A framework for computing TK-equivalence using a large number of environmental compounds

Priyata Kalra, Rafał A. Bachorz, Michael Lawless
Simulations Plus, Inc. Research Triangle Park NC
CONTACT INFORMATION: Priyata.kalra@simulations-plus.com

SimulationsPlus

PURPOSE

The growing imperative to minimize animal testing has driven the adoption of New approach methodologies (NAMs), including read-across (RAx) methodologies[1].

Identifying truly equivalent reference compounds for TK read-across is challenging, as structural similarity alone does not ensure toxicokinetic equivalence.

We developed a novel framework for TK RAx that doesn't use new animal studies. By assuming TK equivalence is defined by plasma concentration-time (Cp-time) curve similarity, our method uses a Recurrent Neural Network (RNN) autoencoder to compress predicted high-throughput-pharmacokinetic curves (HTPK) into a latent space, finding nearest neighbors for comparison with traditional chemical descriptors and TK inputs.

OBJECTIVES

Develop a systematic framework for TK-equivalence in RAx.

Compare nearest-neighbor analogues across:

- Chemical descriptors
- Predicted TK inputs (e.g., pKa, logD, solubility) and TK outputs (bioavailability, Cmax, Tmax, AUC)
- Dynamic Cp-time profiles

Demonstrate how integrated similarity metrics improve reference compound selection for RAx.

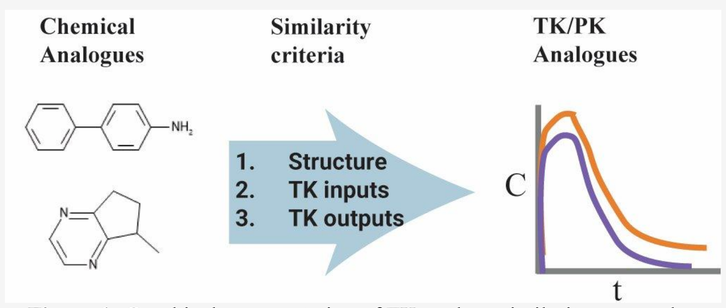


Figure 1: Graphical representation of TK analogs similarity approach

METHODS

Compound dataset and Properties prediction:

- 8117 environmental chemicals (MW < 900 Da, predicted absorption > 10%,) from httk[3] database.
- ADMET Predictor® 12[4] (AP-12) for all physicochemical and biopharmaceutical properties predictions
- HTPK simulations: 1 mg/kg dose, 70-kg human in ADMET Predictor® 12[4] were performed to generate Cp time curves

Similarity Approach:

Descriptors, TK inputs, and TK outputs are calculated with ADMET Predictor. All parameters are scaled to [0,1] range and Euclidean distance is computed between compounds on descriptors, TK inputs, TK outputs. The time domain of the Cp-time curves is standardized and is an input to the RNN-based autoencoder. The distance between two curves is calculated in the latent space.

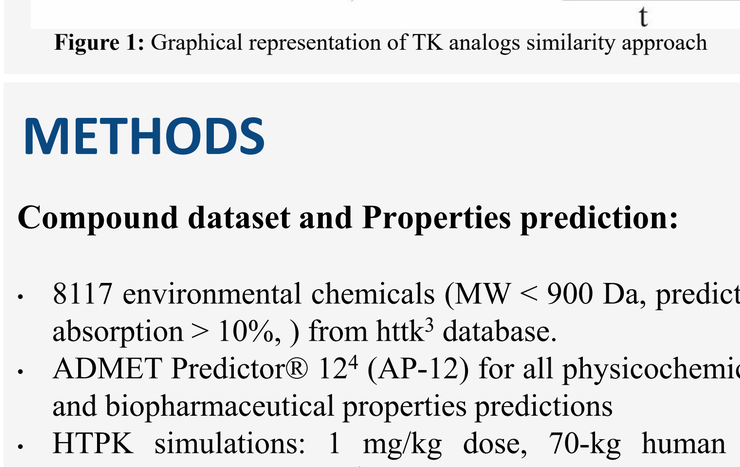


Figure 2: Graphical representation the similarity approach framework

RESULTS

t-SNE maps of chemical space, based on structural descriptors (Ddi,j), TK inputs (Dti,j), and TK outputs (Dtoi,j), demonstrate that the choice of distance metric fundamentally impacts how compounds are clustered. The degree of clustering and separation in each plot provides insight into the limitations of using structural or input-based metrics showing the clusters as heavily diffused.

