

In Silico Prediction of Pharmacokinetic Profle for Human Oral Drug Candidates Which Lack Clinical Pharmacokinetic Experiment Data

Jingchen Zhai1 [·](http://orcid.org/0000-0003-2691-8867) Beihong Ji1 [·](http://orcid.org/0000-0003-0387-4056) Shuhan Liu1 [·](http://orcid.org/0000-0002-6627-4843) Yuzhao Zhang1 · Lianjin Cai¹ · Junmei Wang[1](http://orcid.org/0000-0002-9607-8229)

Accepted: 25 January 2022 / Published online: 16 February 2022 © The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Backgrounds and Objectives In silico methods which can generate high-quality physiologically based pharmacokinetic (PBPK) models for arbitrary drug candidates are greatly needed to select developable drug candidates that escape drug attrition because of the poor pharmacokinetic profle. The purpose of this study is to develop a novel protocol to preliminarily predict the concentration profle of a target drug based on the PBPK model of a structurally similar template drug by combining two software platforms for PBPK modeling, the SimCYP simulator and ADMET Predictor.

Methods The method was evaluated by utilizing 13 drug pairs from 18 drugs in the built-in database of the SimCYP software. All drug pairs have Tanimoto scores (TS) no less than 0.5. As each drug in a drug pair can serve as both target and template, 26 sets were studied in this work. Three versions (V1, V2 and V3) of models for the target drug were constructed by replacing the corresponding parameters of the template drug step by step with those predicted by ADMET Predictor for the target drug. V1 represents the replacement of molecular weight (MW), V2 includes the replacement of parameter MW, fraction unbound in plasma (f_u) , blood-to-plasma partition ratio (B/P), logarithm of the octanol-buffer partition coefficient (log $P_{\text{o:w}}$) and acid dissociation constant (pK_a). In V3, all above-mentioned parameters as well as human jejunum effective permeability (P_{eff}), V_d and cytochrome P450 (CYP) metabolism parameters (K_m , V_{max} or CL_{int}) are modified. Normalized root mean square error (NRMSE) was used for the evaluation of the model performance.

Results We found that the performance of the three versions of the models depends on structural similarity of the drug pairs. For Group I drug pairs (TS \leq 0.7), V2 and V3 performed better than V1 in terms of NRMSE; for Group II drug pairs (0.7 $<$ TS \leq 0.9), 8 out of 10 V3 models had NRMSE $<$ 0.2, the cutoff we applied to judge whether the simulated concentrationtime (*C*–*T*) curve was satisfactory or not. V3 outperformed the V1 and V2 versions. For the two drug pairs belonging to Group III (TS > 0.9), V2 outperformed V1 and V3, suggesting more unnecessary replacement can lower the performance of PBPK models. We also investigated how the prediction accuracy of ADMET Predictor as well as its collaboration with SimCYP infuences the quality of PBPK models constructed using SimCYP.

Conclusion In conclusion, we generated practical guidance on applying two mainstream software packages, ADMET Predictor and SimCYP, to construct PBPK models for drugs or drug candidates that lack ADME parameters in model construction.

1 Introduction

Pharmacokinetics is the study of the time courses of a drug administered to the body, which includes the processes of absorption, distribution, metabolism and excretion (ADME) [\[1](#page-13-0)]. Usually, it is essential to quantitatively measure the concentration of the drug in plasma at diferent time points in a

Jingchen Zhai and Beihong Ji contribute equally to this paper.

 \boxtimes Junmei Wang Junmei.wang@pitt.edu pharmacokinetic study for the analysis of drug behavior and dose adjustment. In addition to clinical trials, which always involve time, cost and ethical considerations, the prediction of concentration profles under various administration conditions can also be achieved by the implementation of physiologically based pharmacokinetic (PBPK) [\[2–](#page-13-1)[4\]](#page-13-2) modeling. Computational tools for both PBPK modeling and pharmacokinetic parameter prediction have been developed, further reducing the experimental expense. By virtue of such tools, quick and convenient in silico prediction of drug behavior in the human body can be easily performed without investing many resources in the experiments, informing further studies in drug toxicity, dosing strategy and potential drug-drug

Extended author information available on the last page of the article

Key Points

We developed an effective computational protocol for the generation of valid physiologically-based pharmacokinetics (PBPK) models for arbitrary molecules, an important task in preclinical drug discovery.

