IN SILICO

Systematic evaluation of high-throughput PBK modelling strategies for the prediction of intravenous and oral pharmacokinetics in humans

René Geci[1,2] · Domenico Gadaleta[3] · Marina García de Lomana[4] · Rita Ortega-Vallbona[5] · Erika Colombo[3] · Eva Serrano-Candelas[5] · Alicia Paini[1] · Lars Kuepfer[2] · Stephan Schaller[1]

Abstract

Physiologically based kinetic (PBK) modelling offers a mechanistic basis for predicting the pharmaco-/toxicokinetics of compounds and thereby provides critical information for integrating toxicity and exposure data to replace animal testing with in vitro or in silico methods. However, traditional PBK modelling depends on animal and human data, which limits its usefulness for non-animal methods. To address this limitation, high-throughput PBK modelling aims to rely exclusively on in vitro and in silico data for model generation. Here, we evaluate a variety of in silico tools and different strategies to parameterise PBK models with input values from various sources in a high-throughput manner. We gather 2000 + publicly available human in vivo concentration–time profiles of 200 + compounds (IV and oral administration), as well as in silico, in vitro and in vivo determined compound-specific parameters required for the PBK modelling of these compounds. Then, we systematically evaluate all possible PBK model parametrisation strategies in PK-Sim and quantify their prediction accuracy against the collected in vivo concentration–time profiles. Our results show that even simple, generic high-throughput PBK modelling can provide accurate predictions of the pharmacokinetics of most compounds (87% of Cmax and 84% of AUC within tenfold). Nevertheless, we also observe major differences in prediction accuracies between the different parameterisation strategies, as well as between different compounds. Finally, we outline a strategy for high-throughput PBK modelling that relies exclusively on freely available tools. Our findings contribute to a more robust understanding of the reliability of high-throughput PBK modelling, which is essential to establish the confidence necessary for its utilisation in Next-Generation Risk Assessment.

Lars Kuepfer and Stephan Schaller share last authorship.

✉ René Geci
rene.geci@esqlabs.com

1 esqLABS GmbH, Saterland, Germany

2 Institute for Systems Medicine with Focus on Organ Interaction, University Hospital RWTH Aachen, Aachen, Germany

3 Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

4 Machine Learning Research, Research and Development, Pharmaceuticals, Bayer AG, Berlin, Germany

5 ProtoQSAR SL, CEEI (Centro Europeo de Empresas Innovadoras), Valencia, Spain

Abbreviations

ADME	Absorption, distribution, metabolism, and excretion
AI	Artificial intelligence
AUC	Area under the curve
CL	Clearance
CLint	Intrinsic hepatic clearance
Cmax	Maximum concentration
HT-PBK	High-throughput PBK
IV	Intravenous
LogD	Logarithm of distribution coefficient
LogMA	Logarithm of membrane affinity partition coefficient
LogP	Logarithm of octanol–water partition coefficient
ML	Machine learning

Tmax	Time to maximum concentration
PK	Pharmacokinetics
TK	Toxicokinetics
PBK	Physiologically based kinetic
PBPK	Physiologically based pharmacokinetic
PBTK	Physiologically based toxicokinetic
PO	Peroral

Introduction

Pharmacokinetics (PK) and toxicokinetics (TK), the study of the distribution of drugs or chemicals within the body over time, are fundamental for understanding the desired or undesired effects of compounds on human health. Physiologically based pharmacokinetic (PBPK) or physiologically based toxicokinetic (PBTK) models, hereafter summarised as physiologically based kinetic (PBK) models, are a well-established computational method for simulating the PK of drugs, or TK of chemicals (Jones and Rowland-Yeo 2013). PBK models incorporate anatomical and physiological knowledge about the body, such as tissue compositions and blood flow rates, as well as compound-specific properties like the lipophilicity or solubility of compounds, to simulate the absorption, distribution, metabolism, and excretion (ADME) processes that determine compound concentrations in the body. This makes PBK models powerful tools that enable a mechanism-based prediction of compounds' concentration–time profiles in plasma and body tissues, even those otherwise inaccessible for direct sampling.

In the pharmaceutical industry, PBK models play a crucial role in drug discovery and development, cross-species extrapolation, the guiding of dosing regimens, extrapolation to special populations, and the prediction of potential drug–drug interactions (Thiel et al. 2015; Krstevska et al. 2022). Their use reduces drug failure rates and optimises trial protocols (Edginton et al. 2008; Lin et al. 2022). Further, PBK models are now also used outside the pharmaceutical field from which they originated. In toxicology, PBK models have become more widely adopted and aid in interpreting human biomonitoring data (Clewell et al. 2008), in the extrapolation from in vitro to in vivo data (Bouvier d'Yvoire et al. 2007; Blaauboer 2010; Yoon et al. 2012) and in chemical safety assessment (Paini et al. 2019, 2021). While pharmacology and toxicology have distinct objectives, both aim to understand the concentrations of compounds in the body, making PBK models a valuable tool in both fields.

Traditionally, generating PBK models is an iterative, time- and labour-intensive process. It is performed on a compound-by-compound basis and requires animal and human data for model parameterisation and validation. This makes the use of PBK models low-throughput and limits their usefulness for chemical risk assessment. Further, there is now a general desire in both the pharmaceutical and toxicological field to reduce the use of animal data to comply with the 3R principles (Törnqvist et al. 2014; Stokes 2015). For these reasons, there has been increasing interest to explore PBK modelling exclusively based on in vitro and in silico methods, as these methods have a greater potential to be applied in high-throughput and promise to reduce dependence on animal data. This new approach to PBK modelling is sometimes called high-throughput PBK (HT-PBK) modelling (Pearce et al. 2017; Breen et al. 2021; Naga et al. 2022; Khalidi et al. 2022) or Next-Generation PBK modelling (Paini et al. 2019; Punt et al. 2022a), either emphasising its rapid nature or that it does not rely on animal data.

The shift towards basing PBK modelling on in silico methods is further supported by recent advances in the fields of machine learning (ML) and artificial intelligence (AI) (Bender and Cortés-Ciriano 2021). ML and AI technologies are now increasingly being explored to provide rapid PK predictions without the need for new animal data. Sometimes, this is pursued using AI/ML methods to directly predict summary PK/TK parameters, like maximum concentration (Cmax) or area under the curve (AUC) values (Miljković et al. 2021; Fagerholm et al. 2021; Li et al. 2023). Other times, it is done by predicting mechanistically relevant compound properties, like the lipophilicity, solubility or clearance of compounds, which can then be used as inputs for making PK predictions using mechanistic models (Danishuddin et al. 2022; Pillai et al. 2022; Fagerholm et al. 2023; Mavroudis et al. 2023; Führer et al. 2024).

Until now, many in silico-based PK prediction efforts have focused on predicting rodent data (Schneckener et al. 2019; Kamiya et al. 2021; Naga et al. 2022; Punt et al. 2022b; Obrezanova et al. 2022; Mavroudis et al. 2023; Handa et al. 2023), presumably due to its greater availability than human data. While valuable, the ultimate goal in pharmacology and toxicology is to yield human-relevant conclusions. In those instances where such approaches were developed for predicting human data, evaluations were usually performed against relatively small datasets, typically not exceeding a few dozen compounds (Punt et al. 2022a; Li et al. 2023; Fagerholm et al. 2023). Moreover, PK prediction validations frequently relied on summary PK parameters, like Cmax or AUC values (Miljković et al. 2021). While they can provide useful insights, summary parameters also obscure prediction inaccuracies and do not fully reflect all intricacies of PBK model quality. This is why PBK models are traditionally evaluated against full concentration–time profiles instead. Furthermore, a number of in silico tools capable of predicting compound properties required for PBK modelling are available already (Benfenati et al. 2013; Mansouri et al. 2018; Xiong et al. 2021). However, to date, there has been no systematic evaluation of the usefulness of these

existing tools, or their various combinations, for HT-PBK modelling.

The aim of this work was to evaluate strategies for the high-throughput generation of human PBK models with input parameters from already available in silico-based property prediction tools. To this end, we compiled a large dataset of human in vivo concentration–time profiles after intravenous (IV) and oral (PO) administration (210 compounds, 2235 concentration–time profiles). Then, we systematically evaluated the predictive performances of PBK models when parameterised with the different in silico tools, and further compared their results to in vitro and in vivo benchmark references. This allowed us to identify which of the in silico tools provide the best input parameters for PBK modelling, as well as to validate the overall accuracy of such HT-PBK models for predicting IV and PO PK profiles in humans.

Materials and methods

Retrieval of pharmacokinetic data

An extensive description of the PK data retrieval process can be found in the Supplementary Information (SI-1). In short, we downloaded all concentration–time data from the CvT-DB (Sayre et al. 2020), OSP Observed Data Repository (Lippert et al. 2019), and PK-DB (Grzegorzewski et al. 2021) databases. We selected data measured in healthy adult humans after single IV or PO administration, excluding studies potentially dedicated to special patient cohorts like paediatric, geriatric, or diseased populations. Further, studies were excluded when concomitant treatments that might influence compounds' PK had occurred, as in drug-drug interaction studies or with special treatments such as grapefruit administration. Next, we manually extended this dataset by adding available literature data for compounds of relevance to the EU-funded ONTOX project (Vinken et al. 2021). We searched the literature for corresponding PK studies, for example using PKPDAI (Gonzalez Hernandez et al. 2021), and also did the same for compounds for which we had only obtained a single concentration–time profile before. From the retrieved literature, we manually extracted concentration–time profiles using WebPlotDigitizer version 4.6 (https://automeris.io/WebPlotDigitizer; Rohatgi 2022). Eventually, we had compiled a large dataset with multiple concentration–time profiles after IV and PO administration for most compounds. To ensure the consistency of the PK data, we visually checked dose-normalised plots of each compound and route and ensured that none of the PK study data were in extreme disagreement with each other. In the case of severe outliers, we manually investigated the causes of these differences, which in most cases were digitisation errors. We then either corrected such errors or, when it was

unexplainable why individual studies were in disagreement with the other studies, we excluded those outliers from the dataset.

Retrieval of compound property data

To evaluate different PBK model parameterisation strategies, we retrieved or predicted the required input properties of all compounds of which we had obtained PK data. The minimal input properties required for model parameterisation in PK-Sim are a compound's lipophilicity, pKa values, plasma or hepatic intrinsic clearance and its fraction unbound (Fu) (Kuepfer et al. 2016). Additionally, the simulation of oral administration also requires values for the solubility and intestinal permeability.

