Toxicology in Vitro 110 (2026) 106150

Contents lists available at ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

Collaborative evaluation of *in silico* predictions for high throughput toxicokinetics

John F. Wambaugh [a], Nisha S. Sipes [a], Gilberto Padilla Mercado [a,b], Jon A. Arnot [c], Linda Bertato [d,e], Trevor N. Brown [c], Nicola Chirico [d], Christopher Cook [a,b], Daniel E. Dawson [a,f], Sarah E. Davidson-Fritz [a], Stephen S. Ferguson [g], Michael-Rock Goldsmith [a,h], Chris M. Grulke [a], Richard S. Judson [a], Kamel Mansouri [i], Grace Patlewicz [a], Ester Papa [d], Prachi Pradeep [a,j], Alessandro Sangion [c], Risa R. Sayre [a], Russell S. Thomas [a], Rogelio Tornero-Velez [a], Barbara A. Wetmore [a], Michael J. Devito [a,*]

[a] *Center for Computational Toxicology and Exposure, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, NC 28311, USA*
[b] *Oak Ridge Institute for Science and Education, Oak Ridge, TN 38331, USA*
[c] *ARC Arnot Research and Consulting Inc, Toronto, Ottawa, Canada*
[d] *Department of Theoretical and Applied Sciences (DiSTA), University of Insubria, Varese, Italy*
[e] *Department of Science and High Technology (DiSAT), University of Insubria, Como, Italy*
[f] *Integral Consulting, Seattle, WA 98104, USA*
[g] *Division of Translational Toxicology (DTT), National Institute of Environmental Health Sciences, Research Triangle Park, NC 2770927709, USA*
[h] *Congruence Therapeutics, Chapel Hill, NC 27517, USA*
[i] *NTP Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC 28317, USA*
[j] *German Federal Institute for Risk Assessment (BfR), Berlin-Jungfernheide, Germany*

ARTICLE INFO

Editor: Dr. P Jennings

Keywords:
Toxicokinetics
PBTK
QSAR
QSPR
HTTK
In silico
ADME

ABSTRACT

High throughput toxicokinetic (HTTK) methods address chemical risk assessment data gaps but require chemical-specific values that can be obtained by *in vitro* measurements or *in silico* models. In this study, seven quantitative structure property relationship (QSPR) models were used to estimate intrinsic hepatic clearance (Cl_{int}), fraction of chemical unbound in plasma (f_{up}), and/or TK elimination half-life ($t_{½}$). Performance of the QSPR models was evaluated using literature time-course *in vivo* TK data, mainly from rats. Simulations of the *in vivo* data were made with a high throughput physiologically based TK (HT-PBTK) model using the different QSPR model predictions as inputs. We estimate that using rat *in vivo* data to evaluate QSPR models trained on human *in vitro* measured data might inflate error estimates by as much as root mean squared $\log_{10}$ error (RMSLE) 0.8. A sensitivity analysis showed that Cl_{int} and f_{up} parameters inform predictions of area under the curve (AUC) and steady-state concentration (C_{ss}). We estimate that AUC can be predicted by HTTK with RMSLE 0.9 using *in vitro* measurements and 0.6–0.8 using QSPR model values. We anticipate that, for some novel compounds, QSPRs for HTTK input parameters will give predictions of TK similar to those based on *in vitro* measurements.

1. Introduction

Toxicokinetics (TK) describes chemical absorption, distribution, metabolism, and excretion (ADME) by the body as a function of time (O'Flaherty, 1981). TK is critical information for assessing health risks posed by chemical exposures (Rotroff et al., 2010; Tonnelier et al., 2012). While TK has been part of human pharmaceutical safety assessment for decades, the data requirements under most non-pharmaceutical chemical regulatory frameworks do not consistently include TK. Regulations such as the European Union's Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) and the United States's Toxic Substances Control Act (TSCA) have no specific

* Corresponding author.
E-mail address: devito.michael@epa.gov (M.J. Devito).

https://doi.org/10.1016/j.tiv.2025.106150
Received 15 July 2025; Received in revised form 15 September 2025; Accepted 19 September 2025
Available online 20 September 2025

requirements to generate TK data (Reale et al., 2024; U.S. Congress, 2016). However, TK information is valuable in 1) the interpretation of biomonitoring data (Reale et al., 2024; Sobus et al., 2011), 2) dosimetric anchoring of animal toxicity studies (National Research Council, 1983), 3) substantiating read-across justifications (Escher et al., 2019), and especially 4) quantitative *in vitro-in vivo* extrapolation (QIVIVE) (Wetmore, 2015) of New Approach Methods (NAMs) data (Kavlock et al., 2018) for next generation risk assessment (NGRA) (Moxon et al., 2020; Paul Friedman et al., 2020; Paul Friedman et al., 2025; Punt et al., 2020; Weitekamp et al., 2025).

Of the many thousands of substances that exist in commerce or in the environment, only a small percentage have been characterized in terms of their toxicity and of these, an even smaller portion have relevant TK information (Bell et al., 2018; Chang et al., 2022). Animal studies have been traditionally used to generate TK data. However, given the large number of environmental substances lacking data, the resources required to generate TK data using animals are neither practical nor desirable from an ethics perspective (Fentem et al., 2021). This is especially true given the call to reduce vertebrate testing under regulatory frameworks such as REACH (Lillicrap et al., 2016) and TSCA (U.S. Environmental Protection Agency, 2018).

"High Throughput Toxicokinetics" (HTTK) is a combination of efficiently-obtained chemical-specific data and chemical-agnostic mathematical models (Breen et al., 2021). Pre-clinical pharmaceutical researchers developed HTTK to prospectively estimate key TK summary statistics (such as area under the plasma concentration curve or AUC) (Wang, 2010). HTTK data typically consist of specific high throughput, *in vitro* measures of chemical-specific TK parameters. HTTK models consist of generic "high throughput" PBTK (or HT-PBTK) mathematical models designed for use with HTTK data and physico-chemical properties (including hydrophobicity and ionization equilibria). We refer to the chemical-specific TK parameters needed for HTTK models as "HTTK parameters".

Hazard NAMs such as *in vitro* high throughput screening assays have provided bioactivity data for thousands of chemicals (Jeong et al., 2022; Richard et al., 2016; Thomas et al., 2018). With IVIVE it is possible to use hazard NAMs to predict the exposures necessary to produce *in vivo* tissue concentrations equivalent to concentrations that have been found to be bioactive *in vitro* (that is, "reverse dosimetry") (Paul Friedman et al., 2020; Paul Friedman et al., 2025; Tan et al., 2006). Given that hazard NAMs provide data on thousands of chemicals, HTTK is needed to permit IVIVE for similar numbers of chemicals (Bell et al., 2018; Chang et al., 2022; Wetmore, 2015).

The U.S. National Academies of Science, Engineering, and Medicine have recognized that HTTK enables "first-tier risk-based rankings of chemicals on the basis of margins of exposure—the ratio of exposures that cause effects (or bioactivity) to measured or estimated human exposures" (National Academies of Sciences and Medicine, 2017). Further, the ability to simulate susceptible and highly exposed populations is important for incorporating TK into probabilistic risk assessments (Breen et al., 2022; Koman et al., 2019; Maertens et al., 2022). HTTK tools exist that can inform human health risk through simulation of human variability (Breen et al., 2022; Jamei et al., 2009a; Kreutz et al., 2024), susceptible life-stages (Kapraun et al., 2022; Truong et al., 2025), and varied exposure scenarios (including oral and inhalation) (Linakis et al., 2020). HTTK methods also exist for propagating chemical-specific uncertainty from *in vitro* measurements (Wambaugh et al., 2019) or *in silico* models for those measurements (Dawson et al., 2021).

To date, *in vitro* experiments have generated HTTK data for approximately 1000 chemicals (Black et al., 2021; Lynn et al., 2025; Paini et al., 2020; Rotroff et al., 2010; Tonnelier et al., 2012; Wambaugh et al., 2019; Wetmore et al., 2015; Wetmore et al., 2013; Wetmore et al., 2012). Although this work has significantly enhanced the throughput and decreased both cost and animal usage, the number of chemicals requiring TK information is still large from a cost and time perspective. This challenge motivated the study presented here: evaluate quantitative structure-property relationship (QSPR) models that predict the HTTK parameters required to use HT-PBTK models. Note that while we use "QSPR" here, these models are often also referred to as quantitative structure-activity relationship (QSAR) models.

We describe the HTTK parameters of interest in Table 1. Multiple QSPR models have been developed that predict *in vitro* TK parameters such as intrinsic hepatic clearance (Cl_{int}) and fraction of chemical unbound in plasma (f_{up}) (Chirico et al., 2021a; Dawson et al., 2021; Kirman et al., 2015; Pradeep et al., 2020; Sipes et al., 2017). The QSPR models range from freely available, open-source models based on public data to proprietary models underpinned by large proprietary data sets. In addition, some QSPR models have been validated following OECD methods for applications in regulatory decision-making (Organisation for Economic Co-operation and Development, 2004, 2014) to directly predict *in vivo* human TK properties, such as the terminal elimination half-life ($t_{½}$) (Arnot et al., 2014; Papa et al., 2018). While directly relevant to *in vivo* conditions, predictions of properties like *in vivo* $t_{½}$ do not yet permit the simulation of human variability that is possible by HT-PBTK models parameterized with more basic *in vitro* TK parameters such

Table 1
Relevant Toxicokinetic Parameters – *Indicates parameters that are typically predicted from combinations of other, more fundamental parameters. CL_{GFR} is approximated as f_{up} * Q_{GFR} (the glomerular filtration rate in L/h/kg bw). $CL_{exhalation}$ is approximated as $Q_{alv}/K_{blood:air}$ (the alveolar blood flow divided by the blood:air partition coefficient). k_{elim} is the terminal elimination rate of chemical from the body.

HTTK Parameter	Definition	Units	Typical Measurement	Typical Calculations
Cl_{int}	*In vitro* measured Intrinsic Hepatic Clearance	μL/min/10^6 hepatocytes	*In vitro*	Scaled to whole liver using liver volume and hepatocellularity
f_{up}	*In vitro* measured Fraction unbound in plasma	Fraction	*In vitro*	None needed
$t_{½}$	Chemical TK half-life	hours	*In vivo*	$t_{½} = \ln(2)/k_{elim}$
C_{ss}	Steady state plasma concentration resulting from 1 mg/kg bw/day	mg/L (per mg/kg bw/day)	*In vivo*	$C_{ss} = f_{bio} / CL_{tot}$
C_{max}	Peak plasma concentration	mg/L	*In vivo*	Estimated from CvT curve
AUC	Time-integrated plasma concentration	mg*h/L	*In vivo*	Estimated from CvT curve
V_d	Volume of distribution	L/kg body weight (bw)	*In vivo*	Can be estimated from f_{up}, logP, ionization, and tissue composition
Cl_{hep}	Effective liver clearance	L/h/kg bw	Calculated	Can be estimated from Cl_{int}, f_{up}, hepatic blood-flow
CL_{tot}	Whole body clearance	L/h/kg (bw)	*In vivo*	$CL_{tot} = V_d$ * ln (2)/$t_{½}$ or, for model "gas_pbtk": $CL_{tot} = CL_{hep} + CL_{GFR} + Cl_{exhalation}$
f_{bio}	Systemic oral bioavailability	fraction	*In vivo*	$f_{bio} = AUC_{oral}/AUC_{iv}$
k_{gutabs}	Oral absorption rate	1 / h	*In vivo*	

as Cl_{int} and f_{up}. Each QSPR model has its own training and testing datasets comprising a range of pharmaceutical and non-pharmaceutical chemicals with different physicochemical properties. Distinct performance metrics used for optimizing each QSPR. Comparing and contrasting the ability of these different models is challenging due to differences in how the QSPR models were built.

Here, we have evaluated seven QSPR models developed by six modeling groups. We started from an initial list of researchers interested in HTTK who participated in the ExpoCast community of practice (Wambaugh and Rager, 2022) and Tox21 (Thomas et al., 2018). A presentation was then made to 19th International Workshop on (Q)SAR in Environmental and Health Sciences in June 2021 to further solicit involvement from the HTTK QSPR community. Participants represent different international academic, regulatory, and commercial entities. Five modeling groups used QSPR models to predict Cl_{int} and/or f_{up} (the TK parameters typically measured *in vitro*) while two groups predicted $t_{½}$ (typically measured *in vivo*). The QSPR modeling groups were provided with the chemical identities (including their chemical structures) and SMILES structure descriptors (Weininger, 1988) and OPEn (quantitative) structure-activity/property Relationship App (OPERA) physicochemical predictions (Mansouri et al., 2018) but were not provided with the actual *in vivo* evaluation data. The QSPR model predictions, where applicable, were then evaluated with *in vitro* and *in vivo* data in a series of contexts outlined in Fig. 2 and Table 2.

2. Materials and methods

Table 2 summarizes the data and the statistical approaches used for evaluation. Table 3 describes the human HTTK parameter QSPR models that were evaluated. QSPR evaluation was conducted at three levels. Level 1 compares the QSPR-predicted HTTK parameters to the *in vitro* measured values for a subset of chemicals where measured values are 1) available and 2) not included in a QSPR model's training set. The Level 2 analysis evaluates the goodness of fit of QSPR-parameterized HT-PBTK simulations of plasma and blood concentration *vs.* time (CvT) values. Next, Level 3 evaluates TK summary statistics (for example, peak concentration C_{max}, AUC, $t_{½}$) derived using the QSPR models. The Level 2 and 3 analyses use *in vivo* data mostly for rat (with some human) from the CvTdb (Sayre et al., 2020). Therefore, we also estimate the error introduced by using human-based HTTK QSPR models for rats *via* a set of chemicals with *in vitro* measured values for both rat and humans. A final analysis examined the sensitivity of the CvT data to HTTK parameters in the different phases of ADME by systematically substituting parameters optimized to the *in vivo* data with HTTK-derived values.

All analyses were performed in the free, open-source statistical analysis language R (R Core Team, 2025) v4.6.0 (the current developer version). Heatmaps were generated using ggplots::heatmap.2 (Warnes et al., 2024) with data clustered on the basis of Euclidean distance *via* base R function stats::dist (R Core Team, 2025). All code is documented using RMarkdown (Baumer and Udwin, 2015) and available as supplemental material at https://github.com/USEPA/CompTox-ExpoCast-HTTKQSPRs.

2.1. Evaluation data and analysis

2.1.1. Initial chemical list

Varying amounts of chemical-specific data were available for the different evaluations performed. Table 2 describes the number of chemicals and types of evaluation performed at each step. The specific chemicals studied are available in Supplemental Table 1. To be included for analysis, we required that time course *in vivo* blood or plasma concentration data be available for either rat or human following single oral gavage and/or intravenous dose(s). For each chemical at least one study was available in either rat or human. In some cases, multiple studies or species were available. Using the initial public release of CvTdb, such data were available for 101 chemicals (Sayre et al., 2020). As discussed in the Data Filtering sections below, only a subset of these chemicals met suitability requirements for QSPR model evaluation (Supplemental Table 2). The *in vivo* measured time-course values are available as Supplemental Table 3. As described later in the Methods, chemical- and species-specific parameters were estimated for empirical (compartmental) TK models. As described in Results, chemical-species

Table 2
Chemical counts and evaluations.

Step	Description	Chemicals for Evaluation	QSPR Models Evaluated	Evaluation Data Reference
Challenge CvT Data Preparation				
Initial Chemical List	Chemicals with mostly rat CvT data following oral and/or intravenous dosing. Modelers were provided with chemical identity and physico-chemical properties to make human HTTK QSPR model predictions.	101	None	Sayre et al. (2020)
CvT Data Suitability Evaluation				
CvT Pre-Screen	Chemical data sets were eliminated for having too few points above the LOD, being radiolabeled, or failing CvTdb quality assurance	84	None	Padilla Mercado et al. (2025)
CvT Data Filtering	Chemical data sets that could not be described with a one- or two-compartment empirical TK model were withheld from further analysis.	81 with good empirical model fits	None	Padilla Mercado et al. (2025)
Investigation of Interspecies Bias				
Interspecies Evaluation	Chemicals with Cl_{int} and F_{up} measured in both rat and human were used to determine the impact of using human QSPR models to make predictions in rat	115	None	Honda et al. (2019); Wetmore et al. (2013)
Collaborative QSPR Evaluation				
Level 1	Direct evaluation of HTTK QSPR model predictions using *in vitro* TK Measurements (f_{up}, Cl_{int})	50 with measured *in vitro TK* parameters and CvT Data	4 (f_{up}), 5 (Cl_{int})	Pearce et al. (2017b)
Level 2	Evaluation of predictions for full TK time course using HT-PBTK with HTTK QSPR (all time points)	Up to 81, depending on QSPR	5 + Ensemble	Sayre et al. (2020)
Level 3	Evaluation of predicted TK summary statistics (C_{max}, time-integral/AUC, V_d, $t_{½}$, Cl_{tot}, C_{ss})	Up to 81, depending on QSPR	6 + Ensemble	Sayre et al. (2020)

QSPR models evaluated.

Table 3
QSPR models evaluated (note fup is unitless).

Model	Predictions	Original Units	Modeling Approach	Reference
Simulations Plus ADMET Predictor®	f_{up} and Cl_{int}	f_{up}: Unitless Cl_{int}: μL/min/mg of microsomal protein	Sum of CYP-specific Artificial Neural Network Ensemble (ANNE)	Sipes et al. (2017)
Pradeep 2020	f_{up} and Cl_{int}	f_{up}: Unitless Cl_{int}: μL/min/10^6 hepatocytes	Random forest and support vectors method	Pradeep et al. (2020)
Dawson 2021	f_{up} and Cl_{int}	f_{up}: Unitless Cl_{int}: μL/min/10^6 hepatocytes	Random forest, clearance organized by categories	Dawson et al. (2021)
OPEn (quantitative) structure-Activity / property Relationship App (OPERA)	f_{up} and Cl_{int}	f_{up}: Unitless Cl_{int}: Log10 μL/min/10^6 hepatocytes (subsequent versions use arithmetic scale)	Weighted distance K Nearest-neighbors (kNN)	Mansouri et al. (2021); Mansouri et al. (2018)
Iterative Fragment Selection QSAR (IFS-QSAR)	$t_{½}$	h	Fragment-based Multiple Linear Regressors (MLR)	Arnot et al. (2014)
QSAR by the Insubria Group (QSARINS-Chem)	$t_{½}$	h	Ordinary Least Squares MLR	Chirico et al. (2021b); Papa et al. (2018)
In vitro Biotransformation Prediction-Suite (IVBP-Suite)	Cl_{int} only	log_{10} mL/h/10^6 hepatocytes	Ordinary Least Squares MLR	Chirico et al. (2021a)

combinations that could not be adequately described by simple TK models were excluded from subsequent analysis.