This method utilizes ADME Predictor to calculate pharmacokinetic parameters as inputs for PBPK modeling and simulation using Simcyp simulator.

More than 60% of compounds have satisfactory performance utilizing this method.

interactions. As such, this PBPK modeling can be particularly useful in the preclinical phase and can serve as a tool to help select drug candidates that are more likely to have desirable pharmacokinetic profles.

In the literature, one study predicted the bioavailability (Fa%) of a structurally diverse group of drugs using theoretical descriptors and neural network modeling [[5\]](#page-13-10). Another study applied a genetic algorithm to optimize the prediction model for drug Fa%, plasma protein binding and urinary excretion [[6\]](#page-13-11). There are also studies predicting the Fa% of a chemical series with GastroPlus [[7,](#page-13-12) [8](#page-13-13)]. Evaluation of the Fa% prediction performances from diferent software platforms, SimCYP and GastroPlus, has also been conducted focusing on low-solubility drugs [\[9\]](#page-13-14). Collectively, these studies focused on the value of Fa% and area under the curve (AUC) as the most important parameters for drugs after administration, but these parameters cannot fully explain the shape of the drug concentration-time $(C-T)$ profile. Therefore, how a drug is absorbed, distributed, metabolized and excreted in the course of time still lacks systematic prediction guidance.

This research aims to develop a pure in silico method to predict the pharmacokinetic profile of a compound efficiently, taking advantage of the available high-quality PBPK models in the Simcyp compound library and public domain. This method has potential application in selecting drug candidates with favorable pharmacokinetic profles to enter the next stage of drug development.

2 Methods

In this study, we developed a novel method to predict the plasma concentration profle of a target compound based on PBPK models constructed using the model of a structurally

similar drug that serves as the template. We utilized the Sim-CYP simulator (V19, Release 1; Shefeld, UK) [[10](#page-13-3)] software to construct PBPK models for a target drug by only substituting the predicted ADME parameters of the target drug for those from the PBPK model of the corresponding template drug. We applied ADMET Predictor (V9.5, Simulation Plus) [\[11,](#page-13-4) [12](#page-13-5)], a software developed by Simulation-Plus Inc. to predict the ADME properties of target drugs, which include physiochemical parameters such as fraction unbound in plasma (f_n) and blood-to-plasma partition ratio (*B*/*P*) and ADME input parameters such as volume of distribution (V_d) , Michaelis-Menten constant (K_m) and maximal metabolism rate (V_{max}) of common enzymes. To better validate our constructed PBPK models as well as evaluate the performance of the two software tools, we selected 18 drugs from the SimCYP compound library (including substrates and inhibitors) as the template drugs. In total, 13 drug pairs were formed based on their structural similarity. For each pair of drugs, one serves as the template and the other serves as the target drug. For the target drug in a drug pair, we pretended that no PBPK model was available, and new PBPK models were constructed based on the PBPK model of the template drug. We tested three protocols by introducing ADMET Predictor predicted ADME properties into the template PBPK model and evaluated the model performance using the observed pharmacokinetic profle of the target drug. The corresponding PBPK models constructed using the three protocols were called V1, V2 and V3 models.

2.1 Drug Preparation

Drugs selected for the construction of in silico PBPK models come from the built-in drug database of the SimCYP software. Simplifed Molecular-Input Line-Entry System (SMILES) [\[13](#page-13-6)] strings of all drugs from the SimCYP builtin library, including substrates and inhibitors, were collected from the DrugBank database [[14](#page-13-7)]. The SMILES strings of drugs were used not only for their structural similarity calculation on a web platform, but also as inputs for the generation of their properties using the ADMET Predictor.

2.2 Structure Similarity Calculation

Tanimoto scoring is a commonly used method to compute the fngerprint-based similarity between two compounds $[15]$ $[15]$. In this study, we applied the maximum common substructure based (MCS) Tanimoto algorithm for the similarity calculation. The Tanimoto score (TS) is defned by the function below (Eq. [1\)](#page-1-0) $[16]$ $[16]$ $[16]$:

$$
TS(X, Y) = \frac{N_Z}{N_X + N_Y - N_Z} \tag{1}
$$

where N_X and N_Y are the numbers of bits in fragment bit strings of the two compounds and N_Z is the intersection set, i.e., the number of common substructures shared by these two compounds. TS (*X*, *Y*) ranges from 0 to 1, measuring the structural similarity between two compounds from the lowest to highest (when the two molecules are identical). TS scores were calculated using ChemMine for all combinations of drugs in the SimCYP compound database [\[17](#page-13-15)].