We generated lipophilicity predictions using six different in silico tools: LogP values by OCHEM (https://ochem.eu/home/show.do) and VEGA ALOGP (Ghose and Crippen 1987; Benfenati et al. 2013), LogD values by ADMETLab 2.0 (Xiong et al. 2021) and ADMETPredictor version 11.0.0.3 (henceforth called "SimPlus"; https://www.simulations-plus.com/software/admetpredictor), and predictions from Bayer's in-house models of LogD and LogMA (Göller et al. 2020). We further converted LogP or LogD values to LogMA-type values using regression relationships taken from Yun and Edginton (2013), Endo et al. (2011), and Loidl-Stahlhofen et al. (2001). pKa values were predicted using ChemAxon (Lee and Crippen 2009) and ADMETPredictor, aqueous solubility values using four in silico tools named ProtoPRED (https://protopred.protoqsar.com/), OPERA (Mansouri et al. 2018), ADMETPredictor, and ADMETLab 2.0. Further, Fasted State Simulated Intestinal Fluid (FaSSIF) and Fed State Simulated Intestinal Fluid (FeSSIF) solubility was also predicted using ADMETPredictor.

Intestinal permeability was predicted in the form of CACO2 values using ADMETLab 2.0, ProtoPRED and OPERA, as well as in the form of MDCK permeability using ADMETLab 2.0 and ADMETPredictor. Fu values were predicted using seven different in silico tools: ADMETLab 2.0, ADMETPredictor, pkCSM (Pires et al. 2015), Watanabe et al. (2018), OPERA, VEGA logK and VEGA CORAL (Toma et al. 2018). Plasma clearance values were predicted using ADMETLab 2.0, pkCSM and the ScitoVation clearance tool (https://scitovation-testing.shinyapps.io/clearancetoolgui). Hepatocyte CLint values were predicted with OPERA, ADMETPredictor and ADMET-AI (Swanson et al. 2023).

We also retrieved experimentally measured benchmark reference values of solubility, Fu, and clearance from public sources when they were available. In particular, aqueous solubility (LogS) data were retrieved from Hughes et al. (2008), and the CompTox Chemicals Dashboard (Williams

et al. 2017). Fu values were collected from Tonnelier et al. (2012), Yamazaki and Kanaoka (2004), Lombardo et al. (2002), Riley et al. (2005), Sohlenius-Sternbeck et al. (2010), Votano et al. (2006), Zhu et al. (2013), from the CompTox Chemicals Dashboard and the R httk package version 2.2.1 (Pearce et al. 2017). We also retrieved intrinsic hepatocyte clearance values from the R httk package version 2.2.1 (Pearce et al. 2017), as well as fitted PK-Sim intestinal permeability values from Willmann et al. (2004). Finally, we retrieved in vivo observed plasma clearance values from the literature, preferentially from IV studies, except for compounds for which no IV studies were available. When multiple values of a property were acquired for a single compound, the average values were used. And when experimental values of a compound were found in the previously described in silico tool's training set, those values were also added to the experimental data retrieved from the aforementioned datasets.

Software, model parameterisation and performance metrics

All PBK model simulations were performed with the standard whole-body PBK model implemented in PK-Sim version 11.1.137 (Willmann et al. 2003) executed from R using the ospsuite package version 11.1.143 and R version 4.2.2 (R Core Team 2022). For every retrieved PK study, a corresponding PBK model simulation was performed by parameterising a generic PBK model with (a) all compound-specific parameter values as provided by the different tools in the different parameterisation strategies, (b) the study-specific parameters like route of administration, dose and infusion duration (IV) or formulation (PO), if provided in the PK data, and (c) demographic parameters like sex, age, weight, and height of subjects, if provided in the PK data. When demographic parameters were not available, PK-Sim default settings (healthy adult male) were used for simulation.

Predicted or measured hepatic intrinsic clearance (CLint) values were scaled to in vivo liver clearance as

$$\text{In vivo intrinsic liver clearance} \left[\frac{\mu l}{\min} \right] = \text{CLint} \left[\frac{\mu l}{\min * 10^6 \text{ cells}} \right]$$
$$* \text{Hepatocellularity} \left[\frac{\text{cells}}{g} \right] * \text{Liver density} \left[\frac{g}{ml} \right] * \text{Liver volume [ml]},$$
$$(1)$$

Using the same values as Pearce et al. (2017) with a hepatocellularity of 1.1×10^8 cells/g (Ito and Houston 2004), a liver density of 1.05 g/ml (Snyder et al. 1979) and the PK-Sim default liver volume.

For oral simulations, we also set a formulation-specific dissolution parameter (80% dissolution time of Lint80 formulation as defined in PK-Sim), as some of the oral data was not obtained by administration of drugs in solution but,

for example, in the form of capsules or tablets which are not dissolved immediately upon administration. The values used for this were 25 min for "capsule" and 40 min for "tablet" formulations.

After simulating PBK models, the simulation results and the observed concentration–time data of corresponding PK studies were compared against each other and different model performance metrics were calculated to quantify PBK model quality: Log2-fold changes of predicted/observed values of Cmax, Tmax, AUC 0-last calculated using the linear up-log down method of the PKNCA package version 0.10.0 (Denney et al. 2015), as well as the percentage of datapoints within different fold ranges (1.5-/2-/3-/5-/tenfold). Further, we evaluated the goodness of the entire concentration–time profile prediction against all datapoints in the in vivo PK profile calculating the Relative and Absolute Log2 Errors as

$$\text{Relative Log2 Error} = \frac{1}{N} * \sum_{i=1}^{N} \text{Log2} \left(\frac{\text{Predicted value}_i}{\text{Observed value}_i} \right),$$
$$(2)$$

$$\text{Absolute Log2 Error} = \frac{1}{N} * \sum_{i=1}^{N} \left| \text{Log2} \left(\frac{\text{Predicted value}_i}{\text{Observed value}_i} \right) \right|,$$
$$(3)$$

for all N datapoints in every concentration–time profile such that the Relative Log2 Error captures biases in prediction, i.e., over- or underestimates (systematic error), while the Absolute Log2 Error quantifies overall closeness of PBK model simulations to the observed concentration–time values (random error).

When we had obtained multiple concentration–time profiles of a single compound, we selected the median performance value of all studies as the representative value for the compound to ensure that individual PK study outliers would not distort our results. To later summarise the overall performance of full parameterisation strategies, we followed the same approach to integrate the results of the various compounds in our dataset, so that when we refer to the Median Log2 Error, this means the median of the Log2 Error values of all the compounds in the evaluation dataset.

Results

PK data extraction and simulation strategy

After the retrieval of PK data from the different databases and the literature (SI-1), we initially obtained a total of 2235 healthy human adult in vivo concentration–time profiles of 210 unique compounds (Fig. 1A). For all compounds in this dataset, we then collected the various input data we intended to evaluate for PBK model parameterisation, for instance, lipophilicity values or predicted and measured
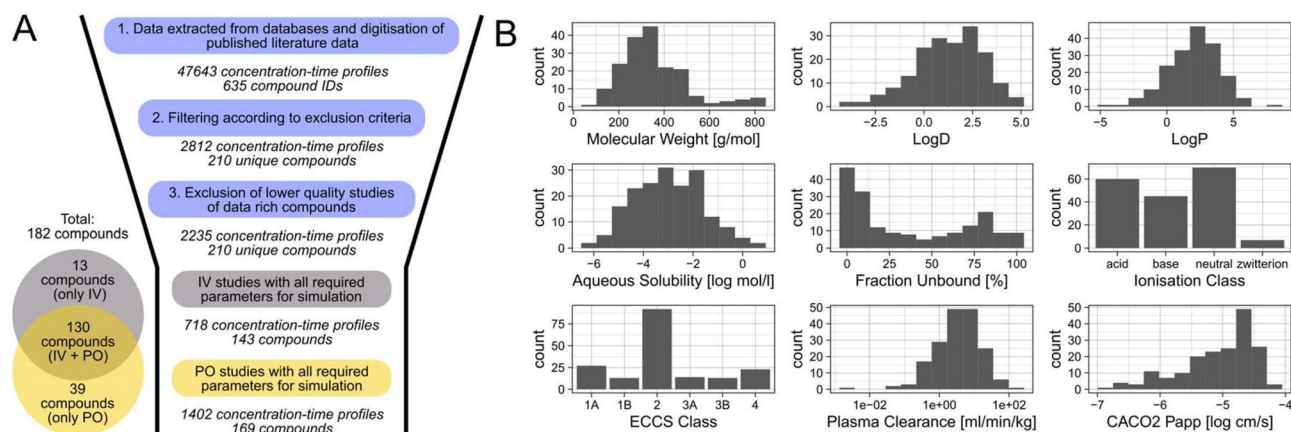
Fig. 1 Summary statistics of the collected PK dataset and of various compound properties. **A** Overview of the number of concentration–time profiles and compounds during and at the end of the PK data retrieval process. **B** Physico-chemical and pharmacokinetic properties of compounds in the retrieved PK dataset. LogD, LogP, aqueous solubility and CACO2 permeability are given as the mean of values predicted in silico by the different prediction tools used in this study. Plasma clearance values are the median of in vivo values collected from the literature. Ionisation at physiological pH (7.4) was predicted using ChemAxon

plasma clearances. This resulted in a complete dataset of 718 IV profiles (143 compounds) and 1402 PO profiles (169 compounds) for which all compounds also had the required input data. Notable exceptions to this were in vitro measured intrinsic hepatic clearance values (CLint), which were taken from httk and were available for only 97 compounds, and measured solubility values which were only available for 167 compounds. The 182 compounds in the final dataset were a diverse set of small molecules with a molecular weight of less than 900 Dalton, different ionisation states at physiological pH, and mostly belonged to ECCS class 2, suggesting that metabolism is driving their clearance (Fig. 2B).

Evaluating the performance of all PBK models parameterised from all the different input data sources against a large number of in vivo PK data requires many model simulations and leads to long computation times. For this reason, we split the simulation and analysis of the HT-PBK model parameterisation strategies into three steps, to be able to systematically evaluate all relevant parameterisation decisions one-at-a-time while keeping the number of required model simulations manageable (here 15 + million).

Briefly summarised, the rationale for our simulation and analysis strategy was as follows (Fig. 2). In the first step, we investigated which of the physico-chemical parameter sources performed best for predicting the passive distribution of compounds within the body. We generated, simulated, and evaluated every PBK model parameterisation strategy for all compounds of which IV data were available. To initially limit the variability, we only used in vivo observed plasma clearance and in vitro measured Fu values as high-quality benchmark reference values in the first step. In the second step, we then used the same IV data, but now

only using the best physico-chemical parameter predictions as determined in the first round, and then tested various Fu and clearance prediction tools to understand which of these would result in the best PK predictions. In the third step, we finally used the oral PK data for evaluation, along with the best physico-chemical, Fu and clearance prediction sources as determined in the previous steps. Then, we systematically varied the various solubility and intestinal permeability values to evaluate how to best predict oral absorption and to assess how well the best full high-throughput strategies would perform overall.