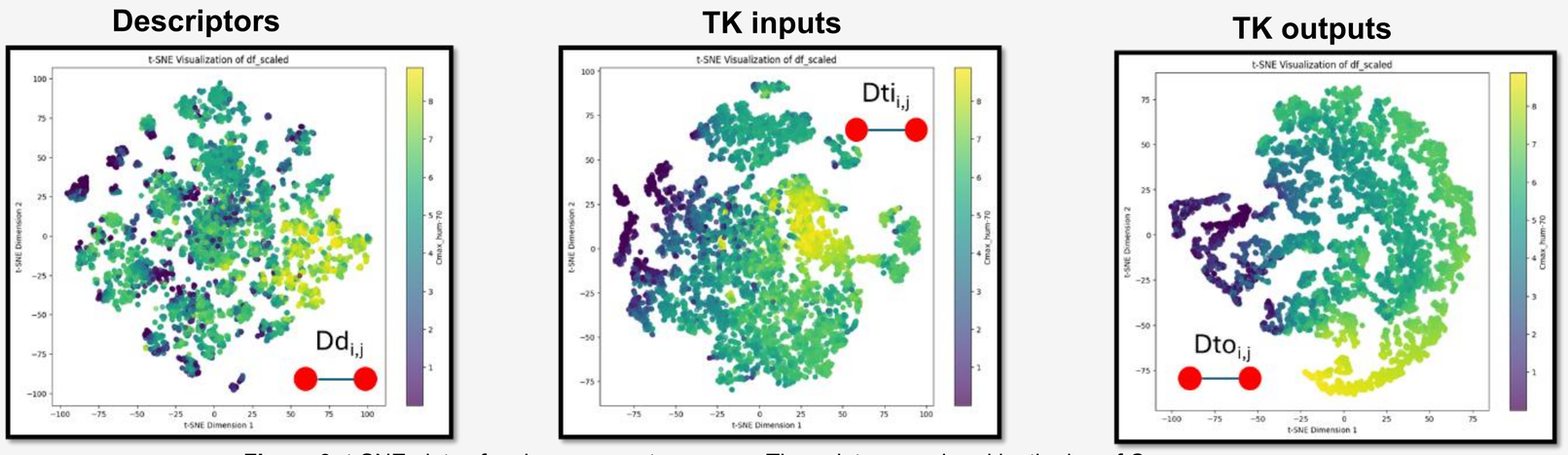
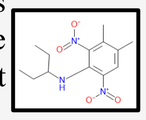


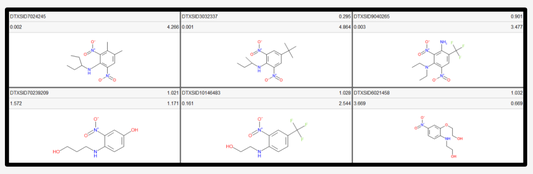
Figure 3: t-SNE plots of various parameter spaces. The points are colored by the log of Cmax.

CASE STUDY: Pendimethalin

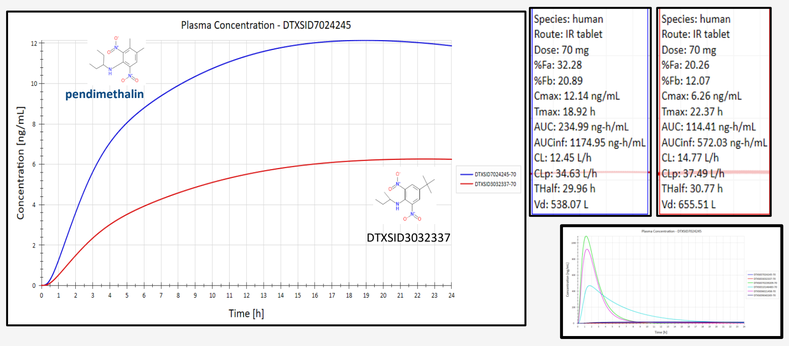
Pendimethalin – selective pre-emergent herbicide. Widely used but considered a persistent chemical in water and soil. It has been identified as chronically toxic to fish. Pendimethalin is positive in 8/11 of AP-12 Ames mutagenicity models. Here we found that structurally close analogues gave divergent Cp–time curves, while a less obvious analogue provided the best kinetic match.



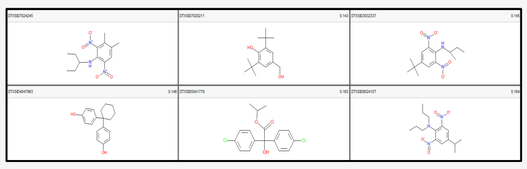
Pendimethalin Descriptor Nearest Neighbor (NN)



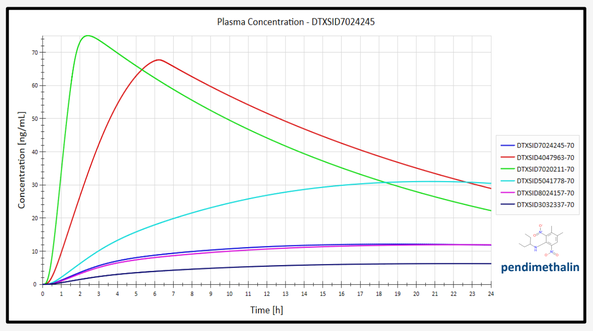
Structurally similar compounds can exhibit divergent TK profiles. Subtle side-chain differences impact key parameters like solubility and permeability. This can conceal activity cliffs where minor structural changes lead to dramatically different kinetic behaviors.