2.3 Validation of PBPK Models for Drug Templates

We frst validated the PBPK models of all 18 selected drugs by utilizing their observed data from the literature. In detail, we utilized the original built-in models of those drugs in SimCYP to run the simulation. In terms of the trial design, the dose regimens, simulation time as well as population information including age, weight and health condition were the same as those reported in the clinical study of pharmacokinetics measurement. Meanwhile, the parameters of the built-in PBPK model, such as the drug's ADME properties, remained the same for all the drugs except fuoxetine. As a racemate, we adjusted some of its ADME and pharmacokinetic parameters according to the literature to make the predicted curve ft the experimental data much better [\[18–](#page-13-16)[20\]](#page-14-0). The key ADME parameters predicted by ADMET Predictor for the 18 drugs are all listed in Table S1, including the details of the adjusted parameters of fuoxetine. The observed drug concentration data of each template drug were extracted from published concentration-time (C-T) curves using WebPlotDigitizer [\[21\]](#page-14-1). The *C*–*T* curves from simulations were then overlaid to the observed drug concentrations. The predicted pharmacokinetic profles of each template drug, including the maximal concentration (C_{Max}) , time at which C_{Max} was observed (T_{Max}) and area under the curve (AUC), were compared to the observed ones.

2.4 Evaluation of Inherent Diferences Among Software Platforms

The quality of models constructed for target drugs is not only afected by the structural similarity between the template drug and the target drug but also relies on the prediction quality of ADMET Predictor and how good the collaboration is between the software. There may be some inherent diferences among diferent software platforms, including but not limited to the training set data and algorithms for constructing models. More importantly, the prediction accuracy of ADMET Predictor for an individual ADME parameter is unknown. Thus, we utilized parameters predicted by ADMET Predictor for the 18 drugs to simulate their pharmacokinetic profles using SimCYP and then compared them to those predicted using SimCYP built-in parameters. Since the calculation of molecular weight (MW) must be

very accurate, the reliability of this parameter from ADMET Predictor for each drug will not be evaluated (Category I). The following ADME parameters predicted by ADMET Predictor belong to Category II: *B*/*P*, *F*u, the logarithm of octanol-buffer partition coefficient (log $P_{\text{o:w}}$) and acid dissociation constant (pK_a); ADME parameters in Category III include human jejunum effective permeability (P_{eff}) , V_{d} and cytochrome P450 (CYP) metabolism parameters (K_m , V_{max}) or CL_{int}). The prediction accuracy decreases from Category I to Category II and then to Category III. The values of these ADME parameters for 18 drugs are listed in Table S1. To investigate the diferent qualities of the calculated parameters, we modifed the template step by step by introducing more and more ADME predicted parameters. Specifcally, in substitution protocol Version 2 (V2), we replaced log $P_{\text{o:w}}$, pK_a , B/P and F_u values in the SimCYP drug template with the calculated results from ADMET Predictor. In substitution protocol Version 3 (V3), all the above-mentioned ADME parameters, which not only include the parameters mentioned by V2 but also P_{eff} in absorption, V_d in distribution and CYP metabolism parameters of template drug, were replaced by predicted values of ADMET Predictor.

2.5 Model Construction for Target Drugs

The parameter substitution plan is the same as that for ADME Predictor software evaluation in Sect. [2.4](#page-2-0). In total, three versions of PBPK models for a target drug were built by modifying the models of the template drug: (1) in Version 1 (V1), only the MW of template drug was changed to that of the target one; (2) in Version 2 (V2), in addition to the MW, other parameters of template drug, which are the same as in the above-mentioned Version 2, were replaced by those predicted for the target drug; (3) in Version 3 (V3), in addition to MW and physiochemical properties, P_{eff} , V_d and CYP of templates were also replaced with the calculated ones for the target drug, in accordance with above-mentioned Version 3. All the ADME properties of the target drugs are predicted by ADMET Predictor, a software tool that can predict > 140 properties based on its built-in quantitative structure-activity relationship (QSPR) models [[22](#page-14-2)]. Information about the experimental subjects and trial design of each target drug during simulations was derived from the corresponding clinical reports.