Step 1: evaluation of physico-chemical property predictions

In the first step, we systematically evaluated how to best set the physico-chemical PBK model parameters that determine the passive distribution of compounds within the body. In PK-Sim, these are primarily the lipophilicity of a compound, its pKa values, and the method used to predict the compound's partitioning coefficients. The lipophilicity values of compounds were predicted with six different in silico tools: three LogD prediction tools (SimPlus, ADMETLab, Bayer), two LogP tools (OCHEM and VEGA), and one LogMA tool (Bayer). pKa values were predicted using ChemAxon and SimPlus, and for comparison we also tested not providing pKa values, effectively assuming that all compounds were neutral. The five tested partitioning methods available in PK-Sim were PK-Sim (Willmann et al. 2005), Schmitt (2008), Rodgers and Rowland (2006), Poulin and Theil (2002), and Berezhkovskiy (2004). To limit the variability in this first analysis, we used in vivo observed plasma clearance and experimentally measured Fu values as high-quality
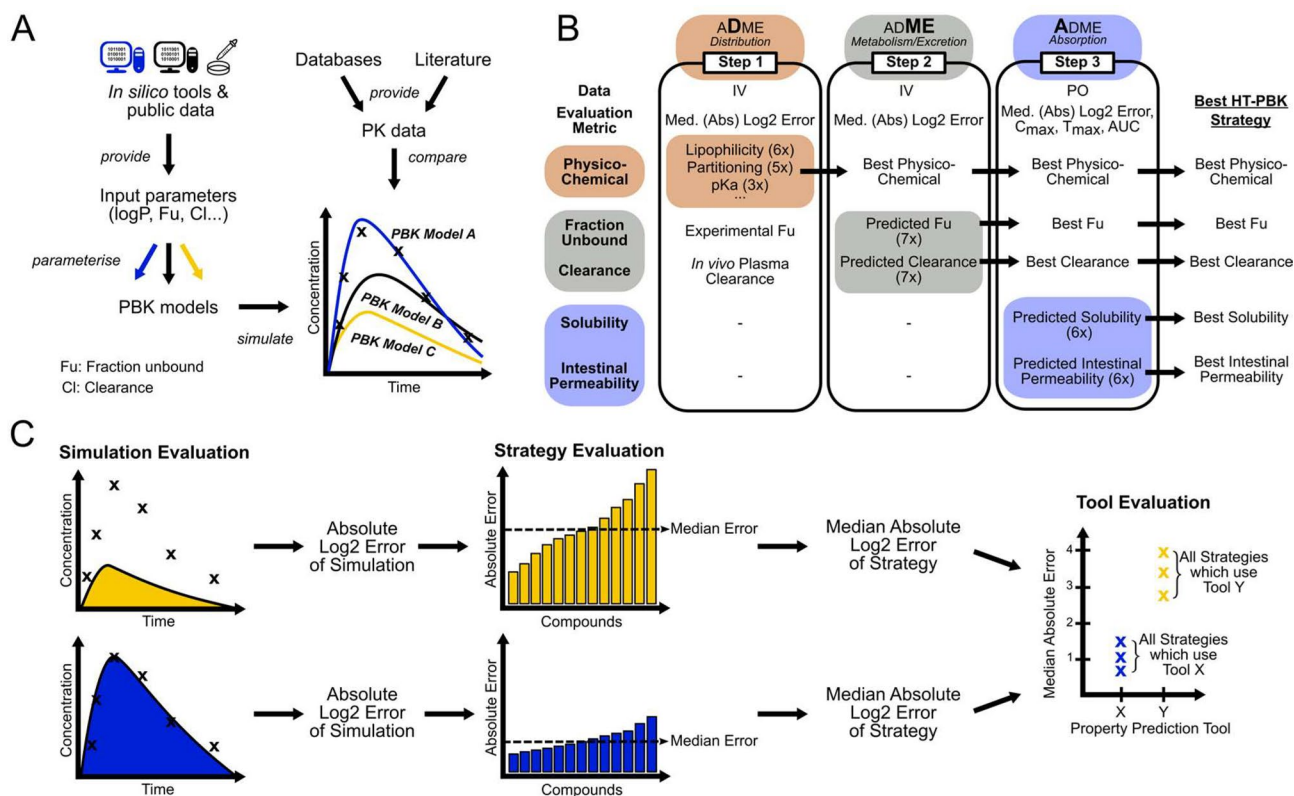
Fig. 2 Conceptual overview of the simulation and analysis strategy. **A** For each compound various PBK models were generated using all combinations of available input parameter sources. Their performance was then compared against gathered concentration–time profiles. **B** The simulation and evaluation of PBK models built from the various parameterisation sources was split into three steps, each focusing on the evaluation of a separate level of ADME processes. **C** Model performance was first evaluated at the level of individual simulations as the difference between model simulation and observed data. Then, for every parameterisation strategy, the median error of all compounds was used as a measure of the performance of that parameterisation strategy. Finally, comparing all performances of all strategies using the same input property prediction tools allowed benchmarking of the various tools against each other

benchmark values for simulation, so that any remaining PBK model simulation inaccuracies would only be due to mispredictions of passive compound distribution alone. Then, we evaluated which physico-chemical prediction tools resulted in the best PBK model simulations by systematically testing all combinations of input parameter sources against our 718 collected concentration–time profiles after IV administration. For each simulation, we calculated Median Relative and Absolute Log2 Errors as measures of prediction bias (systematic error) and precision (random error), respectively.

Out of the tested parameters, we found the strongest factor determining PBK model accuracy was which lipophilicity values were used for PBK model building (Fig. 3A). Tools that predicted LogP performed overall worse than those predicting LogD or LogMA lipophilicity values. Likewise, the performances of tools predicting the same type of lipophilicity also differed. For example, LogD values predicted by the Bayer tool worked better for PBK model parameterisation than the ones coming from ADMETLab or SimPlus (ADMETPredictor). The higher errors of the

LogP tool based predictions were correlated with a general bias for underprediction of the PK data. When investigating this effect on the individual compound level (Fig. 3E–J), it became apparent that for some compounds the LogP tools predicted very high lipophilicity values (> 5), which then led to a severe underprediction of those compounds' plasma concentrations, while the same compounds' PK was predicted reasonably well when using LogD or LogMA values for PBK model parameterisation.

The results of the partitioning methods were less straightforward. We observed a stable hierarchy in the predictive performances of the different methods, with the Berezhkovskiy method performing best and the Schmitt method performing worst under most circumstances (Fig. 3A). However, this difference in performance was only observed clearly when using lipophilicity values from the less well-performing LogP prediction tools, whereas when using the better lipophilicity values (LogD and LogMA Bayer) the difference in prediction precision between the different partitioning methods was only marginal (Fig. 3C). We
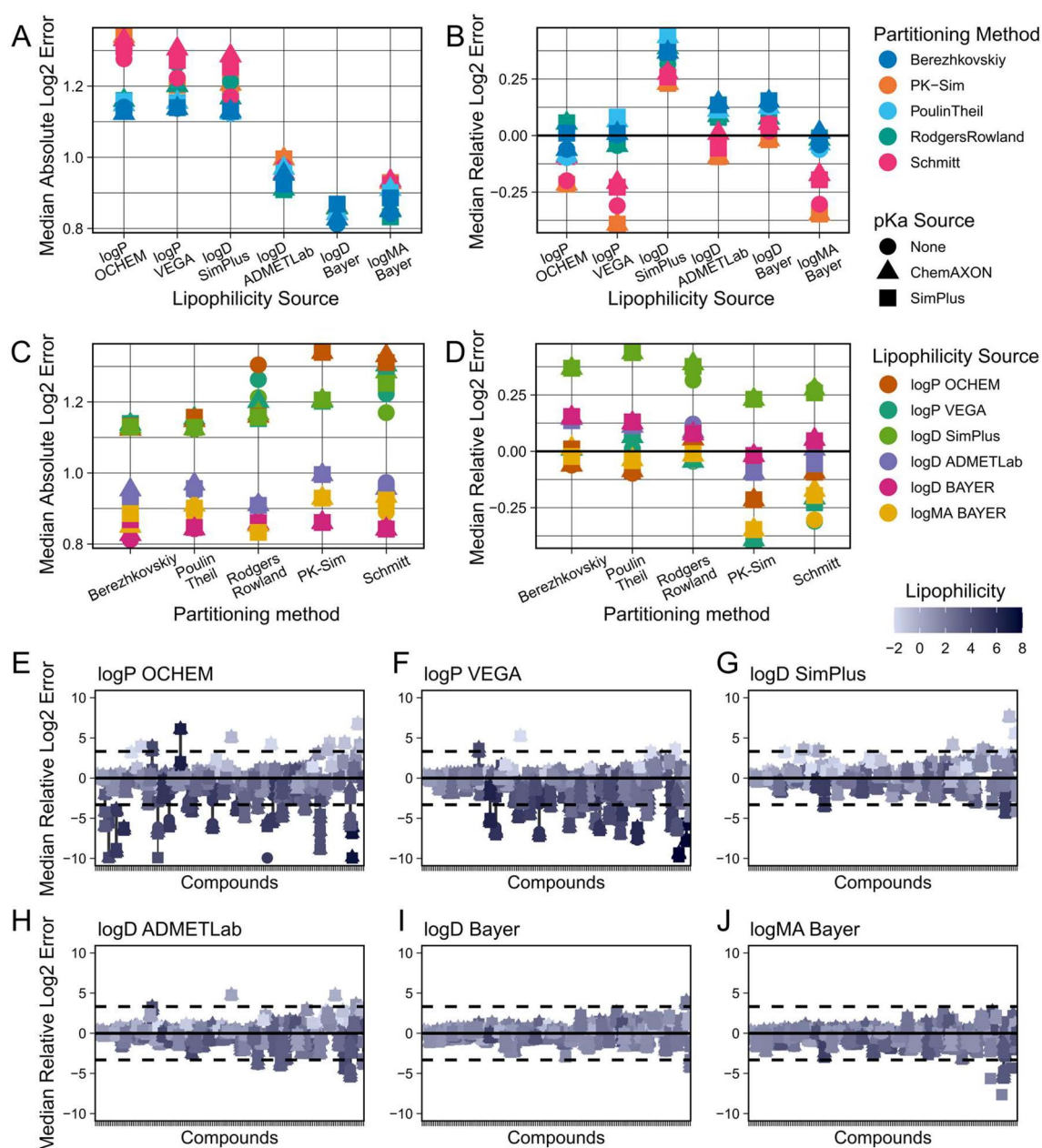
Fig. 3 Comparison of predictive performances of different physico-chemical parameter sources (step 1). Combinations of all available PBK model parameterisation sources were evaluated against the collected IV dataset. Clearance and fraction unbound were parameterised using in vivo and in vitro benchmark reference values, respectively. The top row shows Median Absolute Log2 Errors (**A**) and Median Relative Log2 Errors (**B**) for different lipophilicity prediction methods. The middle row shows Median Absolute Log2 Errors (**C**) and Median Relative Log2 Errors (**D**) for different partitioning methods. The bottom rows (**E–J**) show the Relative Log2 Errors of predictions for every individual compound for the different lipophilicity prediction tools used. Dashed lines indicate tenfold errors

further investigated this at the individual compound level (SI-Fig. 1), and we observed that the methods of Poulin & Theil and of Berezhkovskiy were not generally more predictive for the majority of compounds. Rather, we found that Poulin & Theil and Berezhkovskiy were less negatively impacted by the very high lipophilicity values predicted by the LogP tools for some compounds (SI-Fig. 2). It was only

their robustness to these high lipophilicity outliers, which decreased the observed performance of the other partitioning methods but not theirs, that made them appear to be superior overall (SI-Fig. 3).