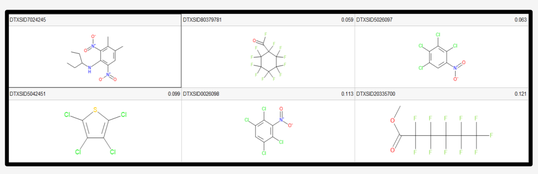
Pendimethalin TK input Nearest Neighbor (NN)



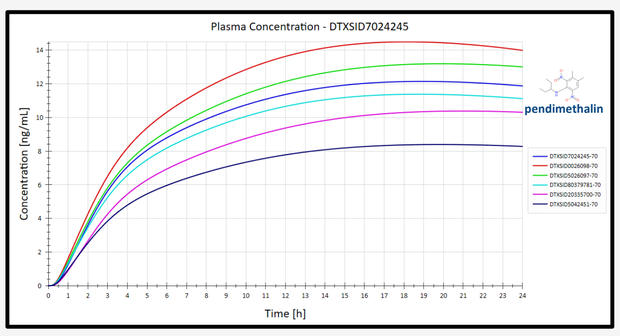
Using predicted input properties (**Peff, Clint, Sol** etc.) helps identify neighbors that reflect the underlying mechanistic drivers of TK. While this approach improves TK equivalence, it still shows significant mismatches. For e.g. DTXSID8024157 has the closest TK, but is furthest on euclidean distance.



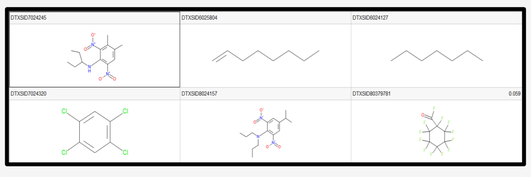
Pendimethalin TK Output Nearest Neighbor (NN)



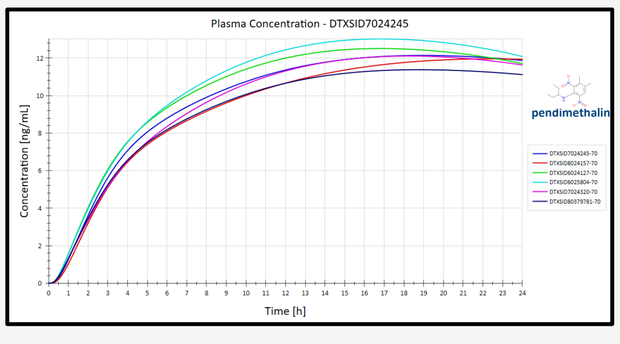
Using predicted output parameters (**AUC, Cmax** etc.) leads to the closer agreement in TK profiles. For instance, DTXSID80379781 and DTXSID5026097 show strong TK equivalence to the reference. The top 5 compounds show a fold difference in Cmax of ~2.3.



Pendimethalin Cp- time Nearest Neighbor (NN)



Using predicted HTPK Cp-time profiles with RNN encoders provide the most accurate TK read across. This method captures the entire Cp-time dynamics across both absorption and elimination phases, an improvement over other methods. **Latent space mapping enables the identification of structurally diverse compounds with highly similar TK profiles.**



MAIN FINDINGS AND FUTURE WORK

Developing TK models without in vivo data is challenging; we tested different methods to identify nearest neighbors and compared Cp-time curve similarity on predicted curves using AP-12.

Descriptor space prediction is fast and common, but it is not sufficient for TK equivalence. It misses kinetic similarity making it prone to activity cliffs where structural matches do not converge to TK equivalence. While TK inputs and outputs improve analogue selection, they miss full absorption-elimination dynamics.

Latent space Cp–time analysis using HTPK time curves provides the tightest TK equivalence, though results may shift once in vitro or in vivo data are incorporated.

Future work will test incorporation of real world in- vitro and in vivo data for a concrete comparison with previously known and reported TK analogs.

REFERENCES

[1] Ellison C. *Regulatory Toxicology and Pharmacology*, **2018**, 99, 61-77.
[2] Kohonen T. *Biol Cybernetics* **1982**, 43:59-69.
[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

EUROTOX ATHENS GREECE 14th–17th September 2025

SimulationsPlus