2.6 Evaluation of Models for Target Drugs

To evaluate the performance of PBPK models with input parameters from ADMET Predictor, the experimental data of target drugs were overlaid by the simulated *C*–*T* curves. To quantitively evaluate how well the experimental and simulated curves overlaid each other, we calculated the root mean square error (RMSE) [\[23\]](#page-14-3) of the observed and predicted concentrations at diferent time points. The formula for the RMSE calculation is as follows (Eq. [2\)](#page-3-0):

RMSE =
$$
\left[\sum_{i=1}^{N} (C_{pi} - C_{oi})^2 / N\right]^{\frac{1}{2}}
$$
 (2)

where C_{oi} and C_{pi} represent the observed and predicted drug concentration at time point *i*. *N* is the number of time points $(N > 1)$ from the extracted observed data. Specifically, in this study, to facilitate the comparison between models for diferent drugs with various concentration scales, we introduced normalized root mean square error (NRMSE) to evaluate the performance of PBPK models, which is calculated using the following formula $(Eq. 3)$ $(Eq. 3)$ $(Eq. 3)$:

$$
NRMSE = \frac{RMSE}{C_{\text{max}} - C_{\text{min}}} \tag{3}
$$

where C_{max} and C_{min} are the maximum and minimum values among the observed and predicted concentrations using all three versions of models.

The flowchart of the experiment protocol is shown in Fig. [1](#page-3-2).

3 Results

3.1 Drug Pair Selection and Validation of PBPK Models for Drug Templates

Thirteen pairs out of 18 drugs, which have calculated TS \geq 0.5, were selected for in silico PBPK modeling. Drug pairs with TS < 0.5 were not considered to be structurally similar and were excluded in this study. The calculated TS for 13 selected pairs (Groups A–M) is listed in Table [1](#page-4-0). Since both drugs in a pair will in turn serve as the template and target drug for cross validation, we used X-1 and X-2 to label two drugs in the pair, respectively, where X can be A to M.

The predicted mean plasma concentration-time profles overlaid with observed data of all 18 template drugs are shown in Fig. [2](#page-5-0). Accordingly, Table [2](#page-6-0) [\[20,](#page-14-0) [24](#page-14-4)–[39\]](#page-14-5) exhibits the predicted pharmacokinetic parameters $(C_{\text{Max}}, T_{\text{Max}})$ AUC) versus observed values. From Table [2,](#page-6-0) excluding the drugs with unavailable observed pharmacokinetic parameters (dextromethorphan, mephenytoin and fluoxetine), the predicted pharmacokinetic parameters of most drugs are within the standard deviation ranges of their observed values. The predicted values of C_{Max} , T_{Max} and AUC for theophylline are all slightly beyond the margin of error but still within the range of two-fold standard deviation. Overall, as shown in Fig. [2,](#page-5-0) the observed *C*–*T* profles are within the 95% confdence interval (CI) ranges (upper and lower gray dashed curves) of the simulated *C*–*T* curves. Therefore, the PBPK models for the template drugs have been well validated.

3.2 Evaluation of Inherent Diferences Among Software Platforms

The predicted pharmacokinetic parameters of the 18 modifed drug templates by replacing the ADME parameters with those predicted by ADMET predictor are listed in Table [2.](#page-6-0) The *C*–*T* profles of those 18 drugs are shown in Fig. [3](#page-8-0) (V2)

Fig. 1 Flowchart of experiment protocol. pK_a acid dissociation constant, $log P_{o:w}$ logarithm of octanol-buffer partition coefficient, *B*/*P* blood-to-plasma partition ratio, F_u fraction unbound in plasma, V_{ss}

volume of distribution at steady state, P_{eff} human jejunum effective permeability, K_m Michaelis-Menten constant, V_{max} maximal metabolism rate, *V1* version 1, *V2* version 2, *V3* version 3