For the provision of predicted pKa values, there was no strong trend observable, even though one may have expected that partitioning methods that use pKa values

as input should perform better when those are provided. However, this was only consistently the case for Rodgers & Rowland partitioning, and only when it was used with LogP lipophilicity values (SI-Fig. 4). The other methods using pKa values, namely the methods of Schmitt, Poulin & Theil and Berezhkovskiy, performed sometimes better, sometimes worse, depending on which lipophilicity prediction tool was being used for simulation.

Finally, we evaluated two approaches to further improve lipophilicity predictions for PBK modelling. The first approach was that of using consensus values of the different prediction tools, e.g., for LogP, simply by taking the mean of the values predicted by the different tools. We did this with both the tools for LogP, as well as for LogD, respectively, and, interestingly, observed opposite effects (SI-Fig. 5). Averaging the LogP predictions indeed resulted in better predictivity of the PBK models than any of the individual LogP tools. For the LogD, however, averaging produced better results than the worst tool (SimPlus) but worse results than the better tools (Bayer, ADMETLab).

The second strategy we tested to improve lipophilicity predictions was to use regression equations that empirically relate LogP or LogD values to membrane affinity (LogMA). We obtained two equations for converting LogP values (Yun et al. 2014; Endo et al. 2011) and generated a comparable equation for LogD values based on the data presented in Loidl-Stahlhofen et al. (2001). But we found that only one of the three strategies, namely using the Yun et al. (2014) equation, consistently improved PK predictions, regardless of which LogP tool or partitioning method it was used with (SI-Fig. 6). Whereas the two other methods showed at best mixed results, or even worsened predictions in the case of our self-derived equation based on the Loidl-Stahlhofen et al. (2001) data. The improvement in PK predictions, achieved by converting LogP to LogMA values using the equation from Yun et al. (2014), suggests that the tested LogP prediction tools may not be inherently less accurate than the LogD tools. Rather, they may just provide a type of lipophilicity value that is less suitable for the PBK modelling of certain compound classes.

Given these results, we concluded that there was no obviously superior partitioning method, nor that providing the pKa values was consistently providing better predictions. However, the lipophilicity values provided by the Bayer tools (LogD and LogMA) did appear to give superior predictions compared to the other lipophilicity prediction sources. For this reason, we proceeded with the mean of those tools as the best lipophilicity prediction, as well as all partitioning and pKa prediction methods into the next round for the evaluation of clearance and Fu prediction tools (step 2).

Step 2: evaluation of clearance and fraction unbound predictions

After evaluating the physico-chemical properties determining passive distribution, we continued with the tools predicting key parameters depending on organism biology, specifically the Fu and clearance of compounds. We used the same IV PK data for validation as in the first step, but only using the best physico-chemical predictions as determined previously, while this time varying the Fu and clearance predictions.

For the prediction of the Fu, we had obtained values from seven in silico tools, as well as in vitro measured benchmark values. As expected, we found that the importance of Fu predictions depended on the clearance prediction approach used. When using in vivo plasma clearance benchmark values for model parameterisation, only marginal differences between the performances of the different Fu prediction tools were observed (SI-Fig. 7). But when predicting in vivo clearance using in vitro measured hepatic CLint values, we observed larger differences between the different Fu prediction tools (Fig. 4A, B). Our experimentally determined Fu values yielded better PK predictions than any in silico tool, which confirmed the validity of our benchmark reference values. However, the differences between the prediction qualities were overall relatively small. All Fu prediction tools led to Median Absolute Errors within the two- to threefold range when using in vitro CLint values and there was no obvious systematic bias for under- or overprediction for any of the Fu prediction tools.

For the prediction of compound clearance, three in silico tools directly predicting plasma clearance, as well as our own previously used in vivo plasma clearance benchmark values, were available. Further, we had retrieved in vitro measured hepatic intrinsic clearance (CLint) values from httk, as well as in silico predictions of CLint values from OPERA, SimPlus and ADMETAI.

Before comparing the performances of the different clearance prediction strategies, we first evaluated whether activating passive renal excretion would improve or worsen PBK model simulations. Plasma clearance values already represent the total effect of all systemic clearance processes, so that adding passive renal clearance on top of them should theoretically lead to less accurate results. Whereas PBK models are expected to yield better PK predictions when passive renal excretion is incorporated if their in vivo clearance prediction is scaled up from hepatocyte-derived CLint values.

Overall, our results were consistent with these expectations. When adding passive renal clearance on top of the in vivo observed plasma clearance, prediction quality became worse and shifted from no bias to underprediction, whereas in vitro hepatocyte-based clearances improved
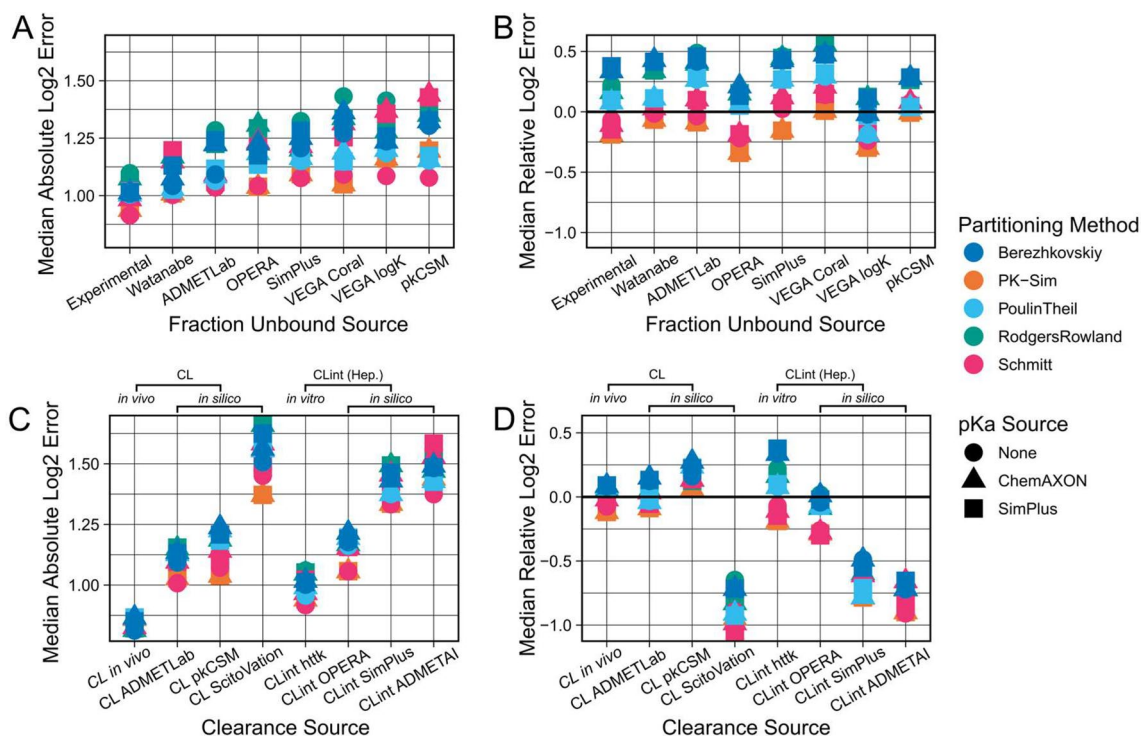
Fig. 4 Comparison of predictive performances of different fraction unbound and clearance prediction sources (step 2). Combinations of all available parameterisation sources were evaluated against the collected IV dataset. Results shown were generated using benchmark values for parameterisation of the other parameters, i.e., in vitro CLint values from httk when comparing Fu predicting sources, and in vitro measured Fu values when comparing clearance predicting sources, as well as the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. The top row shows the Median Absolute Log2 Errors (**A**) and the Median Relative Log2 Errors (**B**) for different fraction unbound prediction methods. The bottom row shows the Median Absolute Log2 Errors (**C**) and the Median Relative Log2 Errors (**D**) for different clearance prediction methods. CL refers to plasma clearance values (measured or predicted), and CLint signifies hepatocyte intrinsic clearance (measured or predicted). In vitro measured CLint values from httk were only available for a subset of compounds (97 out of 143)

from stronger to weaker overprediction of the PK data (SI‑Fig. 8). For in silico predicted clearances, the situation was less straightforward. For instance, in silico CLint values predicted by SimPlus and ADMETAI already led to underpredictions of PK, which was then further exacerbated by additionally adding passive renal clearance. However, given the theoretical considerations, we continued our simulations by adding passive renal clearance on top of hepatocyte‑scaled CLint, but not plasma clearance values.

Similar to the Fu, the in vivo observed plasma clearance benchmark values were the best input source for PBK model parameterisation (Fig. 4C). All clearance prediction strategies yielded profoundly worse results than the benchmark in vivo clearance‑based strategy, and almost all of them gave Median Absolute Errors worse than the twofold range. However, the differences between the clearance prediction tools were much larger than those between the Fu prediction tools. Out of the in silico predicted plasma clearance tools we found that ADMETLab gave the best predictions, followed by pkCSM and ScitoVation. While ADMETLab and pkCSM plasma clearance predictions led to a slight overprediction

of the PK data, ScitoVation plasma clearance values led to a severe underprediction. We further confirmed these findings by directly comparing our in vivo plasma clearance values to the in silico predicted values of the different tools (SI‑Fig. 9), which showed that ADMETLab's plasma clearance values correlated best with our in vivo measured benchmark values.

In vitro hepatocyte CLint values from httk were the second‑best clearance prediction source, after our in vivo plasma clearance benchmark values. Similar to what was observed for the in silico tools predicting plasma clearance values, the values from in silico CLint prediction tools also resulted in substantially worse PK predictions than the in vitro benchmark values. The best‑performing CLint prediction tool was OPERA, followed by SimPlus and ADMETAI. However, we noted that for some compounds OPERA provided CLint values identical to the in vitro CLint values retrieved from httk (SI‑Fig. 10). This suggested that those values were not true in silico predictions, which implies that the OPERA predictions may not be directly comparable to the other tools. Overall, when comparing in silico tools

predicting plasma clearance values to the tools predicting hepatocyte CLint, most plasma clearance tools resulted in better PK predictions than most hepatocyte CLint prediction tools.