2.1.2. Data filtering based on suitability for modeling

The CvTdb aims to fully document TK concentration *vs.* time experiments in the public domain. Some of these experimental data sets are not suitable for TK modeling. TK concentration data may be confounded by limit of detection (LOD) and limit of quantification (LOQ). LOD refers to the minimal signal the chemical analysis method can detect, while LOQ refers to the minimal signal that can be interpreted quantitatively as a chemical concentration. TK concentrations below the LOQ are qualitative and make estimating quantitative TK parameters difficult.

2.1.3. Data filtering based on empirical parameter estimates

Evaluation data were screened for suitability based on their ability to be described by empirical (that is, compartmental) TK models. Parameters were estimated for empirical one- and two-compartment TK models for chemicals with available CvT data using R package "invivoPKfit" (https://CRAN.R-project.org/package=invivoPKfit) (Padilla Mercado et al., 2025). Three models were considered: one- and two-compartment empirical TK models and a flat "null hypothesis" where there was no systematic change in concentration *vs.* time. The model with the lowest Akaike Information Criterion (AIC) value – indicating model parsimony – was selected (Akaike, 1974). Data sets where the flat model was selected were omitted from further analysis. The empirical model fit was then used for 81 chemicals as a "best case" prediction scenario for comparison with HT-PBTK models parameterized by either *in vitro* or QSPR model predictions.

For both one- and two-compartment models an elimination $t_{½}$ was calculated from the terminal elimination rate constant, as in Table 1. For the two-compartment model, the volume of distribution at steady-state ($V_{central} + V_{deep}$) was used as V_d. For both models, total clearance (Cl_{tot}) was calculated as $Cl_{tot} = V_d * k_{elim}$. The estimated TK parameters for both models are provided as Supplemental Table 4.

2.2. QSPR models evaluated

The QSPRs evaluated are summarized in Table 3. Individual QSPR model predictions are available in f_{up} from each QSPR. Table 4 summarizes the training sets for each QSPR modeling approach and predicted endpoint. Training sets were evaluated for relative fraction of pharmaceutical chemicals and number of CvT evaluation chemicals included. The identity of pharmaceuticals was determined by presence among the 8600 pharmaceuticals on the ZINC 15 list (Sterling and Irwin, 2015), as accessed from the CompTox Chemicals Dashboard (CCD) (Williams et al., 2017).

Simulations Plus ADMET Predictor values are sourced from Sipes et al. (2017). In that study, f_{up} was estimated using ADMET Predictor's "S + PrUnbnd" model. Cl_{int} was calculated as the sum of Cl_{int} predictions for the five major drug metabolizing enzymes: CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. Each compound is first assessed by a classification model that determines whether it is a substrate or non-substrate of the enzyme. If classified as a substrate, the model predicts its

Table 4
Summary of the QSPR training sets. a: The training set is proprietary, consisting of pharmaceutical data. b: The identities of the training set chemicals were anonymized in the 1.0 alpha version of the software, and chemical identities could not be algorithmically cross-referenced.

				CvT Evaluation Chemicals	
	Endpoint	Number of Training Chemicals	Fraction of Training Chemicals that are Pharmaceuticals	Number of QSPR model predictions Inside Domain of Applicability	Number of Evaluation Chemicals in Training Set
ADMET	Cl_{int}	a	1	51	a
ADMET	f_{up}	a	1	51	a
Pradeep 2020	Cl_{int}	642	0.27	47	50
Pradeep 2020	f_{up}	1139	0.44	47	51
Dawson 2021	Cl_{int}	1600	0.18	50	45
Dawson 2021	f_{up}	1305	0.35	50	38
OPERA	Cl_{int}	1346	0.23	77	49
OPERA	f_{up}	3229	0.3	74	53
IFS-QSAR	$t_{½}$	1105	0.63	58	23
QSARINS-Chem	$t_{½}$	1105	0.63	58	23
IVBP-Suite	Cl_{int}	560	b	63	b

sites of metabolism; otherwise, Cl_{int} is set to zero. For atoms identified as metabolic sites, the software estimates individual Cl_{int} values, and the total Cl_{int} is obtained by summing these atomic values. Thus, ADMET Predictor's Cl_{int} predictions integrate three models: substrate classification, metabolic site prediction, and atomic Cl_{int} estimation.

The Cl_{int} and f_{up} models for Dawson et al. (2021) were constructed using the method of Random Forests (Breiman, 2001) and four different open source chemical descriptor sets: PaDEL Descriptors (Yap, 2011), OPERA (Mansouri et al., 2018) physico-chemical properties, ToxPrints (Yang et al., 2015), and MACCS (Guha, 2007). A continuous (regression) model was developed for f_{up}. A classification model was developed for Cl_{int} with categories (slow <3.9 μL/min/10^6 cells), moderate (3.9–9.3 μL/min/10^6), and fast (>9.3 μL/min/10^6). The transition point between the moderate and fast rates of metabolism (9.3 μL/min/106 cells) is biologically relevant as the typical rate of blood flow to the human liver. The training set included a mix of ToxCast and pharmaceutical chemicals. By design, the number of training chemicals varied between the three categories, such that greater emphasis was on identifying extremes (slow/fast) rather than emphasizing the most common, moderate class.

The Cl_{int} and f_{up} models for Pradeep et al. (2020) were constructed using an ensemble model based on LASSO regression (Tibshirani, 2011), support vector machine (Cortes, 1995), random forest, and neural network multiple layer perceptron (Schapire and Freund, 2012) with open source descriptors from PubChem fingerprints (Goldsmith et al., 2014), ToxPrints, and OPERA physico-chemical properties. Regression models were built for f_{up}, while both regression and classification models were built for Cl_{int}. Only the regression Cl_{int} model is evaluated here. The training set used focused on chemicals available through the R package "httk" as of 2018.

Both Cl_{int} and f_{up} OPERA models were built using datasets combined from different literature sources. After several rounds of automated and manual curation to reduce errors, variability, and outliers, the Cl_{int} and f_{up} datasets consisted of 1056 and 1873 chemicals, respectively. The QSPR modeling was conducted using k nearest neighbors (kNN) (Taunk et al., 2019) coupled with a genetic algorithm (Alhijawi and Awajan, 2024) to select the most relevant PaDEL and CDK (Steinbeck et al., 2006) descriptors for each of the endpoints.

The IFSQSAR (Arnot et al., 2014) and QSARINS-Chem (Papa et al., 2018) models for predicting human whole-body level $t_{½}$ were developed from a compilation and review of *in vivo* human adult $t_{½}$ data obtained from clinical trials and human biomonitoring studies for environmentally relevant chemicals including halogenated chemicals (Arnot et al., 2014). The dataset includes 1105 chemicals with $t_{½}$ spanning approximately 7.5 orders of magnitude from 0.05 h (0.002 d) for nitroglycerin to 2×10^6 h (83 000 d) for 2,3,4,5,2′,3′,5′,6′-octachlorobiphenyl. The same training (50 % of chemicals) and testing (50 % of chemicals) data sets were used for the validation of both models (Papa et al., 2018) and both models were developed following OECD guidance for the application of QSARs in regulatory decision-making (Organisation for Economic Co-operation and Development, 2004, 2014). The IFQSAR model is a fragment-based approach (Arnot et al., 2014), while the QSARINS-Chem model uses whole molecular descriptors (Papa et al., 2018).

The software *In vitro* Biotransformation Prediction-Suite (IVBP-Suite) (Chirico et al., 2021a) was created within the CEFIC LRI ECO44 Project, from curated *in vitro* hepatic clearance data collected from different literature sources (https://arnotresearch.com/eas-e-suite/). IVBP-Suite QSARs were developed for chemicals grouped according to their common reactivity pathway based on SMARTCyp ranks (Rydberg et al., 2010), which were tested for CYP3A4, CYP1A2, CYP2C19 substrates, and they should work also for isoforms CYP2A6, CYP2B6, and CYP2C8 (Zaretzki et al., 2012). Models were constructed using Multiple Linear Regression by Ordinary Least Squares (MLR-OLS) coupled with a genetic algorithm to select the most relevant descriptors which were calculated by PaDEL-Descriptors (Yap, 2011). Modeled datasets and QMRF reports for each model are available in the online version of the software (https://dunant.dista.uninsubria.it/qsar/?page_id=670). Predictions by consensus reported in this manuscript were generated by IVBP-Suite alpha version 1.0 (June 2021).

In Table 4 we indicate the number of evaluation chemicals with predictions inside each QSPR model's domain of applicability. In Table 4 we also identify how many of the evaluation chemicals were included in the various training sets. Since many of the evaluation chemicals were in the various training sets, we still included QSPR model predictions in the CvT evaluations when they do not appear to be direct retrievals of a training set value. We define "direct retrievals" as predictions for training set chemicals that are within 1 % of the measured value (as in a nearest-neighbor algorithm identifying the compound itself). We are interested in evaluating the QSPR model predictions if the method of prediction is the same as what would be used for a chemical without a measured value.

QSPR-specific predictions were incorporated into the HT-PBTK model by altering the values used by the httk R-package. httk draws values from a single table which stores the f_{up} and Cl_{int} values for all chemicals (httk::chem.phys_and_invitro.data). Whenever changes are made to the table, the httk functions subsequently proceed using the new values. The HTTK data can be returned to their default values *via* the command "httk::reset_httk()". By default, no QSPR predictions are included in the table. However, QSPR model predictions can be loaded with the commands "httk::load_sipes2017(overwrite = TRUE)", "httk::load_pradeep2020(overwrite = TRUE)", or "httk::load_dawson2021(overwrite = TRUE)" (Dawson et al., 2021; Pradeep et al., 2020; Sipes et al., 2017). The argument "overwrite = TRUE" is needed so that *in vitro* measured data are overwritten whenever a chemical-specific prediction is available. To facilitate comparisons, a custom function "clear_httk()" is included in the supplemental material which deletes all human Cl_{int} and f_{up} values.

2.3. Ensemble model

Ensemble predictions were constructed after removing predictions that appear to be direct retrievals of experimental values rather than QSPR model predictions (Supplemental Table 5). An ensemble QSPR was constructed for f_{up}. To make the discrepancies approximately normal, the logit transform was used to move f_{up} from the range [0,1] to $[-\infty,\infty]$. The ensemble f_{up} was then calculated as the inverse logit of the mean of the logit-transformed f_{up} values. The Cl_{int} QSPR differences originate with the various training sets and approaches used. Many of the Cl_{int} QSPRs are trained with human hepatocyte data reflecting the metabolism of one or more unidentified enzymes. ADMet, however, combines multiple CYP-specific models to identify the enzyme most likely responsible for the majority of metabolism. That is, for ADMet, metabolism is the sum of clearance by each enzyme, rather than an average across those enzymes that are substrates and those that are not. The ensemble Cl_{int} prediction was similarly constructed by using the maximum Cl_{int} value predicted by any QSPR. This ensemble approach represents a sort of logical "OR" gate or a "tenth man strategy" (Adday et al., 2023) for rapid metabolism – even when other predictors agree that the compound is stable, the job of remainder is to play "devil's advocate". Only if all approaches agree that there is no metabolism do we predict no/slow metabolism. In this way, metabolism prediction may be more similar to structure alerts for toxicity which might identify potential toxicity based on a single structural alert (Yang et al., 2020).

2.4. Analyses performed

2.4.1. Interspecies concordance

The evaluation data in Table 2 are largely drawn from *in vivo* experiments with rat, while the HTTK parameter QPSR models in Table 3 are trained to predict values for human. The HT-PBTK model scales the measured *in vitro* parameters to *in vivo* quantities using species-specific physiological data (Pearce et al., 2017b). To investigate the impact of the mismatch between using human HTTK QSPR model predictions for

Cl_{int} and f_{up} and rat *in vivo* data, a set of chemicals with *in vitro* measurements of Cl_{int} and f_{up} for both human and rat were investigated. The rat-specific HTTK data, reported by R package httk, were largely drawn from Wetmore et al. (2013), Honda et al. (2019), Black et al. (2021), and Lynn et al. (2025). Chemicals were limited to those where the Cl_{int} assays were approximately concordant – that is, only those chemicals where the assays either agreed that metabolism occurred (both rat and human $Cl_{int} > 0$) or agreed that there was no metabolism (rat and human Cl_{int} both $= 0$). Discordant chemicals were omitted from the analysis.

First, the error in using human-measured Cl_{int} and f_{up} to predict rat-measured Cl_{int} and f_{up} was characterized. Then the two sets of measurements were used to predict C_{max} and C_{ss}, in both cases using the same rat-specific physiological parameters and scaling. This allowed estimation of the error introduced by using human-derived HTTK measurements to make rat-specific HTTK predictions.

2.4.2. Level 1 analysis

Chemicals for which *in vitro* TK measurements (f_{up}, Cl_{int}) were available allowed direct evaluation of QSPR model predictions. These data are drawn from the peer reviewed scientific literature and included in the R package "httk". Predictions for chemicals whose data are known to be in the training set of a QSPR were removed from the Level 1 analysis but were used in Levels 2–4. Unfortunately, training sets were not available for all QSPR models; therefore, to estimate a chemical's inclusion in one of these unknown training sets, any prediction within 1 % of the measured value was flagged as a possible training set chemical. Unless the training set was available and it could be confirmed that the prediction was not a "direct retrieval" of a measured value from the training set, these predictions that were "too good" were removed from all subsequent analysis. See Supplemental Table 5 for predictions that were removed. The *in vitro* measured values are available in Supplemental Table 2. Kolmogorov-Smirnov tests for differences in the distribution of fold errors for the QSPR model predictions (also in Supplemental Table 2) were performed using R function *ks.test*.

2.4.3. Level 2 analysis

The chemical-specific HTTK QSPR parameters were then used in an HT-PBTK model. The HT-PBTK model predicted CvT curves as a function of outcomes in plasma (and other tissues, although the evaluation data were all plasma) for different species, dose route, dose amount, and observation times. The physiological aspects of the model were varied based on species-specific scaling estimates (Davies and Morris, 1993) in units of $L/h/kg^{3/4}$ for flows and L/kg for volumes. Body weights of 70 and 0.25 kg were used for human and rat, respectively.

R package "httk" (Pearce et al., 2017b) v2.6.2 was used to provide HT-PBTK model simulations for the Level 2 analysis. Note that we distinguish between the general research area of HTTK from the specific R package "httk" through capitalization. "httk" can parameterize a HT-PBTK model based on chemical-specific values for f_{up} (unitless) and Cl_{int} ($\mu L/min/10^6$ hepatocytes). The HT-PBTK model that was used ("httk" model "gas_pbtk" from Linakis et al. (2020)) consists of well-mixed compartments for the gut, kidney, liver, lung, and rest of body. The model allows exposure by oral gavage and intravenous doses, and clearance by glomerular filtration (kidneys), metabolism (liver), and exhalation (lungs). Though the model allows for inhalation exposures, this route of exposure was not present in the CvT data used; however, the gas model makes exhalation a possible route of clearance for more volatile chemicals. The model is parameterized for a chemical using f_{up} and Cl_{int} plus equilibrium tissue:plasma partition coefficients predicted with a modified Schmitt's method (Pearce et al., 2017a; Schmitt, 2008) including aspects of Peyret et al. (2010). The *in vitro* measured or QSPR model-predicted $Cl_{int}^{in\ vitro}$ (units of $\mu L/min/10^6$ hepatocytes) is scaled to a physiological Cl_{int}^{liver} (L/h/kg) using species specific values for liver volume, density, and hepatocellularity using function httk::calc_hep_clearance (Breen et al., 2021). Because portal vein flow from the gut to the liver is modeled explicitly, a well-stirred correction is not necessary for metabolism. Oral dosing is subject to first-pass metabolism by the liver before the compound distributes systemically. Oral absorption was characterized in terms of two parameters: fraction absorbed from the gut and gut absorption rate. Fraction absorbed was assumed to be 100 %, with no gut metabolism. The same absorption rate was used for chemicals based on the mean absorption rate observed across the chemicals profiled in Wambaugh et al. (2018). The model was simulated using command httk::solve_gas_pbtk with option default.to.human = TRUE. That is, since no rat-specific values are predicted by the models under evaluation, comparisons to data from rats were done using rat physiology but human *in vitro* TK parameters. Time-course HTTK simulations using the underlying QSPR-predicted f_{up} and CL_{int} were evaluated against rat and human *in vivo* plasma and blood concentrations. Observations were divided into "early" and "late" times by calculating a midpoint time equal to half the time of the latest recorded observation.

2.4.4. Level 3 analysis

The QSPR model predictions were evaluated by comparing the HT-PBTK predicted TK summary statistics with values estimated from the *in vivo*-calibrated empirical TK models. The TK summary statistics include C_{max}, AUC, C_{ss}, CL_{tot}, $t_{½}$, and V_d – see Table 1 for definitions. Based on QSPR-derived and *in vitro* measured parameters, these summary statistics were calculated using various functions of R package "httk". The V_d was calculated using the same calibrated Schmitt partition coefficient algorithm as for the HT-PBTK model, but with all tissues lumped into a single compartment. V_d depends largely on physico-chemical properties (see Sensitivity Analysis below) with some input from f_{up} (Pearce et al., 2017a) and was calculated using httk::calc_vdist. C_{ss} depends upon both Cl_{int} and f_{up} and was calculated using httk::calc_css(model="gas_pbtk") to include exhalation as a route of elimination for semi- and volatile chemicals. For CL_{tot}, the expression $C_{ss} = f_{bio} / CL_{tot}$ was converted to $CL_{tot} = f_{bio} / C_{ss}$ so that the hepatic bioavailability (first-pass hepatic metabolism from httk::calc_hep_bioavailability) and the inverse of the steady-state concentration yielded the effective total clearance rate. For k_{elim}, CL_{tot} was converted to elimination rate using the estimated V_d ($k_{elim} = CL_{tot} / V_d$).