Table 1 Calculated Tanimoto coefficient between each pair of drugs

Group	Drug 1	Drug 2	Tanimoto score
A	Bupropion	Dextromethorphan	0.50
B	Bufuralol	Bupropion	0.52
C	Dextromethorphan	Ouinidine	0.57
D	Lorazepam	Midazolam	0.63
E	Alprazolam	Lorazepam	0.65
F	Lorazepam	Triazolam	0.69
G	Mephenytoin	Phenobarbital	0.74
H	Atomoxetine	Fluoxetine	0.78
I	Simvastatin	Pravastatin	0.82
J	Triazolam	Midazolam	0.84
K	Midazolam	Alprazolam	0.88
L	Theophylline	Caffeine	0.93
М	Imipramine	Desipramine	0.95

and Fig. [4](#page-10-0) (V3). In V2, most drugs exhibit satisfying prediction results. As Fig. [3](#page-8-0) shows, in 14 of 18 drugs most of the experimental data points lay within the predicted confdence interval. Only in triazolam, atomoxetine, simvastatin and pravastatin do nearly or more than half of the data points exceed the confdence interval, showing poor prediction performance. V3 shows that bupropion, cafeine and phenobarbital have a very good overlay between the clinical report and predicted result from the modifed drug template, with the observed data lying within the confdence interval of predicted curve. For fuoxetine, alprazolam, quinidine and triazolam, although the predicted results do not show an excellent overlay with the experimental data, most of the clinical data points lay within the confdence interval of the prediction profles. For lorazepam, although the observed data were all at or around the upper confdence interval of the predicted profle, the shape of the predicted curve is very similar to that of the observed pharmacokinetic profle. Unfortunately, the other drugs do not show very satisfying prediction results, using clinical data points as reference.

To quantitatively measure the deviation of predicted concentration profles from the experimental data, the diference between observed and predicted values was evaluated by NRMSE (Table [3\)](#page-9-0). The lower the NRMSE value is, the smaller the diference between the predicted and experimental concentration profle, i.e., the better the performance of the created drug model. The average NRMSE of V2 is 0.26 compared with the average value of 0.43 for V3, showing that V2 can introduce less prediction error when combining the two software platforms for prediction. Especially for V2, although dextromethorphan has an NRMSE value as large as 0.45, this should be caused by the deviation of the curve from the frst data point. All the remaining data points are very close to the predicted curve. Fourteen of 18 drugs have NRMSE values < 0.4 , and 7 of them are < 0.2 , showing the satisfying prediction and collaboration quality of the two software tools. For V3, the top three drugs, caffeine, phenobarbital and bupropion, all have very small NRMSE values, which is consistent with the fact that the simulated *C*–*T* curves are well overlain with the experimental data points as shown in Fig. [4](#page-10-0). Interestingly, the NRMSE values of fuoxetine (0.41), alprazolam (0.28), quinidine (0.53) and triazolam (0.29) are quite diferent, even though the simulated *C*–*T* curves of the four drugs are relatively satisfactory. Taken together, both the overlay of simulated *C*–*T* curves with the measured *C*–*T* data points and NRMSE should be used to evaluate the quality of the predicted ADME parameters by the ADMET predictor. Overall, the predicted ADME parameters according to ADMET Predictor can produce satisfactory *C*–*T* curves using SimCYP simulator for about half of the tested drugs.

As illustrated in Figs. [3](#page-8-0) and [4,](#page-10-0) more V2 version models (Fig. [3](#page-8-0)) have better performance than V3 version models (Fig. [4](#page-10-0)), suggesting $\log P_{\text{o:w}}$, pK_a , B/P and F_u are more accurately predicted by ADMET Predictor than P_{eff} , V_{d} and CYP parameters. Regarding a specifc parameter, the prediction performance varies from one compound to another. Thus, we recommend adopting a diferent version of parameter substitution mainly based on the structural similarity between the template and target drugs. When the structural similarity is very high $(TS > 0.9)$, fewer parameter substitutions are prefered, while when the structural similarity is not very high, more parameter substitutions are desirable, as the prediction errors are smaller than the diferences of the parameters between the target and the template.

3.3 Predicted Concentration Profles for the in Silico PBPK Models

The *C*–*T* profles predicted by all three versions (Versions 1, 2, and 3) of PBPK models are shown in Fig. [5](#page-11-0). The NRMSE value is also calculated to measure the diferences between observed and predicted values of three versions, respectively, which are summarized in Table [4](#page-12-0). The table cell is marked with "*" if the NRMSE value of V1, V2 or V3 is < 0.2 . In the following, we grouped all 13 drug pairs/26 drug pair sets into three groups according to their Tanimoto scores for the sake of discussion.