Step 3: evaluation of solubility and intestinal permeability

In the third evaluation step, we investigated how to best predict parameters required for simulating oral administrations. In PK-Sim, these are primarily the solubility and intestinal permeability of a compound. We had obtained experimentally measured benchmark values of aqueous solubility, as well as predictions of aqueous solubility from four in silico tools (OPERA, ADMETLab, ProtoQSAR, SimPlus), and of Fasted State Simulated Intestinal Fluid (FaSSIF) and Fed State Simulated Intestinal Fluid (FeSSIF) solubility from SimPlus. For the intestinal permeability, no benchmark reference values were obtained. Instead, CACO2 permeability predictions from three in silico tools were used (OPERA, ADMETLab, ProtoQSAR), as well as MDCK permeability

predictions from ADMETLab and SimPlus. Finally, we also obtained intestinal permeability predictions using the PK-Sim internal prediction equation, which is based on compounds' molecular weight and lipophilicity (LogMA Bayer).

For the evaluation, we initially only used data from PK studies in which the administered formulation implied that the compound was already dissolved at administration (e.g., labelled as "solution" or "suspension") and not in a solid state (e.g., "tablet" or "capsule"), since this additionally requires knowledge about the dissolution times of these formulations. We tested the mentioned prediction tools against the 286 concentration–time profiles (94 compounds) from those liquid formulation studies (Fig. 5). However, no substantial difference was observed between the different parameterisation sources for either property. In the case of solubility, all in silico tools gave results similar to each other, and also comparable to the results of our experimentally measured benchmark values. Likewise, all intestinal permeability prediction tools gave comparable results.

Further, we observed a general trend for overprediction of the velocity of oral absorption which resulted in a
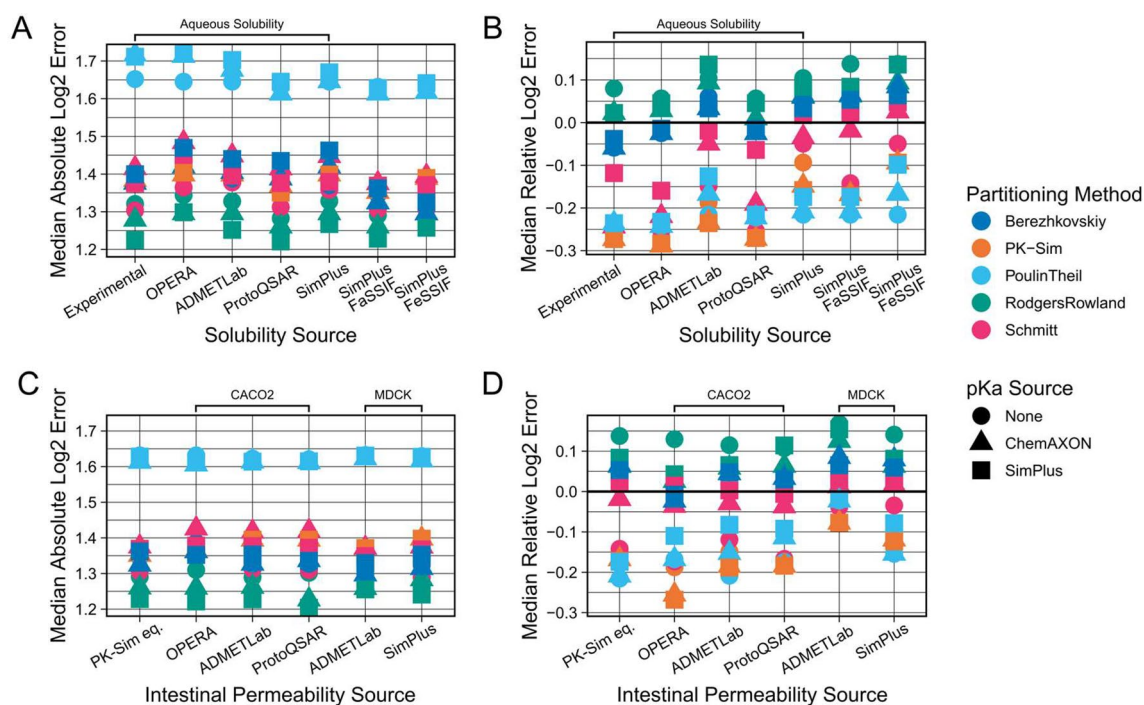

Fig. 5 Comparison of predictive performances of different solubility and intestinal permeability prediction sources (step 3). Combinations of all available parameterisation sources were evaluated against the collected PO dataset (dissolved formulations). Results shown were generated using benchmark values for parameterisation of other parameters, i.e., in vivo plasma clearances, in vitro fraction unbound values and the mean of the two previously determined best lipophilicity prediction tools (LogD and LogMA Bayer) as lipophilicity values. The top row shows the Median Absolute Log2 Errors (**A**) and the Median Relative Log2 Errors (**B**) for different solubility prediction

methods. The bottom row shows the Median Absolute Log2 Errors (**C**) and the Median Relative Log2 Errors (**D**) for different intestinal permeability prediction methods. For comparison of solubility predictions, the intestinal permeability values used were the PK-Sim internal equation values (PK-Sim eq.). For comparison of intestinal permeability sources, the solubility values used were SimPlus FaSSIF values. Results for the PK-Sim internal equation were generated with the Bayer LogMA predictions. Intestinal permeability predictions are either CACO2 or MDCK permeability predictions

consistently strong underprediction of Tmax values, and a slight tendency for overprediction of Cmax values (SI-Fig. 11). We hypothesised that this might be because the in silico tools do not predict the PK-Sim specific intestinal permeability directly but instead were trained to predict in vitro measured CACO2 or MDCK permeabilities. However, when using such in vitro measured permeabilities, the standard procedure would be to scale these values, for example, using reference compounds, to the PK-Sim intestinal permeability parameter. Only when no measurements for reference compounds exist would one use the in vitro measured permeability values directly without scaling.

To take this into account, we extracted fitted PK-Sim intestinal permeability values of 56 compounds from Willmann et al. (2004) and then, for every in silico tool, determined a scaling factor based on the relationship between the values predicted by every tool and the assumed to be optimal values. While we found that there was a clear trend for the CACO2 values to be larger than the optimal PK-Sim intestinal permeability values (SI-Fig. 12), incorporating this scaling did not substantially improve PK predictions overall (SI-Fig. 13). Even though, it did reduce the strength of the bias in the underprediction of Tmax values.

Finally, we evaluated whether our conclusions based on the 94 compounds from liquid formulation studies would also hold true for the compounds of which we only had data from solid formulation studies. The simulation of these formulations required at least one additional parameter to describe the dissolution velocity of the solid formulations, which in reality will vary between different formulations. To at least determine which values might be appropriate average values, we tested different Lint80 dissolution times (10–30 min for capsules, 15–60 min for tablets) and then compared which average dissolution time would yield errors similar to what we had observed for the liquid formulations. Based on this we decided to use 25 min for formulations labelled "capsules" and 40 min for "tablets", which extended the oral dataset for evaluation to 1200 PO concentration–time profiles (161 compounds). Using this larger dataset, all previously outlined conclusions were confirmed.

Predictive performances of full HT-PBK strategies

After evaluating step-by-step how to best predict every compound property required for HT-PBK modelling, we eventually assessed how well different types of HT-PBK strategies would predict the collected PK data overall. We identified the best strategies out of three classes. (1) As a benchmark comparison, we determined the performance of the best strategy overall, using in vivo and in vitro determined benchmark values of plasma clearance and Fu. (2) Additionally, the best fully in silico-based strategy was identified, for which we also considered property predictions coming from proprietary tools. (3) And finally, the best in silico strategy based exclusively on freely available tools was determined. The respective parameterisation strategies and their performances are presented in Table 1. Unsurprisingly, we found the strategy using benchmark reference values to be the most predictive. However, even fully in silico-based strategies yielded acceptable predictivity with 87%, or 89%, of Cmax values being predicted within tenfold when using proprietary, or freely available prediction tools, respectively. Even more importantly, due to overestimation of the velocity of oral absorption in all strategies, the Cmax mispredictions outside the tenfold range were mostly over- not underpredictions and therefore would lead to conservative, health-protective risk assessment conclusions. The performance of the best in silico-based HT-PBK approach is presented in Fig. 6 and that of the other strategies is shown in SI-Fig. 14.

Discussion

We here assembled a large dataset of healthy human in vivo concentration–time profiles (200+ compounds), as well as in vitro and in silico generated property predictions from various sources required for PBK modelling of the corresponding compounds. We systematically compared all possible HT-PBK modelling strategies to understand which prediction tools, and combinations thereof, perform best for parameterising PK-Sim to predict concentration–time data. Thereby, we quantified the expected accuracy of such HT-PBK predictions for typical pharmaceutical compounds.

For some input properties, especially lipophilicity and clearance, we found major differences in PBK model performances, while for other properties there was little variation. This may be due to larger differences in the predictive performances of the respective tools or due to differences in the sensitivity of the PBK model towards the different input parameters. Based on this observation, we conclude that generating and comparing PBK model simulations with different prediction tools for these critical parameters might be required to achieve more robust PBK model predictions. In contrast, for less sensitive parameters that may be of less importance. This could either be achieved by averaging property predictions of different tools, or by simulating ensembles of different PBK model variants. The latter would further allow to generate a distribution of simulation outcomes, thereby providing a confidence interval around simulation predictions. Such an explicit representation of uncertainty could then, for example, also be useful for conducting probabilistic risk assessment.