The QSPR model predictions for human $t_{½}$ (Iterative Fragment Selection QSAR (IFS-QSAR) and QSARINS-Chem (Papa et al., 2018) were first converted to rat predictions using allometric scaling (the ratio of rat body weight, assumed to be 0.25 kg, to human, 70 kg, raised to the ¾ power). $t_{½}$ was converted to elimination rate and multiplied by the V_d from the ensemble model to estimate total clearance and C_{ss}. Summary statistics determined from the HTTK QSPR model predictions for TK statistics C_{max}, AUC, and C_{ss} were compared to values estimated from the *in vivo* data.

2.4.5. Benchmarks used

We use three types of benchmarks to contextualize the QSPR model performance. First, for a best-case performance, we use the empirical (one- or two-compartment TK model) predictions based on parameters optimized (or "fit") to the *in vivo* data, labelled as "*In vivo* Fits". The one-and-two compartment models are simpler than the HT-PBTK model used for all other scenarios, but because they have been optimized to the *in vivo* evaluation data itself, they are expected to outperform the other approaches here. Second, we use the HT-PBTK model with the actual *in vitro* measured Cl_{int} and f_{up} values. Finally, for worst case performance, we use y-randomization so that the *in vitro* measured Cl_{int} and f_{up} values across all chemicals in the R package "httk" library are scrambled and assigned to the incorrect chemicals, labelled "HTTK-YRandom". 25 different sets of y-randomized parameters are drawn for each chemical. We do not y-randomize the physico-chemical properties so as only to examine the parameters being predicted by the HTTK QSPR models in Table 3.

A key unknown of TK IVIVE is whether compounds are subject to restrictive or non-restrictive plasma protein binding. Some compounds are modeled as being restrictively bound to plasma (that is, only the free

fraction of chemical is available for metabolism) while others may be more likely to be non-restrictive (the compound is so weakly bound to the plasma that both the bound and unbound fractions are available for metabolism) (Scheife, 1989). The form for effective intrinsic (flow-independent) restrictive whole-liver clearance is:

$$Cl^{effective}_{intrinsic,restrictive} = f_{up} Cl_{intrinsic,liver}$$

while the effective non-restrictive clearance is not reduced by the unbound fraction in plasma:

$$Cl^{effective}_{intrinsic,nonrestrictive} = Cl_{intrinsic,liver}$$

Restrictive plasma binding is distinct from flow restriction, which is accounted for in the HT-PBTK model using an explicit and finite blood flow from the hepatic portal vein (Wambaugh et al., 2025). Determining restrictive *vs.* non-restrictive chemicals requires *in vivo* data, and very generally restrictive clearance appears to be more common for pharmaceuticals (Wetmore et al., 2012). R package "httk" defaults to restrictive clearance for all chemicals because, for reverse dosimetry applications, it is a more conservative assumption (Rotroff et al., 2010). However, when *in vivo* data are available, as with the level 2 and 3 evaluations, it is possible to make a judgement about the appropriate assumption for a set of chemicals.

2.4.6. Sensitivity analysis

Finally, we evaluated the relative sensitivity of overall prediction error to HTTK parameters in different phases of TK (Fig. 1). Empirical one- or two-compartment models were used. Compartmental model parameters were optimized to *in vivo* data and the better model was selected by invivoPKfit (Padilla Mercado et al., 2025) based on AIC. The *in vivo*-optimized compartmental model provided a baseline for a low error TK model. The source of the parameters for the compartmental models was then varied by TK phase. For each phase of TK (Fig. 1), parameters based on *in vivo* CvT data were replaced by those using the ensemble QSPR model predictions. This analysis used only the ensemble QSPR model predictions for HTTK parameters Cl_{int} and f_{up}. Sensitivity of error was calculated by sequentially substituting values derived from the ensemble HTTK model for the *in vivo*-estimated values for absorption (F_{bio} – httk:: calc_fbio.oral), distribution (V_d – httk::calc_vdist), or elimination rate (from Cl_{tot} and V_d) (k_{elim} – httk::calc_total_clearance). Calculation of F_{bio} was assumed to only include first-pass hepatic metabolism. First pass metabolism for F_{bio} depends on Cl_{int} and f_{up}, and was calculated using httk::calc_hep_bioavailability.

2.5. Evaluation metrics

At each level, multiple statistics were used to evaluate predictions

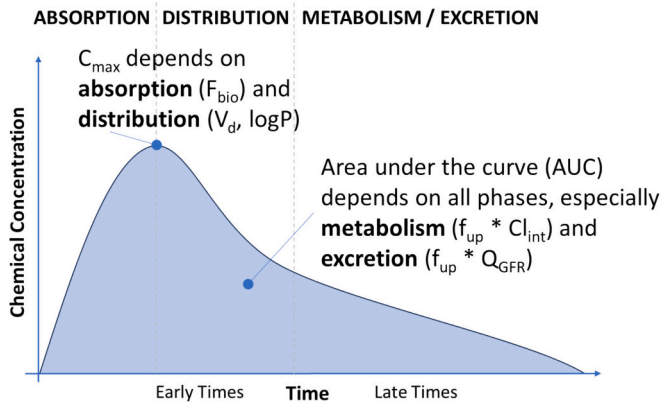

Fig. 1. Conceptual illustration of toxicokinetic concentration vs. time (CvT) data.

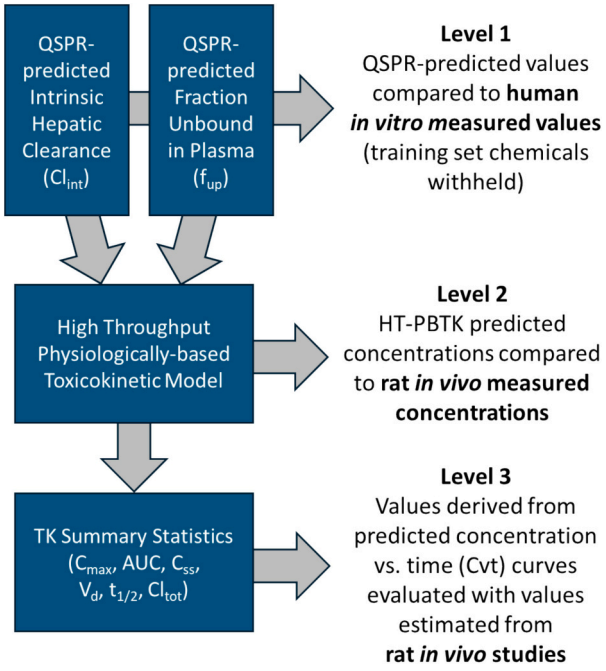

Fig. 2. Overview of the QSPR model prediction analyses performed. See also Table 2.

(*pred*) relative to observed values (*obs*) as appropriate. We note that if the predicted value is 0 then RPE = −1. Average fold error (AFE) was calculated as $AFE = \left\langle \log_{10}\frac{pred}{obs} \right\rangle$. Absolute Average Fold Error (AAFE) was calculated as $AAFE = 10^{\left(\frac{1}{n}\sum\left|\log_{10}\frac{pred}{obs}\right|\right)}$, where if pred = 0 and obs = 0 we assigned $\log_{10}\frac{0}{0} = 0$. Root Mean Squared Log Error (RMSLE) was calculated as $RMSLE = \sqrt{\frac{1}{n}\sum\left(\log_{10}(pred + LOD) - \log_{10}(obs + LOD)\right)^2}$.

3. Results

The primary focus of this analysis is to assess the accuracy of predicted CvT curves using HTTK parameters from QSPR models (this is the "Level 2" analysis). However, of interest to QSPR developers is an *ad hoc* comparison of the QSPR model predictions themselves to *in vitro* measured HTTK parameters where available (Level 1 analysis). Of interest to chemical risk decision makers is the analysis of the accuracy of CvT TK summary statistics (Level 3) based on QSPR model predictions. To place all these analyses in context, we first examine the amount of noise introduced by using human QSPR parameters to predict *in vivo* CvT data that are largely from rats. We conclude by analyzing the sensitivity of HT-PBTK model simulations to the HTTK parameters predicted by the QSPR models.

3.1. Evaluation chemicals and predictions

There are 101 chemicals present in the initial public release of CvTdb (Sayre et al., 2020) that had chemical concentration *versus* time data resulting from either single oral gavage or single intravenous doses given to rats or humans.

As illustrated in Fig. 1, a typical TK concentration *vs.* time curve following a single oral dose consists of separate phases: 1) an initial increase driven by absorption of the chemical into the body; 2) a steep decrease driven by distribution of the chemical from the blood into tissues with varying affinities for the chemical; followed by 3) metabolism and excretion of the chemical (elimination) from the blood. Metabolism and excretion are also present during the earlier phases. The

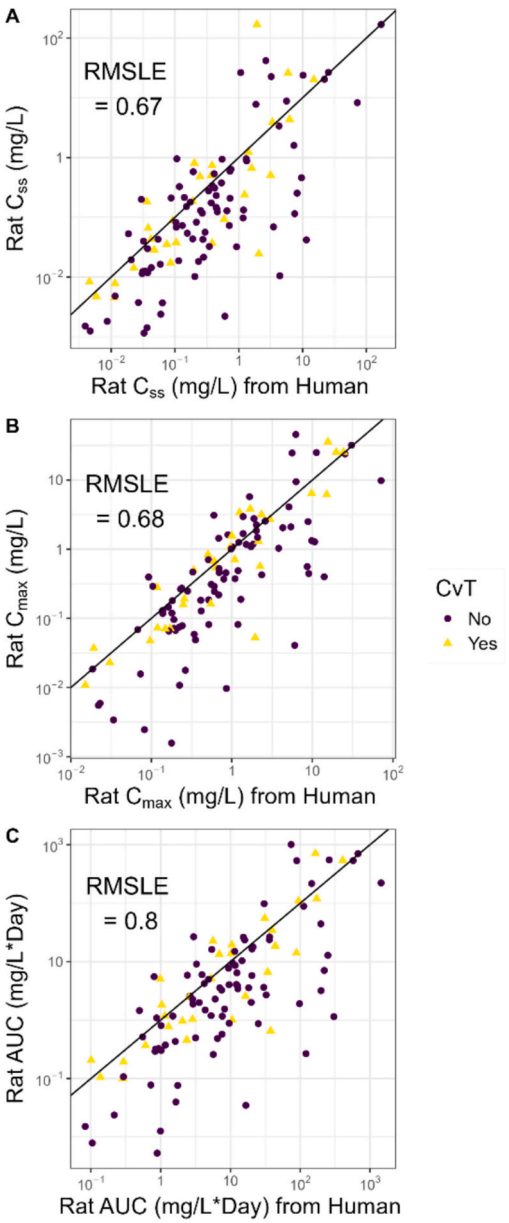

Fig. 3. Interspecies concordance of HT-PBTK model simulations for C_{ss}, C_{max}, and AUC for chemicals with parameters measured in both rat and human: (panel a) predicted steady-state plasma concentration (C_{ss}) predicted with HT-PBTK model parameterized using the *in vitro* measured f_{up} and Cl_{int}, (panel b) predicted peak plasma concentration (C_{max}), and (panel c) area under the concentration *vs.* time (CvT) curve (AUC). Point color and shape indicates whether CvT data were available for use in subsequent analysis here.

distribution phase can be more pronounced depending on how much chemical the tissues can sequester (that is, if V_d is larger). There is no absorption phase following an intravenous dose, although there may be a very short initial phase as the chemical is distributed from the site of injection to the systemic blood circulation. The different phases of TK vary in sensitivity to the parameters Cl_{int} and f_{up} analyzed here – see the sensitivity analysis below. The elimination phase is most sensitive because f_{up} allows the estimation of glomerular filtration in the kidney (excretion) and both Cl_{int} and f_{up} may inform the metabolic clearance in the liver (metabolism).

CvT data sets for eleven chemicals were eliminated in pre-processing before fitting. The eliminated chemicals were: Methanol, Tetrachloroethylene, Methyl tert-butyl ether, tert-Amyl methyl ether, Propylparaben, Phenazone, Flufenacet, Camphor, 2,3,7,8-Tetrachlorodibenzo-p-dioxin, Pentachloroanisole, and 2,3,4,7,8-Pentachlorodibenzofuran. In nine cases, a chemical-species combination has too few observations above the LOQ to allow meaningful modeling. Propylparaben was eliminated because there were no observations above LOQ. Pentachloroanisol had only two timepoints. 2,3,4,7,8-Pentachlorodibenzofuran had all measured concentration values below the LOD. In five cases, the TK data available were from radiolabeling studies or did not have concentration data collected from blood or plasma, which made them unsuitable for HTTK comparisons. The data for Phenazone (antipyrine in CvTdb) were from radiolabeled experiments which only allow estimation of elimination rate (that is, no concentration data are available). Data for four chemicals were excluded because they did not pass recent quality control within the CvTdb database (Sayre et al., 2020).

The CvT data sets were modeled for their suitability to evaluate HT-PBTK model simulations by systematically optimizing parameters for one- and two-compartment models (Padilla Mercado et al., 2025). Separate parameter estimates were made for each combination of compound and species for which there were data. We assumed data sets that could not be described by a one- or two-compartment model did not conform to the general profile depicted in Fig. 1. As shown in Tables 2, 3 out of 84 remaining chemicals were removed from the analysis for being inadequately described by an empirical model. That is, when empirical TK model parameters were estimated, data for the removed chemicals were best described by the "flat" model (indicating that the data are too noisy to estimate empirical TK parameters).

The data for each of the 81 remaining chemicals could be described using either a one- or two-compartment empirical models. CvT data resulting from intravenous dose regimens are needed to estimate all the parameters (in particular, V_d) required to make CvT predictions. There were 12 chemicals where we could not estimate V_d: CvT data resulting from both intravenous and oral dose regimens are needed to estimate f_{bio}. However, without f_{bio} we can still make predictions for intravenous doses. There were 24 chemicals where we could not estimate f_{bio}. Among the 81 remaining evaluation chemicals, 80 had data in rat and 3 in human. The three chemicals with the shortest $t_{½}$ in rat were Imipramine, Simazine, and Propamocarb hydrochloride – mean of 0.009 h. The two chemicals with longest $t_{½}$ in rat were Perfluorooctanoic acid and Potassium Perfluorohexanesulfonate whose mean of 1500 h was almost 40 times longer than the chemical with the next longest $t_{½}$. See Supplemental Table 2.

The 81 chemicals included (according to lists accessed 12/9/24) 25 pharmaceuticals (https://comptox.epa.gov/dashboard/chemical-lists/ZINC15PHARMA). Additionally, the evaluation chemicals included: 13 from the non-confidential Toxic Substances Control Act (TSCA) active inventory (https://comptox.epa.gov/dashboard/chemical-lists/TSCA_ACTIVE_NCTI_0224), 38 pesticide active ingredients (https://comptox.epa.gov/dashboard/chemical-lists/EPAPCS), 16 that are found in consumer products (https://comptox.epa.gov/dashboard/chemical-lists/EPACONS), 7 *per-* and poly-fluoroalkyl substances (PFAS) (https://comptox.epa.gov/dashboard/chemical-lists/PFAS8a7v3), and 73 that are part of the ToxCast screening program (https://comptox.epa.gov/dashboard/chemical-lists/ToxCast_invitroDB_v4_1). Note that a chemical could be in more than one of these categories.

3.2. Interspecies concordance

The evaluation data in Table 2 are largely drawn from rat *in vivo* experiments, while the QSPR models predicting Cl_{int} and f_{up} in Table 3 are trained to predict *in vitro* values using human plasma and primary hepatocytes. Only 27 evaluation chemicals had both human and rat *in vitro* HTTK data. However, there are 113 chemicals with *in vitro* measured f_{up} and Cl_{int} for both human and rat. The potential impact of the species mismatch was characterized using these chemicals. There were no *in vivo* data used in this interspecies analysis. This analysis allowed estimation of the error introduced by using human-derived

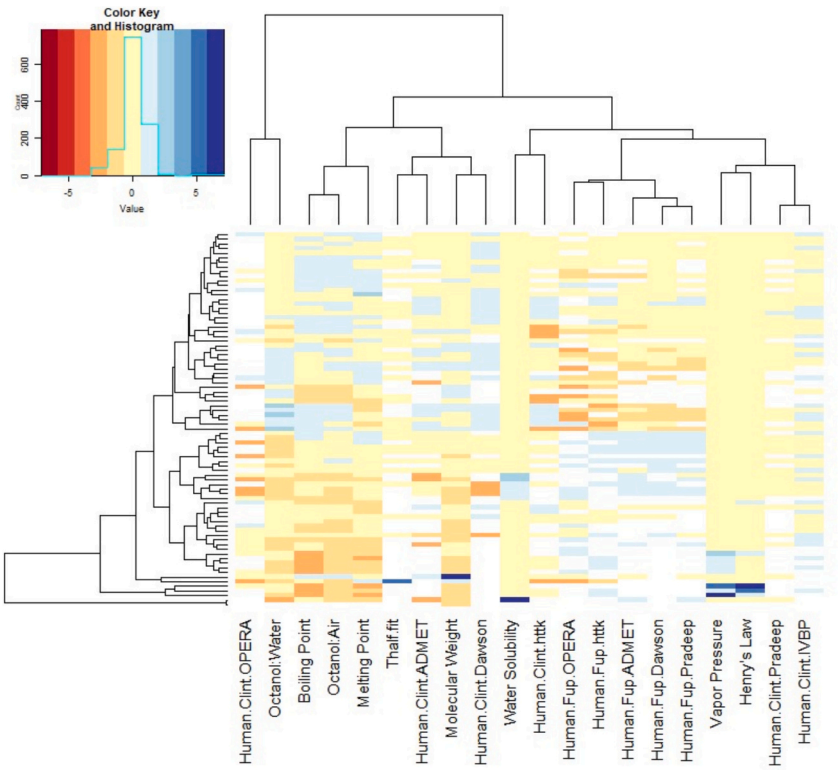

Fig. 4. Columns in this heatmap indicate the physico-chemical properties and measured/predictions for *in vitro* TK (intrinsic hepatic clearance, Cl_{int}, and fraction unbound in plasma, f_{up}). Each row corresponds to one of the 9292 chemicals. The *in vitro* TK measurements ("Human.Clint.InVitro" and "Human.Fup.InVItroInVItro") and predictions for these values from the various QSPR models (Table 3) are indicated by name. Data are normalized on a per column basis by centering (subtracting the mean) and scaling (by standard deviation). Thus, the "Value" of each entry in heatmap indicates the number of standard deviations from the mean. Blank values indicate no prediction. Hierarchical clustering was performed based on Euclidian distance.