Group I (TS \leq 0.7). Six drug pairs, A–F, belong to this group. According to Table [4](#page-12-0), the performance of the three protocols does not show an obvious pattern for Group I. V1, V2 and V3 have two (A-1 and D-1), fve (A-1, A-2, D-1, D-2 and F-1) and three (B-2, C-2 and D-2) pair sets in "*" table cells, respectively. Most of those pair sets also exhibit a good overlay between experimental data points and prediction curves as shown in Fig. [5,](#page-11-0) indicating the collaboration between SimCYP and ADMET Predictor is good. For the other groups from A-1 to F-2, all three protocols **Fig. 2** Predicted concentration profles and observed data of all drugs. Prediction results for all drugs except fuoxetine are from the original SimCYP template. The result for fuoxetine is from the adjusted fuoxetine template. Upper and lower dashed gray curves represent 95% confdential interval

have NRMSE values > 0.2 , and the simulated C-T curves do not overlay the experimental data points well. Interestingly, for the D-2 drug pair set, although the NRMSE of the V2 model is the lowest, the predicted C-T curve by the V3 model has a better shape ftting of the observed data as shown in Fig. [5.](#page-11-0) This phenomenon is caused by the deviation of the frst data point from the predicted curve of V3, which caused its NRMSE to be larger than that of V2. When this outlier is eliminated and the NRMSE value is recalculated, V3 becomes the best for this pair set (NRMSEs are now 0.57, 0.16 and 0.06 for the V1, V2 and V3 protocols, respectively).

Group II ($0.7 <$ TS ≤ 0.9). This group contains five drug pairs, G–K. As shown in Table [4,](#page-12-0) most drug pair sets have at least one version with NRMSE value < 0.2, except H-1 and I-2. Notably, the NRMSE value of I-2 is only 0.21, and the predicted *C*–*T* curve exhibits good consistency with experimental data (Fig. [5](#page-11-0)). The failure of the H-1 model is likely caused by using problematic ADME parameters predicted by ADMET Predictor for the target drug. The "collaboration" between the two software tools should not be a problem for **Table 2** Comparison between predicted and observed pharmacokinetic profles of all drugs

Pred predicted drug pharmacokinetic parameters from the unchanged SimCYP drug template (except fuoxetine; for fuoxetine especially, the SimCYP drug template is modifed to enable the predicted profle ft of the clinically reported curve), *Pred_V2* predicted drug pharmacokinetic parameters using SimCYP with input parameters (log P_{ow} , pK_a, B/P and F_{u}) from ADMET Predictor, $Pred_{\text{v}}$ predicted drug pharmacokinetic parameters using SimCYP with input parameters (log $P_{\text{o:w}}$, pK_a, B/P , F_u , P_{eff} , V_d and CYP parameters) from ADMET Predictor, *Obs* drug pharmacokinetic parameter reported by clinical research, C_{Max} maximal concentration, T_{Max} the time at which C_{Max} is observed, *AUC* area under the curve, pK_a acid dissociation constant, *log* $P_{o:w}$ logarithm of octanol-buffer partition coefficient, *B*/*P* blood-to-plasma partition ratio, F_u fraction unbound in plasma, V_{ss} volume of distribution at steady state, P_{eff} human jejunum effective permeability, K_m Michaelis-Menten constant, V_{max} maximal metabolism rate

this drug pair since the NRMSE values of H-2 are very low for both the V2 and V3 models, which are 0.08 and 0.02 for the two models correspondingly. As shown in Table [4,](#page-12-0) the V3 version models apparently outperform the V1 and V2 models for most drug pair sets, as seven out of ten V3 models have NRMSE values < 0.2 , while none of the V1 models and 2 V2 models have NRMSE values < 0.2. Interestingly, for drug pair set J-2, the V2 and V3 models have highly similar performances with good prediction results as shown in Fig. [5;](#page-11-0) however, for K-2, all three model versions do not exhibit satisfying prediction (Fig. [5](#page-11-0)), even though the NRMSE values of the V1 and V2 models are equal to or lower than the cutoff.

Group III ($TS > 0.9$). This group contains two drug pairs, L and M. As shown in Table [4](#page-12-0), most models have satisfactory NRMSE values. For L-1 and L-2 drug pair sets, the

 Δ Adis

predicted profles of the V2 and V3 models are very close to the clinical data points. Interestingly, for M-1 and M-2 drug pair sets, the performance of the V3 models is very poor. Drug pair M has structural similarity with the TS of 0.95; interestingly, the V3 models perform poorly while the V1 and V2 models have not only satisfactory NRMSE values but also very well-overlain C-T curves with measured data points. This phenomenon may be explained by the prediction error by ADMET Predictor, and error caused by the inherent diference between the two software platforms can be compensated by the small diference in the ADME parameters between the template and target drugs. Indeed, the NRMSE values of the two drugs in drug pair M, 0.51 and 0.70, are very large (Table [4](#page-12-0)).