We also performed a large-scale comparison of the performances of frequently for PBK modelling used partitioning methods and found that all methods implemented in PK-Sim performed similarly well when based on high-quality

Table 1 Overview of different HT-PBK modelling strategies and their predictive performances

|		Best overall (including in vivo and in vitro benchmark values)	Best in silico-proprietary	Best in silico-free
Input parameter sources	Lipophilicity	Mean of Bayer LogD and LogMA	Mean of Bayer LogD and LogMA	ADMETLab 2.0: LogD
	Solubility	SimPlus: FaSSIF	SimPlus: FaSSIF	ADMETLab 2.0: Aqueous solubility
	pKa	ChemAxon	ChemAxon	None
	Partitioning method	PK-Sim	PK-Sim	PK-Sim
	Fraction unbound	In vitro measured	ADMETLab 2.0	ADMETLab 2.0
	Clearance	In vivo observed plasma clearance	ADMETLab 2.0: Plasma clearance	ADMETLab 2.0: Plasma clearance
	Intestinal permeability	ADMETLab 2.0: MDCK	ADMETLab 2.0: MDCK	ADMETLab 2.0: MDCK
Performance metrics	IV: median absolute Log2 error	0.860	1.065	1.273
	PO: median absolute Log2 error	1.245	1.595	1.780
	PO: median absolute Log2 Cmax predicted/observed	0.998	1.127	1.130
	PO: compounds with Cmax within ten-/five-/three-/twofold	86%/78%/69%/50%	87%/78%/65%/44%	89%/71%/62%/42%
	PO: underprediction of compound Cmax of more than ten-/five-/three-/twofold	1.9%/4.3%/8.1%/16%	0%/3.1%/8.1%/19%	3.1%/12%/16%/25%
	PO: median absolute Log2 AUC predicted/observed	0.870	1.285	1.358
	PO: compounds with AUC within ten-/five-/three-/twofold	92%/84%/75%/57%	84%/71%/58%/38%	84%/66%/54%/39%
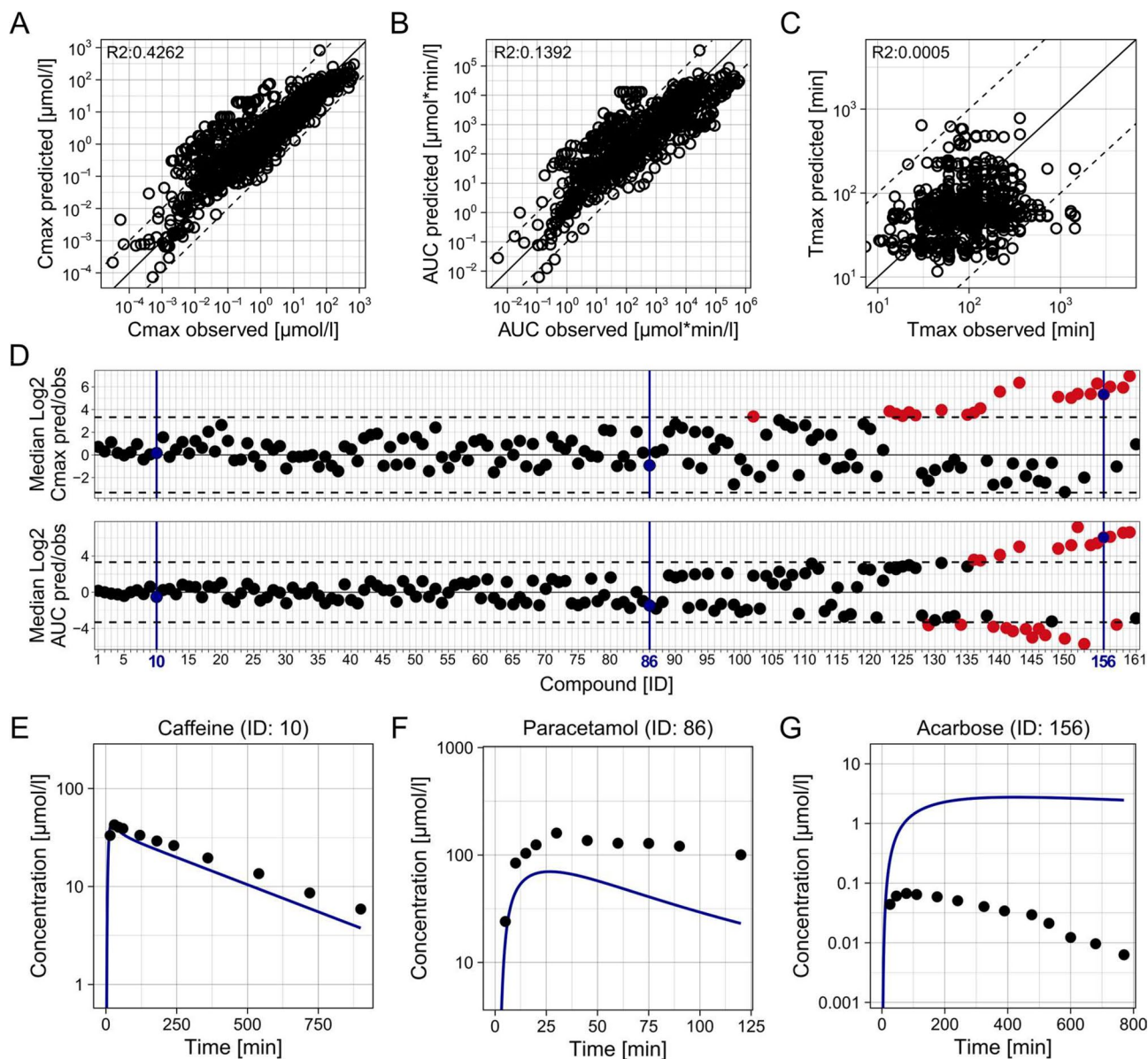
Fig. 6 Predictive performance of the best fully in silico-based HT-PBK modelling strategy (proprietary). HT-PBK models were generated following the approach outlined in Table 1, then model simulations were compared against the collected PO dataset of 161 compounds (dissolved and solid formulations). **A** Predicted against observed Cmax values of all concentration–time profiles in the PO dataset. **B** Predicted against observed AUC values. **C** Predicted against observed Tmax values. Mind that **A–C** show individual con- centration–time profiles with some compounds only represented by a single and others by multiple PK profiles. **D** The median Log2 Cmax and AUC predicted/observed values of each compound in the PO dataset. Dashed lines indicate tenfold errors. **E–G** Representative concentration–time profile predictions, and corresponding in vivo PK data, of three compounds embodying different levels of prediction quality. Blue colour marks the same representative compounds in D (colour figure online)

lipophilicity values. However, for very high lipophilicity values, Poulin & Theil and Berezhkovskiy resulted in better PK predictions than other methods. It is likely though that more subtle differences in the performances of the different partitioning methods were not detectable with our approach, due to the intrinsic variability of our heterogeneous PK dataset gathered from several databases and the literature. Of note, even different PK profiles of the same compound but from different studies were not exactly identical to each other, which implies that our dataset contains intrinsic variability that even an ideal generic PBK model could never capture entirely. Consequently, this may have been the reason why we were unable to detect more subtle performance differences between the partitioning methods when using the highest quality lipophilicity values.

Besides their absolute errors, we also investigated whether any PK prediction strategies had biases for over- or underprediction. Our most noteworthy observation, apart from some tools having their individual biases, was that all HT-PBK modelling strategies seemed to overestimate the velocity of oral absorption, leading to an overprediction of Cmax and an underprediction of Tmax values. It is not fully clear why this occurred, as many factors influence oral absorption in vivo. Some of our fundamental assumptions, such as that "suspension" formulations were fully dissolved at administration, may have been inaccurate. Further, we parameterised our PBK models with a simple passive intestinal permeability parameter, thereby ignoring gut efflux transporters which are known to sometimes have profound impacts on the intestinal absorption of compounds. Ignoring such effects may have led to more rapid and full oral absorption in our simulations than occurred in vivo.

CACO2 or MDCK permeability values should, theoretically, account for such transporter effects (Volpe 2011). However, to be used in PK-Sim, they require scaling. Interestingly, our attempt to perform such a scaling using previously published fitted intestinal permeability values (Willmann et al. 2004) did indeed correct the biases in the underprediction of Tmax values, but the absolute PK prediction's errors did not improve. One possible explanation for this result is that underpredictions of the velocity of oral absorption may have had a more severe impact on PK predictions than its overprediction. And for this reason, we observed lower Absolute Log2 Errors when using CACO2 values directly, despite scaled CACO2 values being potentially more correct representations of in vivo oral permeability. For applications in the pharmaceutical domain, it may therefore be of interest to further investigate such CACO2 scaling approaches. However, for toxicological applications, an overprediction of the velocity of oral absorption may even be preferable since it leads to a health-protective bias in risk assessments.

Similar to gut transporters, we were also unable to explicitly account for the contribution of transporter effects in other body tissues that may potentially alter compounds' PK. Performing high-throughput assessments requires a generic modelling approach and readily available, homogeneous input data from the same sources to be able to systematically compare the predictive performances of strategies. Unfortunately, such data were not available for transporters. Further, our approach solely focused on parent compounds and neglected the issues of bioactivation and metabolism, since the formation of metabolites cannot be predicted quantitatively. Also, it would be valuable to gain more insight into which compound properties (physico-chemical properties, clearance pathways, transporter affinity, etc.) are correlated with lower or higher prediction accuracies, since we do observe large differences between prediction accuracies among different compounds.

Because most compounds in our dataset were typical pharmaceutical compounds, it is further likely that many of them were present in the training data of the in silico tools we evaluated, potentially biasing our analysis. This was most evident for the CLint values gathered from OPERA, which, for some compounds, were identical to the in vitro CLint values from httk. It might be insightful to analyse in more detail how prediction accuracies of in silico tools differ between those compounds on which the models were trained originally and those compounds that were outside their training datasets. Nevertheless, from a practical point of view, the fact that some in silico tools may be able to recall data from larger or higher quality training datasets may also be interpreted as a strength of these tools.

Eventually, we showed that it is possible to generate reliable HT-PBK models for the prediction of IV and PO PK of pharmaceuticals. In principle, it is also possible to apply our approach to other classes of compounds, although the validation of this is hampered by the absence of comparable concentration–time data for validation. Potential applications of such HT-PBK modelling strategies are vast, and there have been many efforts recently in both pharmacology and toxicology to establish such strategies for different use cases and based on different approaches. Many of them, however, relied exclusively on rodent data for their validation (Schneckener et al. 2019; Kamiya et al. 2021; Naga et al. 2022; Punt et al. 2022b; Obrezanova et al. 2022; Mavroudis et al. 2023; Handa et al. 2023; Führer et al. 2024). Others did perform predictions for humans but relied in their validation of prediction quality on summary PK parameters, like Cmax or AUC values (Punt et al. 2022a; Miljković et al. 2021; Li et al. 2023). The problem with such approaches is that they do not consider the full information about the quality of the predicted concentration–time curve as a whole. Hence, they are unable to deconvolute the biases of individual input sources, that can potentially compensate each other, and which may obscure inaccuracies in model parameterisation. This is why we here relied on full concentration–time profiles, and multiple summary parameters, for evaluation, as it would be done by an expert for the traditional development of a PBK model.