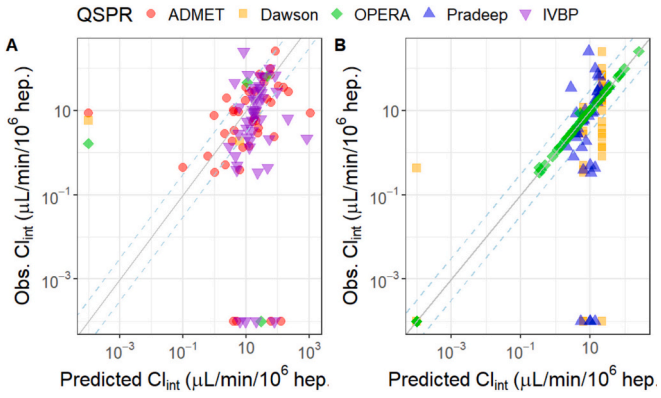

Fig. 5. Evaluation of Predictions for Intrinsic Hepatic Clearance (Cl_{int}). Zero values were plotted at 10^{-1}, the solid line indicates identity (1:1) while the dashed lines indicate 3.2-fold difference. In panel A we show chemicals external to the training set (except for models ADMET and IVBP). In panel B we show chemicals that were in the training set.

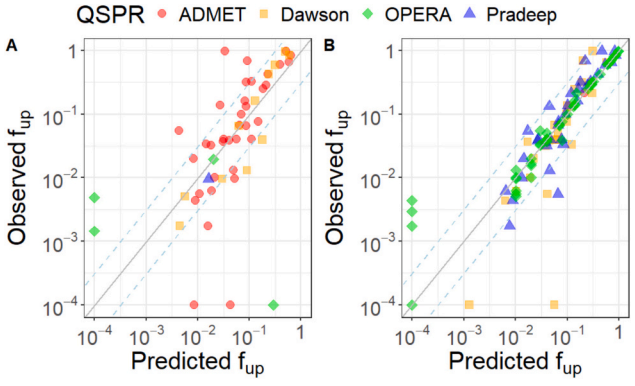

Fig. 6. Evaluation of Predictions for Fraction Unbound in Plasma (f_{up}). Zero values were plotted at 10^{-4}, the solid line indicates identity (1:1) while the dashed lines indicate 3.2-fold difference. In panel A we show chemicals external to the training set (except for model ADMET). In panel B we show chemicals that were in the training set.

HTTK measurements to make rat-specific HTTK predictions.

First, the error in using human-measured Cl_{int} and f_{up} to predict rat-measured Cl_{int} and f_{up} was characterized. As shown in Supplemental Fig. 1, *in vitro* measured values were roughly concordant. The rat-human differences were RMSLE 0.63 for f_{up} and 0.99 for Cl_{int}. There were chemicals with observed clearance in one species and not the other. RMSLE was calculated using only the non-zero Cl_{int} values.

Then the two sets of measurements were used to predict AUC, C_{max}, and C_{ss}. As shown in Fig. 3, RMSLE for C_{ss} was 0.67, RMSLE for C_{max} was 0.68, and for AUC RMSLE was 0.8. Errors were slightly smaller for the 27 chemicals with CvT data used in subsequent evaluations – RMSLE for C_{ss} was 0.6, C_{max} was 0.54, and AUC was 0.69. The RMSLE values indicate that the errors calculated using rat CvT data to evaluate predictions based upon human *in vitro* data may be overestimated due to interspecies differences.

3.3. Performance benchmarks

There were 56 chemicals with *in vitro* measured Cl_{int} and 57 chemicals with *in vitro* measured f_{up}, for a total of 57 unique chemicals.

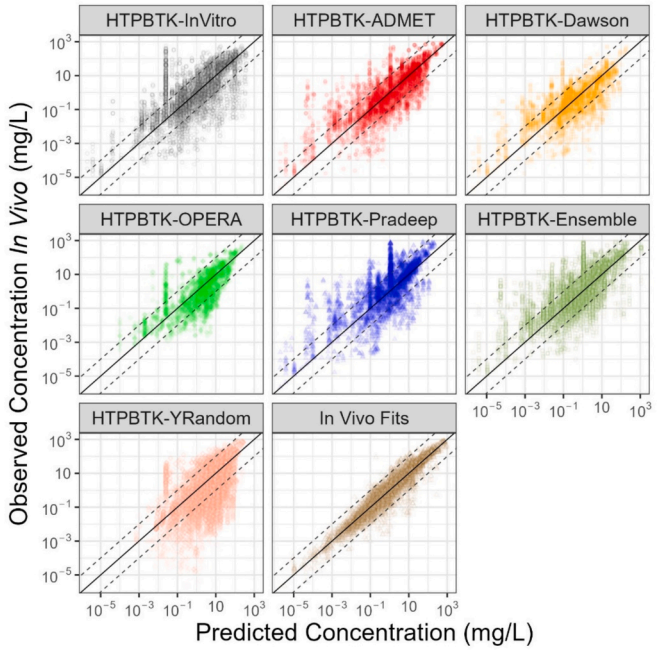

Fig. 7. Comparison of time course in vivo measured chemical concentrations (CvT) (Sayre et al., 2020) vs. predictions for empirical models fit to the data ("In vivo Fits"), and predictions for a HT-PBTK model parameterized with chemical specific Cl_{int} and f_{up} values either measured *in vitro* ("HTTK-InVitro") or predicted with various QSPR models. In each sub-plot. The y-axis shows the measured data while the x-axis shows the predictions made using chemical-specific parameters from the various sources. The solid line indicates identity (1:1) while the dashed lines indicate ten-fold difference.

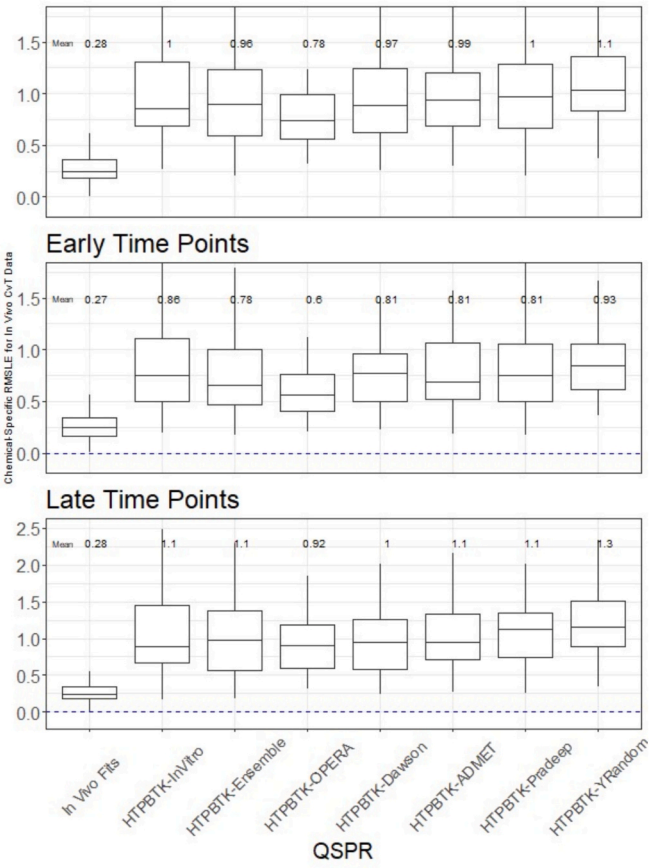

Fig. 8. Chemical-Specific RMSLE for *In vivo* CvT data. The upper and lower extent of the box for each model indicates the 25th to 75th quantiles, the mid-line indicates the median (50th quantile) and vertical line indicates 1.5× the range of the box. The mean is annotated above the box.

However, since there were only 51 chemicals with measured Cl_{int} and $f_{up} > 0$, which are required for the HT-PBTK model, *in vitro*-based predictions could only be evaluated for this set of 51 chemicals.

Hepatic metabolism is often assumed to be "restrictive" (Wetmore et al., 2012) – that is, rate of metabolism is limited to the fraction of chemical unbound in the plasma of the liver. However, for weakly bound chemicals it may be more appropriate to model the chemicals as "non-restrictive" (Krause and Goss, 2018, 2021). Currently there is no *a priori* method for determining plasma binding restriction. Here the HT-PBTK model was initially simulated both ways to identify the assumption that worked best for the evaluation chemicals.

In Table 5, we perform level 2 (CvT) and level 3 (AUC, C_{max}) evaluations of the *in vitro* data. This initial benchmarking informs what sorts of tests we should use in subsequent sections to compare the QSPR predictions. We do not perform a level 1 evaluation (comparing f_{up} and Cl_{int} to *in vitro* data) because we are using the data themselves to make the predictions reported here. The different benchmarks for HTTK vary from a best case of when the *in vivo* data are used to empirically estimate parameters to a worst case of when y-randomized *in vitro* values are used. It appears clear that the majority of chemicals in this set are non-restrictive (RMSLE 1) rather than restrictive (1.2). While the restrictive case represents a benchmark for the performance of default application of "httk" (that is, R package httk defaults to restrictive clearance), for the evaluation of the QSPR models we choose to use the better fitting non-restrictive assumption going forward.

Note that the performance for y-randomization in Table 5 (1.2) is with the assumption of non-restrictive clearance. That result is therefore not directly comparable to the performance of restrictive clearance without y-randomization (1.2). Both can be compared to the performance of non-restrictive clearance without y-randomization (1) but are not meaningfully compared to each other since the y-randomization was only conducted for the non-restrictive case. The error of y-randomized restrictive simulations would be larger. Further note that the y-randomization as performed here only randomizes the f_{up} and Cl_{int} values – that is, the correct, chemical-specific physico-chemical properties, which are critical to predicting TK, are still being used when Cl_{int} and f_{up} are randomized.

The differences between HT-PBTK parameterized with human and rat chemical-specific *in vitro* TK parameters (f_{up} and Cl_{int}) appear to be significantly less than indicated by Fig. 3. That is, for whatever reason, the chemicals in this evaluation of CvT data appear to be less sensitive to the differences between human and rat. While the previous analysis in "Interspecies Concordance" assumed that 100 % of the differences in TK between humans and rats was summarized by differences in Cl_{int} in f_{up}, in fact it appears that that these differences are not as significant for the chemicals analyzed here.

The metrics in Table 5 differ in sensitivity to the HTTK parameters, as indicated by the relative difference between using the values for the correct chemicals and the y-randomized values. For example, the difference in prediction error for AUC is a factor of 1.7-fold (or 50 times) between the appropriate *in vitro* value (0.87) and the y-randomized value (1.5). In contrast, there is little difference for early CvT data – a factor of 1.1-fold between the appropriate *in vitro* value (0.86) and the y-randomized value (0.93). That is, the AUC is sensitive to the chemical-specific *in vitro* TK parameters (f_{up} and Cl_{int}) and early CvT data is insensitive to these parameters (see the Sensitivity Analysis results). We expect the performance of HTTK with QSPR-derived inputs to generally be somewhere between the performance with *in vitro* inputs and the y-randomized inputs, depending on the subset of chemicals with available data.

Since the start of the collaborative trial for f_{up} and Cl_{int}, new *in vitro*

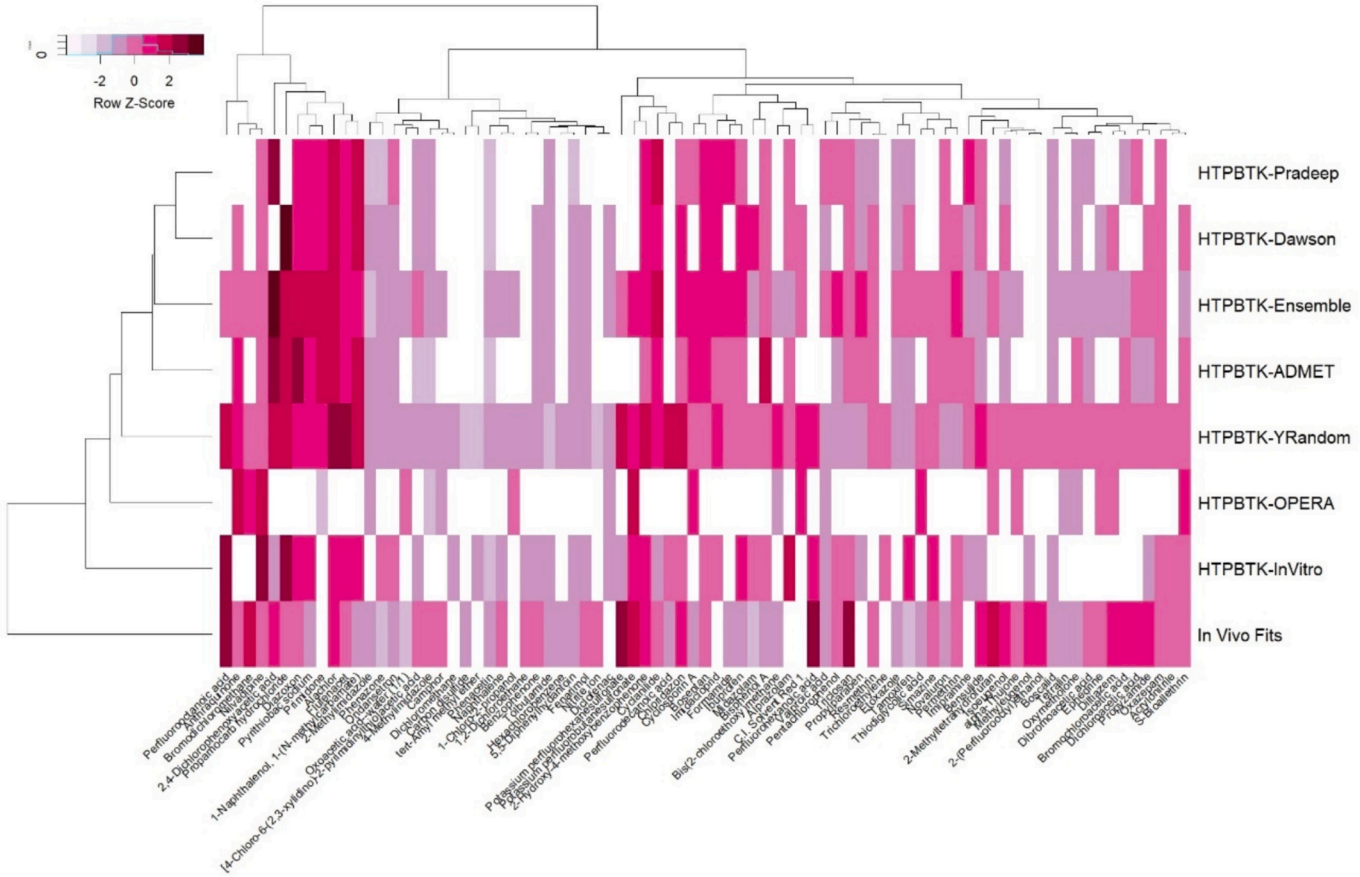

Fig. 9. Individual RMSLE by chemical and method. Lighter values indicate better predictive ability. The columns indicate different evaluation chemicals. The rows give the different prediction methods: The empirical fits to the data are given by "In vivo Fits". All other values are calculated using the HT-PBTK model and either measured values ("*In vitro*"), y-randomized measured values ("Y-Random") or the various QSPR models. White indicates missing values.

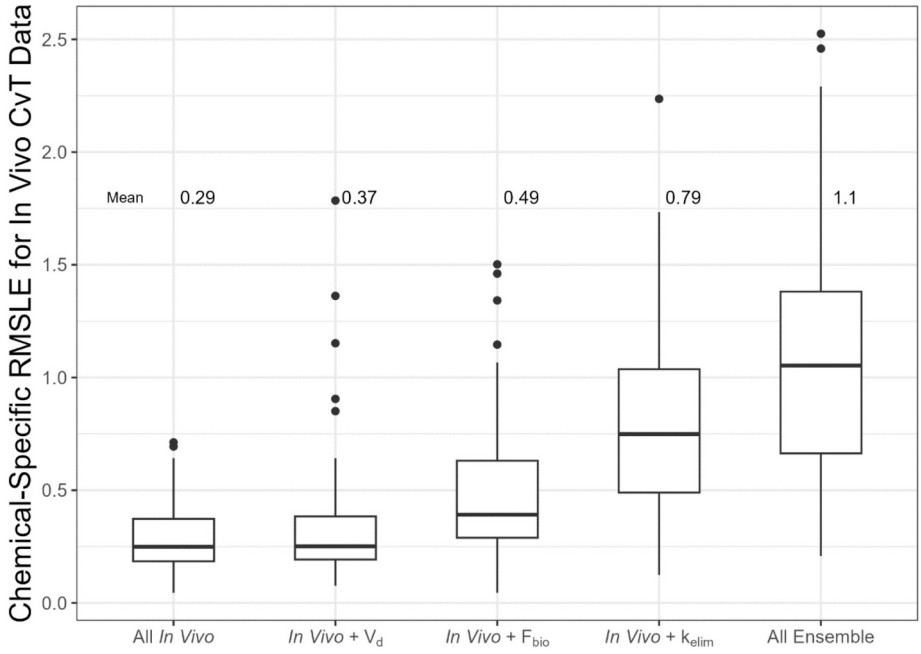

Fig. 10. Replacing *In vivo* Estimates with QSPR-based HTTK. The box-and-whisker plots indicating median, interquartile, 95 %, and outlier RMSLE on a per chemical basis. The lowest median error is when all parameters for an empirical TK model (F_{bio}, V_d, and k_{elim}) are derived from the *in vivo* time course data ("All In vivo"). Substituting values predicted from the Ensemble QSPR with HT-PBTK indicates which aspects of TK are the largest contributor to error. "All Ensemble" indicates the error when all parameters are estimated with the Ensemble QSPR and HT-PBTK.

Table 5
Model performance benchmarks (root mean square log10 error – RMSLE) for level 2 and level 3 statistics. The number of chemicals used in each evaluation is given in parentheses in the column name.