As shown in Fig. [5](#page-11-0), the performance of three parameter substitution versions varied from one drug pair to another **Fig. 3** Predicted concentration profles using SimCYP drug template with input parameters from ADMET Predictor (log $P_{\text{o:w}}$, pK_a, *B*/*P* and F_{u}) and observed data of all drugs. Upper and lower dashed gray curves represent 95% confdential interval. *Po:w* logarithm of octanol-bufer partition coefficient, *B/P* blood-to-plasma partition ratio, F_u fraction unbound in plasma, pK_a acid dissociation constant

mainly depending on the net effect of eliminating two sources of errors, the prediction errors of ADMET Predictor and the errors of applying the template model to describe the target. For the frst source of errors, more and more prediction errors are introduced from V1 to V2 and then to V3. The second source of errors is big for dissimilar drug pairs (Group I) and small for highly similar drug pairs (Group III). For a structurally dissimilar drug pair, V2 or V3 are necessary to overcome the large second source errors, even though more frst source errors are introduced. On the other hand, for a structurally similar drug pair, V1 or V2 is preferrable as the errors from both sources are small. More discussion on choosing proper versions of a parameter substitution scheme is provided below.

4 Discussion

In this study, we developed a novel approach to construct in silico PBPK models for target drugs lacking experimental ADME and other pharmacokinetic parameters using an established PBPK model of a structurally similar drug as the model template. We used 18 drugs, which formed 13 drug pairs (A-M) and 26 drug pair sets (each drug in a pair serves the template and target roles alternatively) to evaluate three ADME parameter substitution protocols, which are corresponding to three versions of PBPK models. The performance of the in silico PBPK models was critically evaluated using experimental pharmacokinetic profles and parameters.

4.1 Practical Guidance on Selecting a Suitable Drug Template

We attempted to obtain guidance on selecting a suitable template drug for a given target drug. We focused on using structural similarity to select the template drugs. It was found that drug pairs with Tanimoto score > 0.70 (Groups II and III) tended to show better prediction performance among the three versions compared with drug pairs with Taminoto score < 0.70 (Group I). It is obvious that the higher structural similarity of two drugs within a drug pair should contribute to the higher possibility of good prediction results. After comparing the model performance of all three versions of models, we developed the following guidance: for Group I drug pairs, V2 or V3 is recommended; for Group II drug pairs, V3 is recommended; for Group III drug pairs, V2 is recommended. Following this practical guidance, 16 out of 26 drug pair sets have NRMSE values < 0.2, the threshold to recognize a good PBPK model. Nevertheless, the prediction accuracy of ADMET Predictor and the extent of inherent diference between it and SimCYP are also crucial factors that affect the model performance. From the evaluation of the error caused by combining the two software tools, the prediction accuracies of each modifed drug template varied from each other, which shows the infuence of the introduced error can be very diferent for diferent drugs.

Additional criteria other than structural similarity between the template and target drugs may be introduced to further improve the computational protocol since the prediction performance of drug pairs with the similar/same TS may have diferent prediction accuracies as indicated by Fig. [5](#page-11-0). This phenomenon is more obvious for drug pair sets with low TS. For example, for drug pair set D (lorazepam/ midazolam), the prediction for midazolam by V3 version of parameter substitution is much more accurate than that

Table 3 Calculated normalized root mean square error (NRMSE) between predicted results by modifed drug template and experimental concentration profles of drugs

for lorazepam. This discrepancy may come from the failure of parameter prediction by ADMET Predictor and/or the imperfect collaboration between the two software platforms. Fortunately, this inconsistency problem becomes less severe when the drug pairs share higher structural similarities as for drug pair sets from G to M.

