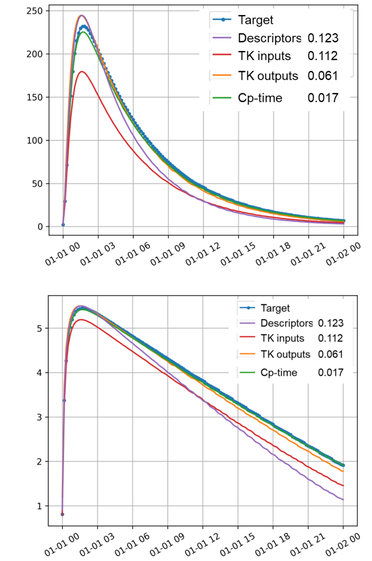


Figure 10 – The Cmax TK inputs NN (red curve) is considerably lower than the other chemicals. The elimination phase of the descriptor NN (purple curve) has the largest negative slope. The Cp-time NN (green curve) and TK outputs (orange curve) curves overlay nicely with the target. Both curves are about the same, each has a mismatch in different parts of the curve.

SUMMARY

Developing TK models in the absence of *in vivo* data is extremely challenging. We compared different methods that find the nearest neighbor of a target compound. The similarity of the Cp-time curves from each method was calculated from an RNN-based autoencoder. In the first and last examples in this poster, the descriptor NN has similar Cp-times to the target compound. In the second example, the descriptor NN was very similar structurally to the target, but the Cp-time curves were quite different. This is a *structural* "activity cliff", i.e., *visually* similar compounds have different Cp-time curves. The simulations allow one to determine which inputs to the TK simulations are responsible for differences in the Cp-time curves. Ideally, these inputs would be from *in vitro* experiments.

REFERENCES

[1] Ellison C. *Regulatory Toxicology and Pharmacology*, **2018**, 99, 61-77.
[2] Escher SE et al. (2021). *Read-Across Methodology in Toxicological Risk Assessment*. In: Reichl, FX., Schwenk, M. (eds) Regulatory Toxicology. Springer, Berlin, Heidelberg.
[3] Pearce RG et al. *J. of Statistical Software*, **2017**, 79(4), 1-25.
[4] ADMET Predictor version 12, Simulations Plus, Inc. Durham NC 27709