		In vitro			
	In vivo Fits (43)	Human Non-Restrictive (51)	Human Restrictive (51)	Rat (33)	Y-Randomized (51)
CvT RMSLE	0.29	1	1.2	1	1.2
Early CvT RMSLE	0.27	0.85	0.95	0.8	0.9
Late CvT RMSLE	0.28	1.1	1.4	1.1	1.3
AUC RMSLE	0.13	0.86	1.3	0.9	1.4
C_{max} RMSLE	0.24	0.75	0.88	0.75	0.86

data and QSPR-predictions characterizing oral absorption (that is, Caco-2 membrane permeability) became available (Honda et al., 2025). As is shown in Supplemental Table 9, inclusion of chemical-specific measures of permeability decreases the RMSLE in the non-restrictive case to 0.96. The default settings for "httk" currently assumes non-restrictive clearance and incorporates Caco-2 information – in this case the RMSLE is 1.1. We do not make use of the Caco-2 related information in the subsequent analysis since these data were not available when the challenge was designed.

3.4. QSPR model predictions

The QSPR models evaluated are listed in Table 3. The number of chemicals for which predictions could be made (that is, domain of applicability) varied across the different QSPR models. Throughout this effort, we report statistics on different subsets of chemicals. Unfortunately, there were no chemicals with predictions from every QSPR. As above, we report QSPR predictions for chemicals where *in vitro* values are available as "*in vitro*" – this subset is the most directly comparable to the benchmarks above. Finally, statistics are reported for the maximal number of chemicals for which each QSPR model could make predictions – denoted as "maximal". There were 12 chemicals with incomplete QSPR predictions (either Cl_{int} or f_{up} missing for all QSPRs): 1,2-Dichloroethane, 8:2 Fluorotelomer alcohol, Acrylonitrile, Carbon disulfide, Dichloromethane, Hexachlorobenzene, Methyleugenol, Nitrite, Perfluorodecanoic acid, Perfluorohexanoic acid, tert-Amyl methyl ether, Trichloroethylene. 4 chemicals had no QSPR-based Cl_{int} predictions, and 10 chemicals had no f_{up} predictions. Note that in some cases there were QSPR predictions removed for these chemicals because of their presence in a QSPR's training set.

We summarize the chemical-specific properties and predictions in Fig. 4, where similar chemicals (rows) and properties/predictions (columns) are clustered together based on Euclidean distance. We assumed that properties/predictions were centered (mean changed to zero) and scaled (divided by standard deviation) such that the value reflects the number of standard deviations from the mean.

3.5. Level 1 analysis

Our first level of evaluation directly compared the predictions of QSPR models with the chemical-specific *in vitro* measured TK values. This Level 1 analysis allows characterization of the differences between the models. There were evaluation data for Cl_{int} for 56 chemicals and f_{up} for 52.

3.5.1. Individual QSPR models

We evaluate model performance for Cl_{int} in Fig. 5. In both Fig. 5 and Fig. 6, wherever possible, we separate data that was in the model training sets (panel B) from external evaluation data that was not used to train the model (panel A). In panel B of both figures, it is clear than in some cases local QSPR models (that usually synthesize information on the most similar molecules to a structure) are making use of a single measured value for their prediction. Essentially, in these specific cases the local QSPR model is directly retrieving the *in vitro* measured values. Unless they are known to not be in the training set for a given QSPR, these low error predictions (Supplemental Table 5) are withheld from the Level 2 and Level 3 analysis since they do not help characterize QSPR performance for a chemical without measured *in vitro* parameters. Fig. 6 shows that all four models perform very well for predicting f_{up}. Predictions are highly correlated with observations.

In Table 6 we summarize the fold errors for the five QSPR models. The QSPR models perform similarly. Note that the Dawson et al. (2021) model is categorical (that is, predicting only three values: slow, moderate, and fast) while the other models are continuous. Supplemental Table 6 gives the number of chemicals for which predictions were made for each QSPR model. As indicated by Table 6, very few evaluation chemicals were not already present in the training sets of the Dawson et al. (2021), Pradeep et al. (2020), and OPERA models. Each QSPR's model predictions are compared separately to observations in Supplemental Fig. 1.

We examined the distributions of fold errors between the predictions and the measured data using a Kolmogorov-Smirnov test. The QSPR errors were highly correlated. For Cl_{int} there were no QSPRs whose distribution of predictions significantly differed (p-value threshold <0.05, see Supplemental Table 7). Pradeep did not have any Cl_{int} predictions for chemicals not in the model's training set. For f_{up}, there were no significant differences (see Supplemental Table 8).

3.5.2. Ensemble model

Due to the varying domains of applicability of the QSPR models and withholding those chemicals where the properties seem to be direct retrievals of experimental values, there was only one chemical in the intersection subset where all QSPR models could make a prediction. However, there were 69 chemicals where at least one of the QSPR models could make a prediction. An ensemble model was constructed that synthesized the predictions of all QSPR models. When multiple QSPR predictions were available, they were combined into a single value. For plasma protein binding the predictions were averaged, however for hepatic clearance the highest value was used. For clearance the assumption was that if any one model predicted rapid metabolism that the compound should be treated as rapidly metabolized.

3.6. Level 2 analysis

For the second level of analysis, we compared predictions based on the QSPR model predictions with actual CvT data. We show all CvT curve fits and predictions for each QSPR model, chemical, species, and route basis in the Supplemental materials (Supplemental Figures SupFig-ChembyQSPRCvTPlots.pdf). In the Level 2 (and 3) analyses chemical-specific predictions for Cl_{int} and f_{up} were still analyzed even if they had been present in the training set for a QSPR model. The only exception were any potential direct retrieval values (that is, predictions within 1 % of measurement) were omitted from Level 2 (and 3) analysis unless they were confirmed to not be present in the training set for a given QSPR model.

All QSPR model predictions were used as inputs to an HT-PBTK model to make CvT predictions. Fig. 7 shows the Level 2 comparisons to *in vivo* observed plasma concentrations across the chemicals. Predictions are made *via* HT-PBTK model simulations using each set of QSPR model predictions for all observed time points. Across the QSPR models, the predictions tended to be within a factor of ten (indicated by the dashed lines). We see chemicals where there are vertical bars in Fig. 7, indicating that the predicted concentrations were relatively

Table 6
Biases of the QSPR models for predicting *in vitro* measured Cl_{int} and f_{up} values in terms of median absolute value (Abs) of the log10 fold error (log10 FE) and log10 fold error (log10 FE). No bias would be log10 FE = 0. The chemicals analyzed are those not in the QSPR training sets (where known).

	Number of Cl_{int} predictions Compared	RMSLE	Median Cl_{int} Abs log_{10} FE	Median Cl_{int} log_{10} FE	Number of f_{up} predictions Compared	RMSLE	Median f_{up} Abs log_{10} FE	Median f_{up} log_{10} FE
ADMet	42	1.61	0.507	0.348	36	0.58	0.294	−0.108
Dawson	3	2.2	0.524	0.376	11	0.406	0.281	−0.0206
Pradeep	0				1	0.226	0.226	0.226
OPERA	4	2.78	3.42	−0.379	4	1.28	3.4	−1.56
IVBP-Suite	46	1.5	0.415	0.312	0			

constant over time while the observed concentrations changed. Typically, these are chemicals where the CvT time course was especially biphasic, with an initial rapid decline and then tails where low levels of the chemical remained detectable for a long time. In the tails (vertical bands in Fig. 7), the predictions tend to underestimate observed concentration. The panel labelled "HTPBTK-InVitro" in Fig. 7 shows the HT-PBTK predictions based on the measured values, where available. Finally, the panel "*In vivo* Fits" in Fig. 7 shows the predictions of empirical model optimized to the *in vivo* data.

For log-normally distributed properties the RMSLE is an empirical estimate of the standard deviation of a prediction. Prediction error as characterized by RMSLE is calculated first on a per chemical basis for each prediction method. That is, error is first calculated for predictions from a single method for each individual chemical aggregating over all doses, routes, and time points for that chemical. Then, for each prediction method, the mean RMSLE across all chemicals is reported in Table 7.

The QSPR models perform roughly equally for the chemicals with *in vitro* data, with RMSLE ranging between 0.9 and 1.1 in Table 7. Notably the QSPR models are close to the performance of the *in vitro* HTTK data. The ensemble predictor (using the mean plasma binding and the maximal predicted metabolism) performs similarly to the *in vitro* measured Cl_{int} and f_{up} data.

In Table 7 we note that error was generally consistent regardless of the subset of chemicals used. The chemical subsets examined were: 1) chemicals with *in vitro* measured Cl_{int} and f_{up} data; 2) chemicals without *in vitro* measured Cl_{int} and f_{up} data; 3) all chemicals predicted by a given method (maximal), 4) pharmaceuticals, and 5) non-pharmaceuticals. In subsequent results, we present statistics from the maximal set of chemicals. In Table 7, we also examined the performance for both pharmaceuticals and non-pharmaceuticals, finding that both *in vitro* measured Cl_{int} and f_{up} data and QSPR models tended to perform similarly or slightly better for pharmaceuticals. Pharmaceuticals are often designed to have short to moderate $t_{½}$ and be cleared by the liver, which would lead them to be better described by the parameters Cl_{int} and f_{up}.

AFE and AAFE results were also calculated and are fully presented in Supplemental Tables 10 and 11. In Table 8 we compare the RMSLE, AAFE, and AFE. AFE indicates bias, with tendency to over-estimate – the AFE of 0.22 for predictions based on measured *in vitro* parameters corresponds to an average over-estimation of 1.7-fold. Predictions based upon the empirical fits to the *in vivo* data slightly underestimated the observed concentrations.

Table 8
Comparison of mean summary statistics for different QSPRs. Comparison of error measures RMSLE, AAFE, and AFE from Level 2 Analysis of predictions based on empirical model fits and a HT-PBTK model parameterized with chemical specific values either measured *in vitro* ("HTTK-InVitro") or predicted with various QSPR models. Values were first calculated on a per chemical basis and then averaged across chemicals. The number of chemicals in each calculation is given in parentheses. *Note that OPERA predictions were available for many more chemicals, but cases where the nearest-neighbor algorithm returned the measured value were withheld to allow evaluation of performance on novel chemicals.

	HT-PBTK-InVitro	HT-PBTK-ADMET	HT-PBTK-Dawson	HT-PBTK-OPERA*	HT-PBTK-Pradeep	HT-PBTK-Ensemble
RMSLE	1 (51)	0.99 (50)	0.97 (48)	0.78 (23)	1 (43)	0.96 (69)
AAFE	15 (51)	10 (50)	12 (48)	5 (23)	11 (43)	13 (69)
AFE	0.22 (51)	0.03 (50)	0.065 (48)	0.11 (23)	−0.11 (43)	−0.21 (69)

In Fig. 8 we examine the distribution of per chemical RMSLE. In the first panel of Fig. 8, all observed time points are valued equally, without consideration of phase (absorption/distribution/metabolism) and measurement accuracy. In Fig. 8, the performance of HT-PBTK with Cl_{int} and f_{up} parameters for a random chemical, while worse, is not a marked departure from the performance HT-PBTK with the parameters for the correct chemical. At early time points (second panel of Fig. 8), all methods are more accurate than for all time points (first panel). This is consistent with Fig. 1 – the early absorption and distribution phases are dominated by the V_d. Prediction of V_d largely depends on physicochemical properties (which have not been randomized) and weakly on f_{up}. See Supplemental Table 11 for RMSLE across subsets and predictive models for early and late times.

Also consistent with Fig. 1, the most discriminating data for judging HT-PBTK-based CvT predictions are the later time points, which characterize metabolism and excretion (the elimination phase of TK). In the

Table 7
Mean per-chemical RMSLE for CvT predictions from different chemical subsets. RMSLE from Level 2 Analysis of predictions based on empirical model fits and a HT-PBTK model parameterized with chemical specific values either measured *in vitro* Cl_{int} and f_{up} ("HTTK-InVitro") or predicted with various QSPR models. RMSLE was first calculated on a per chemical basis and then averaged across chemicals. The number of chemicals in each calculation is given in parentheses. The sets of chemicals refer to those with *in vitro* HTTK parameters ("*In vitro*"), those without *in vitro* HTTK parameters ("No In vitro"), the maximum number of predictions available for each chemical, the pharmaceutical chemicals, and the non-pharmaceutical chemicals. *Note that OPERA predictions were available for many more chemicals, but cases where the nearest-neighbor algorithm returned the measured value were withheld to allow evaluation of performance on novel chemicals.

Chemical Set	HT-PBTK-InVitro	HT-PBTK-ADMET	HT-PBTK-Dawson	HT-PBTK-OPERA*	HT-PBTK-Pradeep	HT-PBTK-Ensemble
In vitro	1 (51)	1 (36)	1 (36)	0.9 (5)	1.1 (30)	1 (43)
No *In vitro*		0.91 (14)	0.84 (12)	0.74 (18)	0.79 (13)	0.83 (26)
Maximal	1 (51)	0.99 (50)	0.97 (48)	0.78 (23)	1 (43)	0.96 (69)
Pharmaceutical	0.92 (21)	0.85 (20)	0.8 (22)	0.58 (6)	0.88 (20)	0.82 (23)
Non-Pharmaceutical	1.1 (30)	1.1 (30)	1.1 (26)	0.85 (17)	1.1 (23)	1 (46)

third panel of Fig. 8, we see that all HT-PBTK model simulations made with either QSPR model predictions or *in vitro* measured parameters perform worse in the elimination phase, which is driven by the estimated Cl_{int}. For the late time points, the specific values of the Cl_{int} and f_{up} parameters (measured or predicted) have greater influence on the accuracy of the predictions – *y*-randomization performs notably worse than the QSPR models. The HT-PBTK model simulations based upon *in vitro*-measured Cl_{int} and f_{up} data have an RMSLE which is indistinct from the QSPR models. Based on the full time series, the ensemble QSPR model predictions, using the most rapid predicted clearance, shows an average RMSLE similar to the *in vitro* measured data of 1; however, the OPERA QSPR shows an even smaller average RMSLE of 0.9. When considering only the later timepoints, several QSPRs have a RMSLE comparable to *in vitro* data (1), while y-randomization has RMSLE 1.2.

Fig. 9 shows the chemical-specific RMSLE values as a function of prediction method. In Fig. 9, each column indicates a different chemical and each row a different method. These data are also provided in Supplemental Table 12. Chemicals and methods have been clustered based upon Euclidean distance. White gaps mark chemicals that were not in the domain of applicability of different QSPR models (or, in the case of *in vitro* measured data, no measurements were available). We see that the ensemble QSPR model provides the best coverage of chemicals because the domain of applicability of the different QSPR models are varied. In general, the RMSLE for a given chemical appears to be similar across model predictions; however, there are several chemicals where the *in vivo* fits are substantially better.

Finally, as part of the Level 2 analysis we contrast the assumption of non-restrictive clearance and the use of the maximum predicted clearance in the ensemble model. For chemicals with moderate to high protein binding (that is, low f_{up}), a non-restrictive assumption enhances the rate of clearance. Similarly, using the maximum predicted clearance across the ensemble of QSPRs acts to boost the metabolism, independent of whether the chemical is highly protein bound. In Table 9 we consider maximum *vs.* mean ensemble QSPR Cl_{int} and restrictive *vs.* non-restrictive clearance. Again, assuming non-restrictive clearance for all chemicals in the evaluation set has lower error. For non-restrictive clearance, using the mean QSPR model prediction for Cl_{int} appears to work slightly better. However, if restrictive metabolic clearance is assumed taking the maximum prediction is slightly better than the mean for late time points, where clearance is most important.

3.7. Level 3 analysis

The Level 3 analysis evaluates prediction of TK summary parameters; specifically, C_{max}, AUC, V_d, $t_{½}$, and Cl_{tot}. Where available, we compare the predictions to the values estimated from the empirical fits to the CvT data. The values predicted for each method are provided in Supplemental Table 13.

Early time points are dominated by the ability to correctly predict peak plasma concentration (C_{max}). In Table 10, we examine each method's accuracy in predicting the C_{max} as determined from invivoPKfit using the CvT data. Empirical fits are again best. Predictions based on *in vitro* measured Cl_{int} and f_{up} data are not that different from QSPR model predictions. As shown in Supplemental Fig. 4, all the models tend to accurately predict C_{max}, which largely depends on V_d (for intravenous doses C_{max} = dose / V_d). Even the y-randomized predictors do reasonably well for C_{max} since V_d depends strongly on physicochemical properties (which are not randomized in our analysis) and weakly on Clint and fup (as in Fig. 1).

The most discriminating data for evaluating HT-PBTK-based CvT predictions depend on the later time points which characterize metabolism and excretion and inform metrics such as AUC. In Table 10, the empirical AUC estimates give a clear best-case scenario, while the y-randomization more clearly gives a worst-case scenario. *In vitro* measured Cl_{int} and f_{up} data predict AUC with an RMSLE of 0.87, while the QSPR models range from RMSLE 0.5 to 0.96. The ensemble model predictions have a RMSLE of 0.83. In Supplemental Table 14 we calculate separate RMSLE for C_{max} and AUC for pharmaceuticals and non-pharmaceuticals. We do not observe differences for the chemicals evaluated here.

As we see in Table 10, C_{max} is estimated more accurately than AUC. From Fig. 1, we know that if AUC is less accurate than C_{max}, we are likely dealing with an issue in metabolism or elimination. From the AFE in Table 8 and Table 10, we see that there is a tendency to over-predict concentrations (that is, predictions larger than observations lead to a positive AFE). All these factors suggest that the HT-PBTK model simulations based on Cl_{int} and f_{up} tend to underestimate metabolism and elimination. The ensemble QSPR model, using the maximum predicted clearance, somewhat overcorrects and has a bias toward underestimating AUC.

In Table 11, we also examine two quantities that inform our ability to predict AUC at late time points: $t_{½}$ and CL_{tot}. The QSPR models for Cl_{int} and f_{up} data predicted $t_{½}$ with RMSLE indicating errors between 0.99 and 1.2 (9.8- to 16-fold). QSARINS-Chem, which was trained specifically to $t_{½}$, data (rather than *in vitro* measured Cl_{int} and f_{up}) produces better predictions, though the similarly-developed IFS-QSAR model performs more like the ensemble QSPR model. Finally, in Table 11 we examine predictions for CL_{tot}, which depends on both elimination rate (inverse of $t_{½}$) and V_d.