Another now frequently used strategy is to use ML and AI techniques to directly predict summary PK parameters, like the Cmax, AUC or bioavailability of compounds (Schneckener et al. 2019; Miljković et al. 2021; Fagerholm et al. 2021; Obrezanova et al. 2022). However, so far, this approach did not perform as well as using in silico predicted properties to then inform mechanistic simulations based on PBK models. For this reason, we here followed the latter strategy of using ML models to predict mechanistically meaningful compound properties to then input

these into the PK-Sim PBK model, which incorporates ab initio expert knowledge about known body physiology into our high-throughput predictions. Besides yielding more accurate predictions of PK parameters, this approach maintains the explainability of our models and their predictions. Since every PBK model parameter has a physiological meaning, the approach further enables investigation into why certain compounds might be mispredicted by different strategies and to understand which property mispredictions are responsible for inaccurate PK predictions. Furthermore, this approach makes it possible to leverage PBK models' ability to mechanistically extrapolate predictions, for example, to special populations or other exposure scenarios.

Finally, the here presented HT-PBK modelling is a promising tool for applications in both pharmacology and toxicology. In drug discovery, HT-PBK models have the potential to aid rapid compound selection and optimisation decisions. While HT-PBK modelling may not initially match the accuracy of traditional PBK models, it can provide a base model which may then be progressively refined throughout the drug development cycle to meet escalating demands for accuracy and detail. For toxicological risk assessment, accurate predictions of internal organ concentrations are key to replace animal testing with in vitro and in silico methods. Such predictions can, for example, assist with the prioritisation and classification of chemicals, or provide valuable information for quantitative in vitro to in vivo extrapolation. For a new methodology like HT-PBK modelling to be used in Next Generation Risk Assessment, however, it is key to validate its predictive performance, to generate the high confidence required for its regulatory use. Using a large, heterogeneous PK dataset, we here showed that the outlined HT-PBK modelling strategies are fit-for-purpose for such applications.

Declarations

References

Bender A, Cortés-Ciriano I (2021) Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. Drug Discov Today 26(2):511–524. https://doi.org/10.1016/j.drudis.2020.12.009

Benfenati E, Manganaro A, Gini G (2013) VEGA-QSAR: AI inside a platform for predictive toxicology. Popularize Artificial Intelligence 2013: Proceedings of the Workshop on Popularize Artificial Intelligence (PAI 2013)

Berezhkovskiy LM (2004) Volume of distribution at steady state for a linear pharmacokinetic system with peripheral elimination. J Pharm Sci 93(6):1628–1640. https://doi.org/10.1002/jps.20073

Blaauboer BJ (2010) Biokinetic modeling and in vitro-in vivo extrapolations. J Toxicol Environ Health Part B Crit Rev 13(2–4):242–252. https://doi.org/10.1080/10937404.2010.483940

Bouvier d'Yvoire M, Prieto P, Blaauboer BJ, Bois FY, Boobis A, Brochot C, Coecke S, Freidig A, Gundert-Remy U, Hartung T, Jacobs MN, Lavé T, Leahy DE, Lennernäs H, Loizou GD, Meek B, Pease C, Rowland M, Spendiff M, Yang J, Zeilmaker M (2007) Physiologically-based kinetic modelling (PBK modelling): meeting the 3Rs agenda. The report and recommendations of ECVAM Workshop 63. Altern Lab Anim: ATLA 35(6):661–671. https://doi.org/10.1177/026119290703500606

Breen M, Ring CL, Kreutz A, Goldsmith M-R, Wambaugh JF (2021) High-throughput PBTK models for in vitro to in vivo extrapolation. Expert Opin Drug Metab Toxicol 17(8):903–921. https://doi.org/10.1080/17425255.2021.1935867

Clewell HJ, Tan YM, Campbell JL, Andersen ME (2008) Quantitative interpretation of human biomonitoring data. Toxicol Appl Pharmacol 231(1):122–133. https://doi.org/10.1016/j.taap.2008.04.021

Danishuddin KV, Faheem M, Woo Lee K (2022) A decade of machine learning-based predictive models for human pharmacokinetics: advances and challenges. Drug Discov Today 27(2):529–537. https://doi.org/10.1016/j.drudis.2021.09.013

Denney WS, Duvvuri S, Buckeridge C (2015) Simple, Automatic noncompartmental analysis: the PKNCA R package. J Pharmacokinet Pharmacodyn 42(1):11–107. https://doi.org/10.1007/s10928-015-9432-2

Edginton AN, Theil F-P, Schmitt W, Willmann S (2008) Whole body physiologically-based pharmacokinetic models: their use in

clinical drug development. Expert Opin Drug Metab Toxicol 4(9):1143–1152. https://doi.org/10.1517/17425255.4.9.1143

Endo S, Escher BI, Goss K-U (2011) Capacities of membrane lipids to accumulate neutral organic chemicals. Environ Sci Technol 45(14):5912–5921. https://doi.org/10.1021/es200855w

Fagerholm U, Hellberg S, Spjuth O (2021) Advances in predictions of oral bioavailability of candidate drugs in man with new machine learning methodology. Molecules (basel, Switzerland). https://doi.org/10.3390/molecules26092572

Fagerholm U, Hellberg S, Alvarsson J, Spjuth O (2023) In silico prediction of human clinical pharmacokinetics with ANDROMEDA by prosilico: predictions for an established benchmarking data set, a modern small drug data set, and a comparison with laboratory methods. Altern Lab Anim: ATLA 51(1):39–54. https://doi.org/10.1177/02611929221148447

Führer F, Gruber A, Diedam H, Göller AH, Menz S, Schneckener S (2024) A deep neural network: mechanistic hybrid model to predict pharmacokinetics in rat. J Comput Aided Mol Des 38(1):7. https://doi.org/10.1007/s10822-023-00547-9

Ghose AK, Crippen GM (1987) Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. J Chem Inf Comput Sci 27(1):21–35. https://doi.org/10.1021/ci00053a005

Göller AH, Kuhnke L, Montanari F, Bonin A, Schneckener S, ter Laak A, Wichard J, Lobell M, Hillisch A (2020) Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. Drug Discov Today 25(9):1702–1709. https://doi.org/10.1016/j.drudis.2020.07.001

Gonzalez Hernandez F, Carter SJ, Iso-Sipilä J, Goldsmith P, Almousa AA, Gastine S, Lilaonitkul W, Kloprogge F, Standing JF (2021) An automated approach to identify scientific publications reporting pharmacokinetic parameters. Wellcome Open Res 6:88. https://doi.org/10.12688/wellcomeopenres.16718.1

Grzegorzewski J, Brandhorst J, Green K, Eleftheriadou D, Duport Y, Barthorscht F, Köller A, Ke DYJ, de Angelis S, König M (2021) PK-DB: pharmacokinetics database for individualized and stratified computational modeling. Nucleic Acids Res 49(D1):D1358–D1364. https://doi.org/10.1093/nar/gkaa990

Handa K, Wright P, Yoshimura S, Kageyama M, Iijima T, Bender A (2023) Prediction of compound plasma concentration-time profiles in mice using random forest. Mol Pharm 20(6):3060–3072. https://doi.org/10.1021/acs.molpharmaceut.3c00071

Hughes LD, Palmer DS, Nigsch F, Mitchell JBO (2008) Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and Log P. J Chem Inf Model 48(1):220–232. https://doi.org/10.1021/ci700307p

Ito K, Houston JB (2004) Comparison of the use of liver models for predicting drug clearance using in vitro kinetic data from hepatic microsomes and isolated hepatocytes. Pharm Res 21(5):785–792. https://doi.org/10.1023/B:PHAM.0000026429.12114.7d

Jones H, Rowland-Yeo K (2013) Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. CPT: Pharmacomet Syst Pharmacol 2(8):e63. https://doi.org/10.1038/psp.2013.41

Kamiya Y, Handa K, Miura T, Yanagi M, Shigeta K, Hina S, Shimizu M, Kitajima M, Shono F, Funatsu K, Yamazaki H (2021) In silico prediction of input parameters for simplified physiologically based pharmacokinetic models for estimating plasma, liver, and kidney exposures in rats after oral doses of 246 disparate chemicals. Chem Res Toxicol 34(2):507–513. https://doi.org/10.1021/acs.chemrestox.0c00336

Khalidi H, Onasanwo A, Islam B, Jo H, Fisher C, Aidley R, Gardner I, Bois FY (2022) SimRFlow: an R-based workflow for automated high-throughput PBPK simulation with the Simcyp® simulator. Front Pharmacol 13:929200. https://doi.org/10.3389/fphar.2022.929200

Krstevska A, Đuriš J, Ibrić S, Cvijić S (2022) In-depth analysis of physiologically based pharmacokinetic (PBPK) modeling utilization in different application fields using text mining tools. Pharmaceutics. https://doi.org/10.3390/pharmaceutics15010107

Kuepfer L, Niederalt C, Wendl T, Schlender J-F, Willmann S, Lippert J, Block M, Eissing T, Teutonico D (2016) Applied concepts in PBPK modeling: how to build a PBPK/PD model. CPT: Pharmacomet Syst Pharmacol 5(10):516–531. https://doi.org/10.1002/psp4.12134

Lee AC, Crippen GM (2009) Predicting pKa. J Chem Inf Model 49(9):2013–2033. https://doi.org/10.1021/ci900209w

Li Y, Wang Z, Li Y, Du J, Gao X, Li Y, Lai L (2023) A combination of machine learning and PBPK modeling approach for pharmacokinetics prediction of small molecules in humans. https://doi.org/10.1101/2023.07.17.549292

Lin W, Chen Y, Unadkat JD, Zhang X, Di Wu, Heimbach T (2022) Applications, challenges, and outlook for PBPK modeling and simulation: a regulatory. Ind Acad Perspect Pharm Res 39(8):1701–1731. https://doi.org/10.1007/s11095-022-03274-2

Lippert J, Burghaus R, Edginton A, Frechen S, Karlsson M, Kovar A, Lehr T, Milligan P, Nock V, Ramusovic S, Riggs M, Schaller S, Schlender J, Schmidt S, Sevestre M, Sjögren E, Solodenko J, Staab A, Teutonico D (2019) Open systems pharmacology community-an open access, open source, open science approach to modeling and simulation in pharmaceutical sciences. CPT: Pharmacomet Syst Pharmacol 8(12):878–882. https://doi.org/10.1002/psp4.12473

Loidl-Stahlhofen A, Eckert A, Hartmann T, Schöttner M (2001) Solid-supported lipid membranes as a tool for determination of membrane affinity: high-throughput screening of a physicochemical parameter. J Pharm Sci 90(5):599–606. https://doi.org/10.1002/1520-6017(200105)90:5%3c599:AID-JPS1016%3e3.0.CO;2-N

Lombardo F, Obach RS, Shalaeva MY, Gao F (2002) Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding Data. J Med Chem 45(13):2867–2876. https://doi.org/10.1021/jm0200409

Mansouri K, Grulke CM, Judson RS, Williams AJ (2018) OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminform 10(1):10. https://doi.org/10.1186/s13321-018-0263-1

Mavroudis PD, Teutonico D, Abos A, Pillai N (2023) Application of machine learning in combination with mechanistic modeling to predict plasma exposure of small molecules. Front Syst Biol. https://doi.org/10.3389/fsysb.2023.1180948