3.8. Sensitivity analysis

Table 11 indicates that different aspects of TK (as illustrated by Fig. 1) are predicted with varying accuracy by QSPR-parameterized HT-PBTK. To evaluate the impact of the errors summarized in Table 11, we used the parameters estimated directly from the *in vivo* data as the "best" case and then substituted QSPR-based values for different parameters. In Fig. 10, we first indicate the distribution of per-chemical RMSLE when a one- or two-compartment model is used with parameters optimized to the *in vivo* data. We then substitute parameters based on the ensemble QSPR and HT-PBTK for the absorption (F_{bio}), distribution (V_d), and elimination (k_{elim}) phases of TK. The results suggest that the current methods used to estimate V_d add relatively little overall error on average to the HT-PBTK model simulations, while the assumption of 100 % oral absorption in the calculation of F_{bio} adds slightly more relative error. For both V_d and F_{bio}, there are small subsets of chemicals where the RMSLE is greater than 1. We see that the largest errors are observed when we use QSPR model predictions for Cl_{int} and f_{up} data in estimating the elimination rate, consistent with the idea that estimation of metabolism and other potential routes of elimination (for example, renal excretion) are the most challenging aspect of parameterizing an HT-PBTK model. In Fig. 10, we also include the distribution of RMSLE when all parameters are derived from the ensemble QSPR model – the errors from each phase of TK combine to produce a larger error, although the mean error is still within a factor of ten (RMSLE <1).

Table 9
Comparison of approaches to hepatic metabolism (restrictive or non-restrictive plasma binding) and aggregating ensemble predictions (mean QSPR model prediction *vs.* maximum QSPR model prediction).

	RMSLE		RMSLE Late	
	Cl_{int} Max	Cl_{int} Mean	Cl_{int} Max	Cl_{int} Mean
Non-Restrictive	0.96 (69)	0.95 (69)	1.1 (69)	1 (69)
Restrictive	1.1 (69)	1.1 (69)	1.2 (69)	1.3 (69)

4. Discussion

Translation of *in vitro* concentrations that cause bioactivity to putative administered *in vivo* doses by HTTK is key to next generation chemical risk assessment. Multiple government agencies and advisory committees have recognized that a high throughput (chemical-agnostic)

Table 10
Mean RMSLE, AAFE, and AFE for C_{max} and AUC TK statistics. Level 3 Analysis of predictions based on empirical model fits and a HT-PBTK model parameterized with chemical specific values either measured in vitro ("HTTK-InVitro") or predicted with various QSPR models. "HTTK-YRandom" indicates the HT-PBTK model parameterized with a random permutation of the measured in vitro parameters but not the physico-chemical properties.

	HTPBTK-InVitro	HTPBTK-ADMET	HTPBTK-Dawson	HTPBTK-OPERA	HTPBTK-Pradeep	HTPBTK-Ensemble	HTPBTK-YRandom	*In Vivo* Fits
C_{max} RMSLE	0.75	0.74	0.76	0.45	0.74	0.67	0.74	0.22
C_{max} AAFE	8.3	7.9	8.9	3.1	9	6.8	12	2
C_{max} AFE	0.18	0.13	0.17	0.017	0.098	−0.095	0.31	−0.12
AUC RMSLE	0.86	0.76	0.76	0.58	0.78	0.74	1.4	0.13
AUC AAFE	53	17	55	5.2	19	21	270	1.4
AUC AFE	0.36	0.14	0.13	0.24	−0.046	−0.22	1.3	−0.072

Table 11
Mean RMSLE for other key TK statistics (Level 3 Analysis). Analysis of predictions based on empirical model estimates and a HT-PBTK model parameterized with chemical specific values either measured *in vitro* ("HTTK-InVitro") or predicted with various QSPR models. "HTTK-YRandom" indicates the HT-PBTK model parameterized with a random permutation of the measured *in vitro* parameters.

	HTPBTK-InVitro	HTPBTK-ADMET	HTPBTK-Dawson	HTPBTK-Pradeep	HTPBTK-OPERA	HTPBTK-Ensemble	HTPBTK-YRandom	QSARINS-Chem	IFS-QSAR
$t_{½}$ RMSLE	1.6	1.2	1.1	0.99	1.2	1.1	2.8	0.55	1.1
Cl_{tot} RMSLE	1.8	1.2	1.2	1.1	1.3	1	3.2	0.62	1
V_d RMSLE	1.1	0.92	0.94	0.94	0.65	0.95	0.89		
C_{ss} RMSLE	1.8	1.2	1.2	1	1.3	0.98	3.2	0.62	1

TK model using chemical-specific measured *in vitro* parameters can be a powerful tool in linking administered to internal dose (Health Canada, 2021; National Academies of Sciences and Medicine, 2017; Paini et al., 2020; U.S. Environmental Protection Agency, 2021). There is an ongoing proliferation of HT-PBTK models developed to make use of these *in vitro* measured parameters to make chemical-specific predictions (Armitage et al., 2021; Bernstein et al., 2021; Breen et al., 2021; Geci et al., 2024; Jamei et al., 2009b; Kapraun et al., 2022; Linakis et al., 2020). While *in vitro* TK data have been generated for more than 1000 chemicals, several thousand chemicals remain in need of TK information, leading researchers and regulators to potentially rely on machine learning to address this gap (Chou and Lin, 2023; Di Lascio et al., 2023).

Our goal is to establish the expected accuracy of time-course and summary-level TK predictions made without any *in vivo* or *in vitro* measurements and using only QSPR models. In this study, we find that a QSPR model approach may yield reasonable predictions of key TK parameters for novel compounds. We believe a QSPR approach might be sufficiently reliable for prioritization when the error introduced by substituting QSPR model predictions for *in vitro* measurements is less than a fixed factor (such as 3 or RMSLE 0.48). Given that resources such as the CCD list more than 1,000,000 chemical substances, QSPRs for HTTK provide a path forward.

We have characterized the accuracy of HT-PBTK modeling approaches for new chemicals based on structure-derived *in silico* predictions of *in vitro* TK parameters. This was accomplished through a collaborative trial of five QSPR models for *in vitro* f_{up} and/or Cl_{int} parameters and two additional QSPR models of *in vivo* TK $t_{½}$. To bracket the performance of the QSPR models, we have characterized the accuracy of 1) using parameters empirically estimated from the *in vivo* data (that is, fits), 2) using *in vitro* measured parameters for f_{up} and Cl_{int}, and 3) using random (incorrect) draws from a large library of *in vitro* measured f_{up} and Cl_{int} parameters (though the physicochemical properties remained correct).

We have focused on the RMSLE statistic because it provides an estimate of the prediction error for novel chemicals. Both RMSE and RMSLE are estimates of the expected error of a prediction (Chai and Draxler, 2014). RMSE is appropriate for errors with a constant standard deviation (that is, a normal distribution) while RMSLE is appropriate for errors with a constant coefficient of variation (that is, a log-normal distribution). Data in biology and medicine are often-log-normally distributed (Limpert et al., 2001). Since Cl_{int}, f_{up}, dose, and concentrations all vary over several orders of magnitude, we believe RMSLE to be most appropriate. However, we have calculated other statistics characterizing accuracy (such as AFE and AAFE) and included them mostly in supplemental material.

QSPR modeling approaches can be either "local" (synthesizing information on a novel chemical structure based only on sufficiently similar chemicals) or "global" (looking for systematic structure-property relationships across all available chemicals). OPERA is an example of a local model because it uses the k-nearest neighbors method. Because the training data for HTTK are relatively sparse (see Table 3), the number of similar chemicals can vary depending on structure. In some cases, a local model, when asked to make a prediction for a chemical in the training set, retrieves the measured value from the training set. In this case, a QSPR model based on a local model should perform very well – the QSPR is providing the measured data. Since the goal of the analyses was to characterize HTTK QSPR performance for novel chemicals, we have removed values that were less than 1 % different from the measured value (as an indicator of a local model with only one similar chemical in the training set). Because the CvT data are also limited (81 chemicals), we have included other training set chemicals in the Level 2 and 3 evaluations.

This collaborative trial used a database of *in vivo* measured TK data to evaluate *in silico* approaches. *In vivo* data had to be carefully reviewed to include only those data that could be well-described by empirical TK models. By comparing predictions with observations, the RMSLE and other key statistics could be calculated. The RMSLE characterizes the expected accuracy for new predictions and can be interpreted as a coefficient of variation for normally distributed errors about the prediction. For log-normally distributed errors, one has 95 % confidence that the actual value will occur within ±2 RMSLE of the prediction.

Here the QSPR models were first (Level 1) evaluated for their ability to predict the chemical specific TK parameters Cl_{int} and f_{up} that are used by HT-PBTK models. These *in vitro* parameters do not directly correspond to *in vivo* TK parameters (for example, V_{dist}, k_{elim}). Even if a QSPR model perfectly reproduces an *in vitro* measurement, the model prediction is only as good as the *in vitro* assays they were trained to estimate. However, the *in vitro* assays for Cl_{int} and f_{up} provide relatively rapid methods for partially characterizing TK (Breen et al., 2021; Wang, 2010).

In the Level 2 analysis, HT-PBTK models parameterized using QSPR model predictions of Cl_{int} and f_{up} were evaluated across the full

concentration time-course. However, TK in the absorption and distribution phases are relatively insensitive to these *in vitro* measured parameters (Fig. 10). For example, distribution depends more on physico-chemical properties and only somewhat on f_{up} for accurate prediction of equilibrium tissue partition coefficients (and, in turn, V_d) (Pearce et al., 2017a). Further, the absorption rate does not depend on f_{up} and Cl_{int}. However, the elimination phase of a TK time course is dominated by metabolism and excretion, which are both characterized to a larger extent by the *in vitro* Cl_{int} and f_{up} parameters.

A key factor to consider is that while the *in vitro* TK data evaluated here were human, the *in vivo* data were largely taken from rats. Comparison of the HT-PBTK predictions made with rat *in vitro* TK parameters against predictions made with human *in vitro* TK parameters indicated that species differences could inflate RMSLE by as much as 0.8. However, a subset of the *in vivo* evaluation chemicals did have *in vitro* TK data measured in rat. For these chemicals, the HT-PBTK model parameterized with *in vitro* TK parameters for rat (the correct species) had error roughly the same as when human *in vitro* TK data were used (Table 5). Using chemicals with Cl_{int} and f_{up} measured both in human and rat, we found that predictions of C_{max} had mean RMSLE of 0.68 across chemicals when human data were used instead of rat. Human-substituted AUC had RMSLE of 0.8. If these errors were the result of independent factors, which they are not, then one could subtract the interspecies HTTK error estimate to get a better estimate of how well the human QSPR models predict *in vivo* data. The observed error for AUC was 0.87 (in Table 10), while the interspecies error was estimated to be 0.8. This means that the values estimated here may be more consistent with Wang (2010) – that is, RMSLE of ~0.5 – when interspecies differences are taken into account. The estimated errors for the time-course data (Table 7) might similarly be overestimated.

To contextualize the Level 2 evaluation, statistics were also calculated for empirical TK models optimized using the *in vivo* data. The empirical TK models were intended to approximate a best case, accounting for the inherent variability of the data themselves. As a worst-case scenario, HT-PBTK models were parameterized using *in vitro* Cl_{int} and f_{up} parameters for random (incorrect) chemicals, while keeping the physico-chemical parameters correct. This y-randomization helps account for potential correlations and a constrained dynamic range within the *in vitro* measured chemical data. During early time points (absorption and distribution), the HT-PBTK models parameterized with QSPR models perform closer to y-randomized predictions than to empirical estimates from the data. The relatively small difference between the error associated with the y-randomized, QSPR models, and *in vitro*-measured values for the early time points indicates the greater importance of physico-chemical properties for predicting TK in the absorption and distribution phases. However, at later time points (elimination phase), the separation between some of the QSPR models, *in vitro* measured values, and y-randomized predictions was larger, highlighting the increased importance of experimental measurements or *in silico* predictions of the Cl_{int} and f_{up} parameters in this phase. It is worthwhile noting that y-randomization does not tend to produce extreme values because the data are clustered around pronounced modes of much more likely values (Dawson et al., 2021; Kirman et al., 2015).

In the Level 2 analysis, we have found the HT-PBTK model performed similarly with either QSPR model predicted or *in vitro* measured parameters. TK model parameters estimated empirically from the full time-course of *in vivo* data had mean RMSLE of 0.29 across all chemicals, while HT-PBTK model simulations based on *in vitro* Cl_{int} and f_{up} measurements had a mean RMSLE of 1. The mean RMSLE for HT-PBTK models parameterized using individual QSPR model predictions ranged from 1 to 1.3. Across the full time-course, both the *in vitro* measured data and QSPR model predictions only performed slightly better than the y-randomized predictions (y-randomized predictions were 1.6 times less accurate). For the elimination phase, the HT-PBTK models parameterized using *in vitro* measurements had a mean RMSLE of 1.1, while the RMSLE values for HT-PBTK models parameterized using individual QSPR model predictions ranged from 0.92 to 1.1. In comparison, the HT-PBTK models parameterized using y-randomized values had a mean RMSLE of 1.3.

Ensemble QSPR model predictions were constructed from the various QSPR models. The number of QSPR predictions varied from chemical to chemical, depending on domain of applicability and presence of the chemical in the QSPR training set. For plasma protein binding, the predictions of the various QSPR models were weighted equally. However, the ensemble model for Cl_{int} used the maximum Cl_{int} value predicted by any QSPR. Using the maximum Cl_{int} corresponds to a logical "OR" function for rapid metabolism – if any of the QSPR models predicts that a compound is rapidly cleared, then we treat it as rapidly cleared in the ensemble model. Only if all the QSPR models agree that metabolism is slow do we treat a compound as being slowly metabolized. This is not a conservative assumption, as more rapid metabolism predicts lower tissue concentrations. As explored here, it is possible that the plasma binding of the chemicals in the evaluation set was "non-restrictive" with respect to metabolism. With restrictive clearance, the more highly bound to plasma is a chemical, the more the rate of metabolism is slowed. Restrictive clearance is considered a more conservative assumption since it produces higher tissue concentration predictions and is the default for R package "httk" (Breen et al., 2021). Restrictive metabolic clearance only impacts chemicals that are 1) metabolized and 2) substantially protein bound. There is no known heuristic for identifying whether a chemical's metabolism is restricted by protein binding. Instead, we must compare to Cvt data and identify the assumption that is more consistent with the data. However, especially for non-pharmaceuticals, there is also a chance of extrahepatic metabolism, which would also act to increase the metabolism of a chemical regardless of whether metabolism was restricted by protein binding. Given that the HT-PBTK model used only includes renal clearance and metabolism in the liver, increasing the rate of liver metabolism might approximate the impact of extrahepatic biotransformation. Non-restrictive clearance seemed to be supported for the chemicals examined here as indicated by the lower RMSLE (Table 5). For the chemicals evaluated here, when we assume non-restrictive clearance there is little difference between the use of the mean and maximum clearance (Table 9) for the ensemble QSPR model. When non-restrictive clearance is assumed, the maximum clearance ensemble QSPR model performs better, but this could be due to the chemicals actually undergoing non-restrictive or extrahepatic metabolism.

In most applications of HTTK modeling using *in vitro* measured Cl_{int} and f_{up} parameters, the metabolism of chemicals is assumed to be "restrictive". The extent of restriction is likely due to the chemical affinity for the plasma proteins to which it binds, with some "highly bound" chemicals likely to quickly disassociate from a binding site and be available for metabolism. Since binding affinity is typically unknown, restrictive clearance is assumed to be a conservative assumption for human health risk assessment since slower clearance results in higher tissue concentrations. However, the AFE suggests that the non-restrictive assumption is still slightly conservative from a human health risk assessment perspective (0.8 for restrictive *versus* 0.22 for non-restrictive; see Supplemental Table 9).

Although *in vivo* TK data for nearly 100 chemicals is a substantial collection, the existing annotated *in vivo* TK data do not constitute machine learning "Big Data", which might rely on thousands, if not millions of observations (Kitchin and McArdle, 2016). While more than one thousand chemicals have available *in vitro* Cl_{int} and f_{up} measurements for QSPR modeling, even that is still a constrained, artifactual set of chemicals reflecting the correlation and the property distribution of the set as well as the dynamic range of the specific assays. For example, to be suitable for *in vitro* measurement, the volatility and solubility of the chemicals must be somewhat constrained (Richard et al., 2025). When chemical space is relatively narrow, the overall statistics may be biased; with the CvTdb, like any data set, one can only evaluate and model things that vary across the dataset. Among the *in vivo* evaluation

chemicals, only two chemicals had *in vitro* measured Cl_{int} above 10^3 μL/min/10^6 hepatocytes and 94 % of measured values are within two-fold of the median.

Seven out of the 56 chemicals with *in vitro* HTTK Cl_{int} measured data have no observed clearance (that is, they are metabolically stable within the short-duration viability of conventional suspension hepatocyte metabolism models) compared with 260 out of 1046 measured Cl_{int} values (25 %) in the httk data set as a whole. The two parameters f_{up} and Cl_{int} further interact in how they influence TK under restrictive clearance assumptions. If a chemical has low metabolic clearance, it may accumulate regardless of how highly the chemical binds. Conversely, if a chemical is highly bound, it may not matter how fast the free chemical clears.

QSPR models for metabolism are limited by many factors, including the limitations of the *in vitro* intrinsic clearance model. The majority of the *in vitro* Cl_{int} assays are based upon hepatocytes that are suspended in media and losing viability over a four-hour measurement (Rotroff et al., 2010; Shibata et al., 2002). The dominant *in vivo* metabolic pathway for a chemical is therefore not necessarily present *in vitro*. It is then unsurprising that the *in vitro* assays underestimate clearance (Sipes et al., 2017). We currently cannot evaluate the impacts of three-dimensional aspects of chemical structures (including chirality) – all the QSPR models are based on two-dimensional structure descriptors because sufficient *in vitro* TK data do not exist to train the models otherwise (for example chiral pairs). Further, the data used to train the models are based on human biological material, but the data used for evaluation here are largely drawn from rat, again due to the much wider availability of human than rat *in vitro* HTTK data (Honda et al., 2019; Wetmore et al., 2013).

Geci et al. (2024) recently evaluated HT-PBTK using a large data set of chemical concentration *vs.* time-course data (including the CvTdb) for which AUC and C_{max} were estimated *via* non-compartmental analysis. They found that the error for HT-PBTK parametrized with measured *in vitro* parameters was 2.0-fold for C_{max} and 1.8-fold for AUC. With *in silico* predictors, the best that Geci et al. (2024) could achieve was 2.2-fold for C_{max} and 2.4-fold for AUC. A key factor considered by Geci et al. (2024) that was not considered in most of the analyses here was intestinal permeability. We did perform limited benchmark simulations using oral absorption assumptions informed by either human Caco-2 permeability data or QSPR predictions (Honda et al., 2025). We found that additional data to inform oral absorption slightly reduces the RMSLE to 0.96 for the non-restrictive assumption and 1.1 for the restrictive assumption. However, the majority of the data analyzed here were for rat and there is little reason to expect concordance between rat and human oral absorption (Musther et al., 2014). The sensitivity analysis (Fig. 10) indicated that while the oral absorption assumptions used here did contribute to the estimated error, the elimination phase was a greater source of our estimated error. A potentially important consideration is that Geci et al. (2024) reports the median, rather than mean error for their results, and for non-normally distributed results (such as log-normal, with a long "tail" of larger values), the mean can be much larger than the median.

Geci et al. (2024) is consistent with Wang (2010) which found that the SimCYP model, parameterized with *in vitro* measured data, could predict AUC within 2.3-fold when evaluated across 54 pharmaceuticals. Wambaugh et al. (2018) evaluated HTTK-based IVIVE for TK using just over forty chemicals including pharmaceuticals and non-pharmaceuticals. They found that C_{max} could be predicted with RMSLE 2.2 and AUC could be predicted with RMSLE 2.0. As summarized in Table 10, here we found that C_{max} could be predicted with an RMSLE of 0.74 with *in vitro* measured data and an RMSLE of 0.67 with the ensemble QSPR model. AUC could be predicted with RMSLE of 0.87 using *in vitro* measured data and 0.76 using the ensemble QSPR model. We speculate that differences with the literature are due to 1) our curation of the CvT data by requiring that they can be fit to empirical models (Geci et al., 2024), 2) the inclusion of many non-pharmaceuticals (Wang, 2010), and 3) the inclusion of more chemicals (Wambaugh et al., 2018). However, for AUC predicted with *in vitro* parameters, we observe an RMSLE of 0.87 for pharmaceutical compounds and the same for non-pharmaceuticals.

The summary TK statistics (Level 3) involved in the elimination phase were hardest to predict. The related quantities AUC, CL_{tot}, and $t_{½}$ are more challenging than C_{max} and V_d. V_d only depends on partitioning (partially characterized by f_{up} and otherwise predicted according to physico-chemical properties) and C_{max} only depends on V_d; neither of these two quantities depend upon Cl_{int}. For AUC, optimized empirical TK model fits to the *in vivo* data indicated an RMSLE of 0.13 while the HT-PBTK model parameterized using *in vitro* values for random (incorrect) chemicals had an AUC RMSLE of 1.5. Using chemical-specific *in vitro* measured data in the HT-PBTK model simulated the AUC with an RMSLE 0.87, while HT-PBTK model simulations using QSPR model predictions ranged in RMSLE values from 0.58 to 0.78. HT-PBTK models simulations based on the ensemble QSPR model predictions, using the maximum clearance predicted across all QSPR models, predicted the AUC with an RMSLE of 0.74. Using the ensemble QSPR model parameter predictions, the total clearance CL_{tot} RMSLE is 1. For C_{max}, the empirical TK model fits had an RMSLE 0.22, while HT-PBTK models parameterized using Cl_{int} and f_{up} for random chemicals had an RMSLE of 0.73. Note that physico-chemical properties, which largely determine C_{max}, were not randomized, which is why the sensitivity analysis in Fig. 10 shows C_{max} is insensitive to Cl_{int} and f_{up}. The accuracy of QSPR models for physico-chemical properties has been examined elsewhere (Cappelli et al., 2015; Gadaleta et al., 2024; Nicolas et al., 2018).

HT-PBTK model simulations using *in vitro* measured data had an RMSLE of 0.74 for C_{max} QSPR models, while the HT-PBTK models parameterized using QSPR model predictions showed RMSLE values for C_{max} ranging from 0.45 to 0.76, with the ensemble QSPR model predictions being 0.67. Overall, predictions of V_d had RMSLE ranging between 0.65 and 0.92. Here, we have assumed that F_{bio} only depends on first-pass metabolism, as informed by Cl_{int}. This bioavailability assumption does not appear to explain the observed discrepancies with *in vivo* data. The uncertainty analysis (Fig. 10) confirmed that improving HTTK estimates of the elimination phase would be most likely to enhance prediction accuracy. Two of the QSPR models predicted $t_{½}$, so a QSPR prediction for V_d was needed to compare the predictions to *in vivo* data. However, as we see in Fig. 10, V_d is predicted relatively accurately from physico-chemical properties. For a set of 1352 pharmaceuticals Bouarar et al. (2019) found that they could directly predict V_d (RMSE 0.208), CL_{tot} (RMSE 0.103), and $t_{½}$ (RMSE 0.154).

Mathew et al. (2021) showed that QSPR models trained to predict TK summary statistics (V_d in that case) outperform more mechanistic PBTK models. Further, QSPR models trained to true human $t_{½}$ values (that is, QSARINS-Chem and IFS-QSAR) are trained to more chemicals (1105 chemicals with empirical *in vivo* $t_{½}$ rather than just those with *in vitro* HTTK data) and capture whole body clearance including whole-body level biotransformation (liver and other tissues), renal clearance, and exhalation. QSPR predicted $t_{½}$ can be directly used in IVIVE of *in vitro* bioactivity (NAMs) and help interpret biomonitoring data and parameterize exposure and bioaccumulation models. The $t_{½}$ predictions with the V_d predicted by the ensemble model gave more accurate predictions of whole-body clearance (and therefore of C_{ss}) compared to the *in vitro* driven HTTK models. However, PBTK models permit greater insight into why values vary between chemicals and, more importantly, allow simulation of human variability (Barton et al., 2007; Breen et al., 2022).

While *in vitro* NAMs aim to decrease reliance on *in vivo* animal testing (Nelson et al., 2024), *in silico* NAMs, such as HTTK QSPR models, have the potential to reduce the need for *in vitro* testing (Pelkonen et al., 2011). For a novel chemical, we hope the prediction errors estimated here provided conservative values of the accuracy of using HTTK with QSPR models. We believe that approaches like these can help inform public health risk-based prioritization of many more chemicals in commerce and the environment.

CRediT authorship contribution statement

John F. Wambaugh: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Nisha S. Sipes:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Gilberto Padilla Mercado:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Jon A. Arnot:** Writing – review & editing, Validation, Software, Investigation. **Linda Bertato:** Writing – review & editing, Software. **Trevor N. Brown:** Writing – review & editing, Software. **Nicola Chirico:** Writing – review & editing, Software. **Christopher Cook:** Visualization, Software, Data curation. **Daniel E. Dawson:** Validation, Software. **Sarah E. Davidson-Fritz:** Software. **Stephen S. Ferguson:** Writing – review & editing, Methodology, Conceptualization. **Michael-Rock Goldsmith:** Writing – review & editing, Methodology. **Richard S. Judson:** Writing – review & editing, Software, Methodology, Conceptualization. **Kamel Mansouri:** Writing – review & editing, Software, Methodology, Conceptualization. **Grace Patlewicz:** Writing – review & editing, Software, Methodology. **Ester Papa:** Writing – review & editing, Software, Methodology. **Prachi Pradeep:** Software. **Alessandro Sangion:** Writing – review & editing, Software. **Risa R. Sayre:** Writing – review & editing, Supervision, Methodology, Data curation, Conceptualization. **Russell S. Thomas:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Rogelio Tornero-Velez:** Writing – review & editing, Supervision, Software, Methodology. **Barbara A. Wetmore:** Writing – review & editing, Methodology, Conceptualization. **Michael J. Devito:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization.

Disclaimer

The views expressed in this publication are those of the authors and do not necessarily represent the views or policies of the U.S. EPA. Reference to commercial products or services does not constitute endorsement.

Funding

Employees of the United States Environmental Protection Agency (EPA) were funded through EPA's Office of Research and Development (ORD). The Oak Ridge Institute for Science and Education provided funding for G. Padilla Mercado and C. Cook. The National Institute of Environmental Health Sciences provided funding for S. Ferguson and K. Mansouri.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Given his role as NAMS Journal Editorial Board Member, John Wambaugh has no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank John DiBella and Dr. Michael Lawless of Simulations Plus for helpful conversations and background on ADMet Predictor. We thank Drs. Marina Evans and Todd Zurlinden for their helpful U.S. EPA internal reviews of the manuscript. We thank Dr. Xiaoqing Chang for her helpful National Institute of Environmental Health Sciences internal review.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tiv.2025.106150.

Data availability

All data and analysis scripts are publicly available at: https://github.com/USEPA/CompTox-Expocast-HTTKQSPRs

References

Adday, G.H., Subramaniam, S.K., Zukarnain, Z.A., Samian, N., 2023. Friendship degree and tenth man strategy: A new method for differentiating between erroneous readings and true events in wireless sensor networks. IEEE Access 11, 127651–127668.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723.
Alhijawi, B., Awajan, A., 2024. Genetic algorithms: theory, genetic operators, solutions, and applications. Evol. Intel. 17, 1245–1256.
Armitage, J.M., Hughes, L., Sangion, A., Arnot, J.A., 2021. Development and intercomparison of single and multicompartment physiologically-based toxicokinetic models: implications for model selection and tiered modeling frameworks. Environ. Int. 154, 106557.
Arnot, J.A., Brown, T.N., Wania, F., 2014. Estimating screening-level organic chemical half-lives in humans. Environ. Sci. Technol. 48, 723–730.
Barton, H.A., Chiu, W.A., Setzer, R.W., Andersen, M.E., Bailer, A.J., Bois, F.Y., DeWoskin, R.S., Hays, S., Johanson, G., Jones, N., 2007. Characterizing uncertainty and variability in physiologically based pharmacokinetic models: state of the science and needs for research and implementation. Toxicol. Sci. 99, 395–402.
Baumer, B., Udwin, D., 2015. R markdown. Wiley Interdiscip. Rev. 7, 167–177.
Bell, S.M., Chang, X., Wambaugh, J.F., Allen, D.G., Bartels, M., Brouwer, K.L.R., Casey, W.M., Choksi, N., Ferguson, S.S., Fraczkiewicz, G., Jarabek, A.M., Ke, A., Lumen, A., Lynn, S.G., Paini, A., Price, P.S., Ring, C., Simon, T.W., Sipes, N.S., Sprankle, C.S., Strickland, J., Troutman, J., Wetmore, B.A., Kleinstreuer, N.C., 2018. In vitro to in vivo extrapolation for high throughput prioritization and decision making. Toxicol. in Vitro 47, 213–227.
Bernstein, A.S., Kapraun, D.F., Schlosser, P.M., 2021. A model template approach for rapid evaluation and application of physiologically based pharmacokinetic models for use in human health risk assessments: a case study on per-and polyfluoroalkyl substances. Toxicol. Sci. 182, 215–228.
Black, S.R., Nichols, J.W., Fay, K.A., Matten, S.R., Lynn, S.G., 2021. Evaluation and comparison of in vitro intrinsic clearance rates measured using cryopreserved hepatocytes from humans, rats, and rainbow trout. Toxicology 457, 152819.
Bouarar, I., Brasseur, G., Petersen, K., Granier, C., Fan, Q., Wang, X., Wang, L., Ji, D., Liu, Z., Xie, Y., 2019. Influence of anthropogenic emission inventories on simulations of air quality in China during winter and summer 2010. Atmos. Environ. 198, 236–256.
Breen, M., Ring, C.L., Kreutz, A., Goldsmith, M.-R., Wambaugh, J.F., 2021. High-throughput PBTK models for in vitro to in vivo extrapolation. Expert Opin. Drug Metab. Toxicol. 17, 903–921.
Breen, M., Wambaugh, J.F., Bernstein, A., Sfeir, M., Ring, C.L., 2022. Simulating Toxicokinetic variability to identify susceptible and highly exposed populations. J. Expos. Sci. Environ. Epidemiol. 32, 855–863.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Cappelli, C.I., Benfenati, E., Cester, J., 2015. Evaluation of QSAR models for predicting the partition coefficient (logP) of chemicals under the REACH regulation. Environ. Res. 143, 26–32.
Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250.
Chang, X., Tan, Y.-M., Allen, D.G., Bell, S., Brown, P.C., Browning, L., Ceger, P., Gearhart, J., Hakkinen, P.J., Kabadi, S.V., Kleinstreuer, N.C., Lumen, A., Matheson, J., Paini, A., Pangburn, H.A., Petersen, E.J., Reinke, E.N., Ribeiro, A.J.S., Sipes, N., Sweeney, L.M., Wambaugh, J.F., Wange, R., Wetmore, B.A., Mumtaz, M., 2022. IVIVE: facilitating the use of in vitro toxicity data in risk assessment and decision making. Toxics 10, 232.
Chirico, N., Bertato, L., Papa, E., 2021a. In Vitro Biotransformation Prediction-Suite (IVBP-Suite).
Chirico, N., Sangion, A., Gramatica, P., Bertato, L., Casartelli, I., Papa, E., 2021b. QSARINS-Chem standalone version: A new platform-independent software to profile chemicals for physico-chemical properties, fate, and toxicity. J. Comput. Chem. 42, 1452–1460.
Chou, W.-C., Lin, Z., 2023. Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. Toxicol. Sci. 191, 1–14.
Cortes, C., 1995. Support-vector networks. Mach. Learn. 20, 273–297.
Davies, B., Morris, T., 1993. Physiological parameters in laboratory animals and humans. Pharm. Res. 10, 1093–1095.