Miljković F, Martinsson A, Obrezanova O, Williamson B, Johnson M, Sykes A, Bender A, Greene N (2021) Machine learning models for human in vivo pharmacokinetic parameters with in-house validation. Mol Pharm 18(12):4520–4530. https://doi.org/10.1021/acs.molpharmaceut.1c00718

Naga D, Parrott N, Ecker GF, Olivares-Morales A (2022) Evaluation of the success of high-throughput physiologically based pharmacokinetic (HT-PBPK) modeling predictions to inform early drug discovery. Mol Pharm 19(7):2203–2216. https://doi.org/10.1021/acs.molpharmaceut.2c00040

Obrezanova O, Martinsson A, Whitehead T, Mahmoud S, Bender A, Miljković F, Grabowski P, Irwin B, Oprisiu I, Conduit G, Segall M, Smith GF, Williamson B, Winiwarter S, Greene N (2022) Prediction of in vivo pharmacokinetic parameters and time-exposure curves in rats using machine learning from the chemical structure. Mol Pharm 19(5):1488–1504. https://doi.org/10.1021/acs.molpharmaceut.2c00027

Paini A, Leonard JA, Joossens E, Bessems JGM, Desalegn A, Dorne JL, Gosling JP, Heringa MB, Klaric M, Kliment T, Kramer NI, Loizou G, Louisse J, Lumen A, Madden JC, Patterson EA, Proença S, Punt A, Setzer RW, Suciu N, Troutman J, Yoon M, Worth A, Tan YM (2019) Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making. Comput Toxicol 9:61–72. https://doi.org/10.1016/j.comtox.2018.11.002

Paini A, Tan Y-M, Sachana M, Worth A (2021) Gaining acceptance in next generation PBK modelling approaches for regulatory assessments - An OECD international effort. Comput Toxicol (amsterdam, Netherlands) 18:100163. https://doi.org/10.1016/j.comtox.2021.100163

Pearce RG, Setzer RW, Strope CL, Wambaugh JF, Sipes NS (2017) httk: R package for high-throughput toxicokinetics. J Stat Softw 79(4):1–26. https://doi.org/10.18637/jss.v079.i04

Pillai N, Dasgupta A, Sudsakorn S, Fretland J, Mavroudis PD (2022) Machine learning guided early drug discovery of small molecules. Drug Discov Today 27(8):2209–2215. https://doi.org/10.1016/j.drudis.2022.03.017

Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem 58(9):4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104

Poulin P, Theil F-P (2002) Prediction of pharmacokinetics prior to in vivo studies. II. Generic physiologically based pharmacokinetic models of drug disposition. J Pharma Sci 91(5):1358–1370. https://doi.org/10.1002/jps.10128

Punt A, Louisse J, Beekmann K, Pinckaers N, Fabian E, van Ravenzwaay B, Carmichael PL, Sorrell I, Moxon TE (2022a) Predictive performance of next generation human physiologically based kinetic (PBK) models based on in vitro and in silico input data. Altex 39(2):221–234. https://doi.org/10.14573/altex.2108301

Punt A, Louisse J, Pinckaers N, Fabian E, van Ravenzwaay B (2022b) Predictive performance of next generation physiologically based kinetic (PBK) model predictions in rats based on in vitro and in silico input data. Toxicol Sci: off J Soc Toxicol 186(1):18–28. https://doi.org/10.1093/toxsci/kfab150

R Core Team (2022) R: a language and environment for statistical computing. https://www.R-project.org/.

Riley RJ, McGinnity DF, Austin RP (2005) A unified model for predicting human hepatic, metabolic clearance from in vitro intrinsic clearance data in hepatocytes and microsomes. Drug Metab Dispos: Biol Fate Chem 33(9):1304–1311. https://doi.org/10.1124/dmd.105.004259

Rodgers T, Rowland M (2006) Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. J Pharm Sci 95(6):1238–1257. https://doi.org/10.1002/jps.20502

Rohatgi A (2022) Webplotdigitizer: Version 4.6. https://automeris.io/WebPlotDigitizer.

Sayre RR, Wambaugh JF, Grulke CM (2020) Database of pharmacokinetic time-series data and parameters for 144 environmental chemicals. Sci Data 7(1):122. https://doi.org/10.1038/s41597-020-0455-1

Schmitt W (2008) General approach for the calculation of tissue to plasma partition coefficients. Toxicol Vitro: Int J Publ Assoc BIBRA 22(2):457–467. https://doi.org/10.1016/j.tiv.2007.09.010

Schneckener S, Grimbs S, Hey J, Menz S, Osmers M, Schaper S, Hillisch A, Göller AH (2019) Prediction of oral bioavailability in rats: transferring insights from in vitro correlations to (deep) machine learning models using in silico model outputs and chemical structure parameters. J Chem Inf Model 59(11):4893–4905. https://doi.org/10.1021/acs.jcim.9b00460

Snyder WS, Cook M, Nasset E, Karhausen L, Howells G, Tipton I (1979) Report of the task group on reference man. Ann ICRP 3(1–4):iii. https://doi.org/10.1016/0146-6453(79)90123-4

Sohlenius-Sternbeck A-K, Afzelius L, Prusis P, Neelissen J, Hoogstraate J, Johansson J, Floby E, Bengtsson A, Gissberg O, Sternbeck J, Petersson C (2010) Evaluation of the human prediction of clearance from hepatocyte and microsome intrinsic clearance for 52 drug compounds. Xenobiotica Fate Foreign Compd Biol Syst 40(9):637–649. https://doi.org/10.3109/00498254.2010.500407

Stokes WS (2015) Animals and the 3Rs in toxicology research and testing: the way forward. Hum Exp Toxicol 34(12):1297–1303. https://doi.org/10.1177/0960327115598410

Swanson K, Walther P, Leitz J, Mukherjee S, Wu JC, Shivnaraine RV, Zou J (2023) ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. bioRxiv. https://doi.org/10.1101/2023.12.28.573531

Thiel C, Schneckener S, Krauss M, Ghallab A, Hofmann U, Kanacher T, Zellmer S, Gebhardt R, Hengstler JG, Kuepfer L (2015) A systematic evaluation of the use of physiologically based pharmacokinetic modeling for cross-species extrapolation. J Pharm Sci 104(1):191–206. https://doi.org/10.1002/jps.24214

Toma C, Gadaleta D, Roncaglioni A, Toropov A, Toropova A, Marzo M, Benfenati E (2018) QSAR development for plasma protein binding: influence of the ionization state. Pharm Res 36(2):28. https://doi.org/10.1007/s11095-018-2561-8

Tonnelier A, Coecke S, Zaldívar J-M (2012) Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model. Arch Toxicol 86(3):393–403. https://doi.org/10.1007/s00204-011-0768-0

Törnqvist E, Annas A, Granath B, Jalkesten E, Cotgreave I, Öberg M (2014) Strategic focus on 3R principles reveals major reductions in the use of animals in pharmaceutical toxicity testing. PLoS ONE 9(7):e101638. https://doi.org/10.1371/journal.pone.0101638

Vinken M, Benfenati E, Busquet F, Castell J, Clevert D-A, de Kok TM, Dirven H, Fritsche E, Geris L, Gozalbes R, Hartung T, Jennen D, Jover R, Kandarova H, Kramer N, Krul C, Luechtefeld T, Masereeuw R, Roggen E, Schaller S, Vanhaecke T, Yang C, Piersma AH (2021) Safer chemicals using less animals: kick-off of the European ONTOX project. Toxicology 458:152846. https://doi.org/10.1016/j.tox.2021.152846

Volpe DA (2011) Drug-permeability and transporter assays in Caco-2 and MDCK cell lines. Future Med Chem 3(16):2063–2077. https://doi.org/10.4155/fmc.11.149

Votano JR, Parham M, Hall LM, Hall LH, Kier LB, Oloff S, Tropsha A (2006) QSAR modeling of human serum protein binding with several modeling techniques utilizing structure−information representation. J Med Chem 49(24):7169–7181. https://doi.org/10.1021/jm051245v

Watanabe R, Esaki T, Kawashima H, Natsume-Kitatani Y, Nagao C, Ohashi R, Mizuguchi K (2018) Predicting fraction unbound in human plasma from chemical structure: improved accuracy in the low value ranges. Mol Pharm 15(11):5302–5311. https://doi.org/10.1021/acs.molpharmaceut.8b00785

Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, Richard AM (2017) The CompTox chemistry dashboard: a community data resource for environmental chemistry. J Cheminform 9(1):61. https://doi.org/10.1186/s13321-017-0247-6

Willmann S, Lippert J, Sevestre M, Solodenko J, Fois F, Schmitt W (2003) PK-Sim®: a physiologically based pharmacokinetic 'whole-body' model. Biosilico 1(4):121–124. https://doi.org/10.1016/S1478-5382(03)02342-4

Willmann S, Schmitt W, Keldenich J, Lippert J, Dressman JB (2004) A physiological model for the estimation of the fraction dose absorbed in humans. J Med Chem 47(16):4022–4031. https://doi.org/10.1021/jm030999b

Willmann S, Lippert J, Schmitt W (2005) From physicochemistry to absorption and distribution: predictive mechanistic modelling and computational tools. Expert Opin Drug Metab Toxicol 1(1):159–168. https://doi.org/10.1517/17425255.1.1.159

Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, Yin M, Zeng X, Wu C, Lu A, Chen X, Hou T, Cao D (2021) ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res 49(W1):W5–W14. https://doi.org/10.1093/nar/gkab255

Yamazaki K, Kanaoka M (2004) Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. J Pharm Sci 93(6):1480–1494. https://doi.org/10.1002/jps.20059

Yoon M, Campbell JL, Andersen ME, Clewell HJ (2012) Quantitative in vitro to in vivo extrapolation of cell-based toxicity assay results. Crit Rev Toxicol 42(8):633–652. https://doi.org/10.3109/10408444.2012.692115

Yun YE, Edginton AN (2013) Correlation-based prediction of tissue-to-plasma partition coefficients using readily available input parameters. Xenobiotica Fate Foreign Compd Biol Syst 43(10):839–852. https://doi.org/10.3109/00498254.2013.770182

Yun YE, Cotton CA, Edginton AN (2014) Development of a decision tree to classify the most accurate tissue-specific tissue to plasma partition coefficient algorithm for a given compound. J Pharmacokinet Pharmacodyn 41(1):1–14. https://doi.org/10.1007/s10928-013-9342-0

Zhu X-W, Sedykh A, Zhu H, Liu S-S, Tropsha A (2013) The use of pseudo-equilibrium constant affords improved QSAR models of human plasma protein binding. Pharm Res 30(7):1790–1798. https://doi.org/10.1007/s11095-013-1023-6