Dawson, D.E., Ingle, B.L., Phillips, K.A., Nichols, J.W., Wambaugh, J.F., Tornero-Velez, R., 2021. Designing QSARs for parameters of high-throughput toxicokinetic models using open-source descriptors. Environ. Sci. Technol. 55, 6505–6517.
Di Lascio, E., Gerebtzoff, G., Rodríguez-Pérez, R., 2023. Systematic evaluation of local and global machine learning models for the prediction of ADME properties. Mol. Pharm. 20, 1758–1767.
Escher, S.E., Kamp, H., Bennekou, S.H., et al., 2019. Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project. Arch Toxicol 93, 3643–3667. https://doi.org/10.1007/s00204-019-02591-7.
Fentem, J., Malcomber, I., Maxwell, G., Westmoreland, C., 2021. Upholding the EU'S commitment to 'animal testing as a last resort' under REACH requires a paradigm shift in how we assess chemical safety to close the gap between regulatory testing and modern safety science. Altern. Lab. Anim 49, 122–132.
Gadaleta, D., Serrano-Candelas, E., Ortega-Vallbona, R., Colombo, E., Garcia de Lomana, M., Biava, G., Aparicio-Sánchez, P., Roncaglioni, A., Gozalbes, R., Benfenati, E., 2024. Comprehensive benchmarking of computational tools for predicting toxicokinetic and physicochemical properties of chemicals. J. Chemother. 16, 145.
Geci, R., Gadaleta, D., de Lomana, M.G., Ortega-Vallbona, R., Colombo, E., Serrano-Candelas, E., Paini, A., Kuepfer, L., Schaller, S., 2024. Systematic evaluation of high-throughput PBK modelling strategies for the prediction of intravenous and oral pharmacokinetics in humans. Arch. Toxicol. 98, 2659–2676.
Goldsmith, M.R., Grulke, C.M., Brooks, R.D., Transue, T.R., Tan, Y.M., Frame, A., Egeghy, P.P., Edwards, R., Chang, D.T., Tornero-Velez, R., Isaacs, K., Wang, A., Johnson, J., Holm, K., Reich, M., Mitchell, J., Vallero, D.A., Phillips, L., Phillips, M., Wambaugh, J.F., Judson, R.S., Buckley, T.J., Dary, C.C., 2014. Development of a consumer product ingredient database for chemical exposure screening and prioritization. Food Chem. Toxicol. 65, 269–279.
Guha, R., 2007. Chemical informatics functionality in R. J. Stat. Softw. 18, 1–16.
Health Canada, 2021. Bioactivity Exposure Ratio: Application in Priority Setting and Risk Assessment. Ottawa, ON, Canada.
Honda, G.S., Pearce, R.G., Pham, L.L., Setzer, R.W., Wetmore, B.A., Sipes, N.S., Gilbert, J., Franz, B., Thomas, R.S., Wambaugh, J.F., 2019. Using the concordance of in vitro and in vivo data to evaluate extrapolation assumptions. PLoS One 14, e0217564.
Honda, G.S., Kenyon, E.M., Davidson-Fritz, S., Dinallo, R., El Masri, H., Korol-Bexell, E., Li, L., Angus, D., Pearce, R.G., Sayre, R.R., Strock, C., Thomas, R.S., Wetmore, B.A., Wambaugh, J.F., 2025. Impact of gut permeability on estimation of oral bioavailability for chemicals in commerce and the environment. ALTEX - Alternat. Anim. Exp. 42, 56–74.
Jamei, M., Dickinson, G.L., Rostami-Hodjegan, A., 2009a. A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of 'bottom-up'vs 'top-down'recognition of covariates. Drug Metab. Pharmacokinet. 24, 53–75.
Jamei, M., Marciniak, S., Feng, K., Barnett, A., Tucker, G., Rostami-Hodjegan, A., 2009b. The Simcyp® population-based ADME simulator. Expert Opin. Drug Metab. Toxicol. 5, 211–223.
Jeong, J., Kim, D., Choi, J., 2022. Application of ToxCast/Tox21 data for toxicity mechanism-based evaluation and prioritization of environmental chemicals: perspective and limitations. Toxicol. in Vitro 84, 105451.
Kapraun, D.F., Sfeir, M., Pearce, R., Davidson, S., Lumen, A., Dallmann, A., Judson, R., Wambaugh, J.F., 2022. Evaluation of a Rapid, Generic Human Gestational Dose Model.
Kavlock, R.J., Bahadori, T., Barton-Maclaren, T.S., Gwinn, M.R., Rasenberg, M., Thomas, R.S., 2018. Accelerating the pace of chemical risk assessment. Chem. Res. Toxicol. 31, 287–290.
Kirman, C.R., Aylward, L.L., Wetmore, B.A., Thomas, R.S., Sochaski, M., Ferguson, S.S., Csiszar, S.A., Jolliet, O., 2015. Quantitative property–property relationship for screening-level prediction of intrinsic clearance: a tool for exposure modeling for high-throughput toxicity screening data. Appl. In Vitro Toxicol. 1, 140–146.
Kitchin, R., McArdle, G., 2016. What makes big data, big data? Exploring the ontological characteristics of 26 datasets. Big Data Soc. 3, 2053951716631130.
Koman, P.D., Singla, V., Lam, J., Woodruff, T.J., 2019. Population susceptibility: A vital consideration in chemical risk evaluation under the Lautenberg toxic substances control act. PLoS Biol. 17, e3000372.
Krause, S., Goss, K.-U., 2018. The impact of desorption kinetics from albumin on hepatic extraction efficiency and hepatic clearance: a model study. Arch. Toxicol. 92, 2175–2182.
Krause, S., Goss, K.-U., 2021. Relevance of desorption kinetics and permeability for in vitro-based predictions of hepatic clearance in fish. Aquat. Toxicol. 235, 105825.
Kreutz, A., Chang, X., Hogberg, H.T., Wetmore, B.A., 2024. Advancing understanding of human variability through toxicokinetic modeling, in vitro-in vivo extrapolation, and new approach methodologies. Hum. Genom. 18, 1–23.
Lillicrap, A., Belanger, S., Burden, N., Pasquier, D.D., Embry, M.R., Halder, M., Lampi, M. A., Lee, L., Norberg-King, T., Rattner, B.A., Schirmer, K., Thomas, P., 2016. Alternative approaches to vertebrate ecotoxicity tests in the 21st century: A review of developments over the last 2 decades and current status. Environ. Toxicol. Chem. 35, 2637–2646.
Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues. BioScience 51.
Linakis, M.W., Sayre, R.R., Pearce, R.G., Sfeir, M.A., Sipes, N.S., Pangburn, H.A., Gearhart, J.M., Wambaugh, J.F., 2020. Development and evaluation of a high throughput inhalation model for organic chemicals. J. Expos. Sci. Environ. Epidemiol. 30, 866–877.
Lynn, S.G., Schultz, I.R., Matten, S.R., Patel, P.R., Watson, S.L., Yueh, Y.L., Black, S.R., Wetmore, B.A., 2025. Cross-species comparisons of plasma binding and considerations for data evaluation. Toxicol. in Vitro 106036.
Maertens, A., Golden, E., Luechtefeld, T.H., Hoffmann, S., Tsaioun, K., Hartung, T., 2022. Probabilistic risk assessment - the keystone for the future of toxicology. Altex 39, 3–29.
Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J. Chemother. 10, 10.
Mansouri, K., Chang, X., Allen, D., Judson, R., Williams, A., Kleinstreuer, N., 2021. OPERA models for ADME properties and toxicity endpoint. In: Society of Toxicology Annual Meeting.
Mathew, S., Tess, D., Burchett, W., Chang, G., Woody, N., Keefer, C., Orozco, C., Lin, J., Jordan, S., Yamazaki, S., 2021. Evaluation of prediction accuracy for volume of distribution in rat and human using in vitro, in vivo, PBPK and QSAR methods. J. Pharm. Sci. 110, 1799–1823.
Moxon, T.E., Li, H., Lee, M.-Y., Piechota, P., Nicol, B., Pickles, J., Pendlington, R., Sorrell, I., Baltazar, M.T., 2020. Application of physiologically based kinetic (PBK) modelling in the next generation risk assessment of dermally applied consumer products. Toxicol. in Vitro 63, 104746.
Musther, H., Olivares-Morales, A., Hatley, O.J., Liu, B., Hodjegan, A.R., 2014. Animal versus human oral drug bioavailability: do they correlate? Eur. J. Pharm. Sci. 57, 280–291.
National Academies of Sciences, E., Medicine, 2017. Using 21st Century Science to Improve Risk-Related Evaluations. The National Academies Press, Washington, DC.
National Research Council, 1983. Risk Assessment in the Federal Government: Managing the Process. National Academies Press, Washington (DC).
Nelson, C.P., Brown, P., Fitzpatrick, S., Ford, K.A., Howard, P.C., MacGill, T., Margerrison, E.E.C., O'Shaughnessy, J., Patterson, T.A., Raghuwanshi, R., Rouse, R., Stromgren, S., Sung, K.E., Valerio, L.G., Ward, J.L., Bumpus, N.N., 2024. Advancing alternative methods to reduce animal testing. Science 386, 724–726.
Nicolas, C.I., Mansouri, K., Phillips, K.A., Grulke, C.M., Richard, A.M., Williams, A.J., Rabinowitz, J., Isaacs, K.K., Yau, A., Wambaugh, J.F., 2018. Rapid experimental measurements of physicochemical properties to inform models and testing. Sci. Total Environ. 636, 901–909.
O'Flaherty, E.J., 1981. Toxicants and Drugs: Kinetics and Dynamics. John Wiley & Sons.
Organisation for Economic Co-operation and Development, 2004. The Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q) SARs] on the Principles for the Validation of (Q) SARs. In: Series on Testing and Assesment, p. 206.
Organisation for Economic Co-operation and Development, 2014. Guidance document on the validation of (quantitative) structure-activity relationship [(Q) SAR] models. Organisation for Economic Co-operation and Development.
Padilla Mercado, G., Cook, C., Adkins, N., Albrecht, L., Cary, G., Edwards, B., Haggard, D. E., Hanley, N.M., Hughes, M.F., Jarnagin, A., Kodavanti, T.D., Korol-Bexell, E., Kreutz, A., Ngo, M., Patullo, C., Rowan, E.G., Huse, L.M., Correa, V.A., Kesic, B., Casey, W., Aboabdo, J., Wolf, K., Sayre, R., Sharma, B., Wall, J.T., Yamazaki, H., Wambaugh, J.F., Ring, C.L., 2025. Informatics for toxicokinetics. J. Pharmacokinet. Pharmacodyn. 52, 30.
Paini, A., Cole, T., Meinero, M., Carpi, D., Deceuninck, P., Macko, P., Palosaari, T., Sund, J., Worth, A., Whelan, M., 2020. In: J.R.C.J. European Commission (Ed.), EURL ECVAM in vitro hepatocyte clearance and blood plasma protein binding dataset for 77 chemicals.
Papa, E., Sangion, A., Arnot, J.A., Gramatica, P., 2018. Development of human biotransformation QSARs and application for PBT assessment refinement. Food Chem. Toxicol. 112, 535–543.
Paul Friedman, K., Gagne, M., Loo, L.-H., Karamertzanis, P., Netzeva, T., Sobanski, T., Franzosa, J.A., Richard, A.M., Lougee, R.R., Gissi, A., 2020. Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. Toxicol. Sci. 173, 202–225.
Paul Friedman, K., Thomas, R.S., Wambaugh, J.F., Harrill, J.A., Judson, R.S., Shafer, T.J., Williams, A.J., Lee, J.-Y.J., Loo, L.-H., Gagné, M., 2025. Integration of new approach methods for the assessment of data poor chemicals. Toxicol. Sci. kfaf019.
Pearce, R.G., Setzer, R.W., Davis, J.L., Wambaugh, J.F., 2017a. Evaluation and calibration of high-throughput predictions of chemical distribution to tissues. J. Pharmacokinet. Pharmacodyn. 44, 549–565.
Pearce, R.G., Setzer, R.W., Strope, C.L., Sipes, N.S., Wambaugh, J.F., 2017b. httk: R package for high-throughput toxicokinetics. J. Stat. Softw. 79, 1–26.
Pelkonen, O., Turpeinen, M., Raunio, H., 2011. In vivo-in vitro-in silico pharmacokinetic modelling in drug development. Clin. Pharmacokinet. 50, 483–491.
Peyret, T., Poulin, P., Krishnan, K., 2010. A unified algorithm for predicting partition coefficients for PBPK modeling of drugs and environmental chemicals. Toxicol. Appl. Pharmacol. 249, 197–207.
Pradeep, P., Patlewicz, G., Pearce, R., Wambaugh, J., Wetmore, B., Judson, R., 2020. Using chemical structure information to develop predictive models for in vitro toxicokinetic parameters to inform high-throughput risk-assessment. Comput. Toxicol. 16, 100136.
Punt, A., Bouwmeester, H., Blaauboer, B.J., Coecke, S., Hakkert, B., Hendriks, D.F., Jennings, P., Kramer, N.I., Neuhoff, S., Masereeuw, R., 2020. New approach methodologies (NAMs) for human-relevant biokinetics predictions: meeting the paradigm shift in toxicology towards an animal-free chemical risk assessment. Altex 37, 607–622.
R Core Team, 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Reale, E., Zare Jeddi, M., Paini, A., Connolly, A., Duca, R., Cubadda, F., Benfenati, E., Bessems, J., Galea, K., Dirven, H., Santonen, T., Koch, H., Jones, K., Sams, C.,

Viegas, S., Kyriaki, M., Campisi, L., David, A., Antignac, J.-P., Hopf, N., 2024. Human biomonitoring and toxicokinetics as key building blocks for next generation risk assessment. Environ. Int. 184, 108474.
Richard, A.M., Judson, R.S., Houck, K.A., Grulke, C.M., Volarath, P., Thillainadarajah, I., Yang, C., Rathman, J., Martin, M.T., Wambaugh, J.F., 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. Chem. Res. Toxicol. 29, 1225–1251.
Richard, A.M., Tao, D., LeClair, C.A., Leister, W., Tretyakov, K.V., White, E.V., Lewis, K. C., Sefler, A., Shinn, P., Collins, B.J., Nguyen, D.-T., Ye, L., Zhao, T., Xu, T., Williams, A.J., Waidyanatha, S., Thomas, R.S., Tice, R., Simeonov, A., Huang, R., 2025. Analytical quality evaluation of the Tox21 compound library. Chem. Res. Toxicol. 38, 15–41.
Rotroff, D.M., Wetmore, B.A., Dix, D.J., Ferguson, S.S., Clewell, H.J., Houck, K.A., LeCluyse, E.L., Andersen, M.E., Judson, R.S., Smith, C.M., Sochaski, M.A., Kavlock, R.J., Boellmann, F., Martin, M.T., Reif, D.M., Wambaugh, J.F., Thomas, R. S., 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol. Sci. 117, 348–358.
Rydberg, P., Gloriam, D.E., Zaretzki, J., Breneman, C., Olsen, L., 2010. SMARTCyp: A 2D method for prediction of cytochrome P450-mediated drug metabolism. ACS Med. Chem. Lett. 1, 96–100.
Sayre, R.R., Wambaugh, J.F., Grulke, C.M., 2020. Database of Pharmacokinetic Time-Series Data and Parameters for 144 Environmental Chemicals. Scientific Data.
Schapire, R.E., Freund, Y., 2012. Foundations of Machine Learning. Boosting: Foundations and Algorithms. The MIT Press, p. 0.
Scheife, R.T., 1989. Protein binding: what does it mean? Ann. Pharmacother. 23, S27–S31.
Schmitt, W., 2008. General approach for the calculation of tissue to plasma partition coefficients. Toxicol. in Vitro 22, 457–467.
Shibata, Y., Takahashi, H., Chiba, M., Ishii, Y., 2002. Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method. Drug Metab. Dispos. 30, 892–896.
Sipes, N.S., Wambaugh, J.F., Pearce, R., Auerbach, S.S., Wetmore, B.A., Hsieh, J.-H., Shapiro, A.J., Svoboda, D., DeVito, M.J., Ferguson, S.S., 2017. An intuitive approach for predicting potential human health risk with the Tox21 10k library. Environ. Sci. Technol. 51, 10786–10796.
Sobus, J.R., Tan, Y.-M., Pleil, J.D., Sheldon, L.S., 2011. A biomonitoring framework to support exposure and risk assessments. Sci. Total Environ. 409, 4875–4884.
Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R., Willighagen, E.L., 2006. Recent developments of the chemistry development kit (CDK)-an open-source java library for chemo-and bioinformatics. Curr. Pharm. Des. 12, 2111–2120.
Sterling, T., Irwin, J.J., 2015. ZINC 15 – ligand discovery for everyone. J. Chem. Inf. Model. 55, 2324–2337.
Tan, Y.-M., Liao, K.H., Conolly, R.B., Blount, B.C., Mason, A.M., Clewell, H.J., 2006. Use of a physiologically based pharmacokinetic model to identify exposures consistent with human biomonitoring data for chloroform. J. Toxicol. Environ. Health A 69, 1727–1756.
Taunk, K., De, S., Verma, S., Swetapadma, A., 2019. A brief review of nearest neighbor algorithm for learning and classification. In: 2019 International Conference on Intelligent Computing and Control Systems (ICCS), pp. 1255–1260.
Thomas, R.S., Paules, R.S., Simeonov, A., Fitzpatrick, S.C., Crofton, K.M., Casey, W.M., Mendrick, D.L., 2018. The US federal Tox21 program: A strategic and operational plan for continued leadership. Altex 35, 163–168.
Tibshirani, R., 2011. Regression shrinkage and selection via the Lasso: A retrospective. J. R. Stat. Soc. Ser. B Stat Methodol. 73, 273–282.
Tonnelier, A., Coecke, S., Zaldívar, J.-M., 2012. Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model. Arch. Toxicol. 86, 393–403.
Truong, K., Wambaugh, J., Kapraun, D., Davidson-Fritz, S., Eytcheson, S., Judson, R., Friedman, K.P., 2025. Interpretation of thyroid-relevant bioactivity data for comparison to in vivo exposures: A prioritization approach for putative chemical inhibitors of in vitro deiodinase activity. Toxicology 154157.
U.S. Congress, 2016. Frank R. Lautenberg Chemical Safety for the 21st Century Act. Pub L 114–182 (114th Congress, 22 June 2016). Washington, D.C.
U.S. Environmental Protection Agency, 2018. Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA (Toxic Substances Control Act) Program.
U.S. Environmental Protection Agency, 2021. U.S. EPA. A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA. Washington, DC.
Wambaugh, J.F., Rager, J.E., 2022. Exposure forecasting–ExpoCast–for data-poor chemicals in commerce and the environment. J. Expo. Sci. Environ. Epidemiol. 32, 783–793.
Wambaugh, J.F., Hughes, M.F., Ring, C.L., MacMillan, D.K., Ford, J., Fennell, T.R., Black, S.R., Snyder, R.W., Sipes, N.S., Wetmore, B.A., Westerhout, J., Setzer, R.W., Pearce, R.G., Simmons, J.E., Thomas, R.S., 2018. Evaluating in vitro-in vivo extrapolation of toxicokinetics. Toxicol. Sci. 163, 152–169.
Wambaugh, J.F., Wetmore, B.A., Ring, C.L., Nicolas, C.I., Pearce, R., Honda, G., Dinallo, R., Angus, D., Gilbert, J., Sierra, T., Badrinarayanan, A., Snodgrass, B., Brockman, A., Strock, C., Setzer, W., Thomas, R.S., 2019. Assessing toxicokinetic uncertainty and variability in risk prioritization. Toxicol. Sci. 172, 235–251.
Wambaugh, J.F., Schacht, C.M., Ring, C.L., 2025. A simple physiologically based Toxicokinetic model for multi-route in vitro–in vivo extrapolation. Environ. Sci. Technol. Lett. 12, 261–268.
Wang, Y.-H., 2010. Confidence assessment of the Simcyp time-based approach and a static mathematical model in predicting clinical drug-drug interactions for mechanism-based CYP3A inhibitors. Drug Metab. Dispos. 38, 1094–1104.
Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M.V., 2024. gplots: various R programming tools for plotting data. R package version 3.1.3.1. https://CRAN.R-project.org/package=gplots.
Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36.
Weitekamp, C.A., Paul Friedman, K., Harrill, A.H., Auerbach, S., Bandele, O., Barton-Maclaren, T.S., Fitzpatrick, S., Mezencev, R., Santillo, M., Simanainen, U., Smith, D., Whelan, M., Thomas, R.S., 2025. Quantitative and qualitative concordance between clinical and nonclinical toxicity data. Toxicol. Sci. 206, 253–272.
Wetmore, B.A., 2015. Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology 332, 94–101.
Wetmore, B.A., Wambaugh, J.F., Ferguson, S.S., Sochaski, M.A., Rotroff, D.M., Freeman, K., Clewell 3rd, H.J., Dix, D.J., Andersen, M.E., Houck, K.A., Allen, B., Judson, R.S., Singh, R., Kavlock, R.J., Richard, A.M., Thomas, R.S., 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125, 157–174.
Wetmore, B.A., Wambaugh, J.F., Ferguson, S.S., Li, L., Clewell 3rd, H.J., Judson, R.S., Freeman, K., Bao, W., Sochaski, M.A., Chu, T.M., Black, M.B., Healy, E., Allen, B., Andersen, M.E., Wolfinger, R.D., Thomas, R.S., 2013. Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. Toxicol. Sci. 132, 327–346.
Wetmore, B.A., Wambaugh, J.F., Allen, B., Ferguson, S.S., Sochaski, M.A., Setzer, R.W., Houck, K.A., Strope, C.L., Cantwell, K., Judson, R.S., LeCluyse, E., Clewell, H.J., Thomas, R.S., Andersen, M.E., 2015. Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. Toxicol. Sci. 148, 121–136.
Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., 2017. The CompTox chemistry dashboard: a community data resource for environmental chemistry. J. Chemother. 9, 61.
Yang, C., Tarkhov, A., Marusczyk, J.R., Bienfait, B., Gasteiger, J., Kleinoeder, T., Magdziarz, T., Sacher, O., Schwab, C.H., Schwoebel, J., 2015. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. J. Chem. Inf. Model. 55, 510–528.
Yang, H., Lou, C., Li, W., Liu, G., Tang, Y., 2020. Computational approaches to identify structural alerts and their applications in environmental toxicology and drug discovery. Chem. Res. Toxicol. 33, 1312–1322.
Yap, C.W., 2011. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32, 1466–1474.
Zaretzki, J., Rydberg, P., Bergeron, C., Bennett, K.P., Olsen, L., Breneman, C.M., 2012. RS-predictor models augmented with SMARTCyp reactivities: robust metabolic regioselectivity predictions for nine CYP isozymes. J. Chem. Inf. Model. 52, 1637–1659.