4.2 Another Possible Method to Evaluate the Prediction Results of the Three Versions

There is also another method to evaluate the prediction results of V1, V2 and V3, which is the fold error in the AUC of the three prediction versions compared to the clinical data. However, the fold error in the AUC can only show the diference between the total area under the prediction curve and the literature-reported pharmacokinetic curve without delineating the concrete shapes of curves. Contrarily, the shape of the predicted drug C-T curve can be refected by the diference between predicted and observed drug concentrations at each time point when using RMSE as an evaluation method. Furthermore, the variation of the dosages can contribute to large RMSE discrepancy among drugs. Therefore, we normalized RMSE to eliminate the infuence of dosages on RMSE values. The utilization of NRMSE can help to reduce the false-positive rate.

Fig. 4 Predicted concentration profles using SimCYP drug template with input parameters from ADMET Predictor $(\log P_{\text{o:w}}, \text{pK}_a, \text{B/P}, F_u, P_{\text{eff}},$ V_d and CYP parameters) and observed data of all drugs. Upper and lower dashed gray curves represent 95% confdential interval. $P_{o:w}$ logarithm of octanol-buffer partition coefficient, *B/P* blood-to-plasma partition ratio, F_u fraction unbound in plasma, pK_a acid dissociation constant, *Pef* human jejunumeffective permeability, V_d volume of distribution, *CYP* cytochrome P450

4.3 Perspective of Applying in Silico PBPK Modeling for Compounds Lack Experimental ADME and Pharmacokinetic Properties

SimCYP simulator is an advanced software with well-constructed drug pharmacokinetic models in its built-in drug library, with each drug template containing comprehensive drug parameters. It can intuitively show simulated drug *C*–*T* curves contributed by these parameters under diferent trial designs. On the other hand, ADMET Predictor can predict many pharmacokinetic parameters of an input compound based on its structural information without giving additional

Fig. 5 Predicted concentration profles of three versions and observed data of all predicted drugs

information. However, constructing a drug pharmacokinetics model needs full-scale pharmacokinetic parameters, and some of them cannot be predicted reasonably. Considering this, we can partially rely on the pharmacokinetic parameters of another compound which shares high structural similarity with the unknown target compound. In this study, we put forward a novel approach to build PBPK models for a target drug with a lack of measured ADME and other pharmacokinetic parameters using the PBPK model of a template drug which is structurally similar to the target drug. Also, we proposed overall guidance on selecting a suitable template drug and using its PBPK model as the model template.

The success of this computational approach depends on two important factors, the availability of a high-quality PBPK model for the template compound and accuracy and consistency of the ADME and pharmacokinetic parameters predicted by ADMET Predictor software for the target drug. Thus, the performance of two software tools can greatly contribute to the experimental results of our study. As a calculator of ADMET properties for compounds, the prediction results of drug properties may not be close enough to the real state, leading to errors when constructing drug models. Additionally, not all the ADME/pharmacokinetic properties can be calculated with the current version of ADMET Predictor. For example, the prediction of metabolism in ADMET Predictor is limited to only fve commonly used enzymes (CYP1A2, CYP2D6, CYP2C9, CYP2C19 and CYP3A4), and the prediction results of the transporters related to the drug can only be reported qualitatively rather than quantitively. On the other hand, there are currently 70 established compounds in SimCYP's drug libraries (including both the substrate and inhibitor libraries), and the libraries are still under development. We tested 18 compounds that shared structural similarity, and this study will be flled out as more clinically validated PBPK models and related parameters for in-use drugs become available.

At the current stage, application of this proposed method to construct a PBPK model for a candidate compound may encounter some difficulties, such as failing to identify a template drug in the SimCYP library that shares high structural similarity with the target compound or several template drugs being identifed in SimCYP library that have similar structural similarity with the target compound. For the frst problem, a practical solution is to greatly expand the library of PBPK models, and we are now constructing high-quality PBPK models for the top-selling drugs. For the second problem, we can add additional criteria to further prioritize the templates; those criteria include but are not limited to key ADME properties (such as aqueous solubility, permeability and metabolism profle) and drug targets.

Nevertheless, we have proposed a practical approach to generate PBPK models for a compound lacking experimental ADME/pharmacokinetic properties. This model can serve as the initial version of the PBPK models for the target compound, and its performance can be improved using the measured pharmacokinetic profles and properties in the future. The computational protocol introduced in this work can have important applications in selecting drugs to enter the drug optimization phase or drug candidates to enter preclinical studies.

Table 4 Calculated normalized root mean square error (NRMSE) between predicted (three versions) and experimental concentration profles of drugs in each drug pair set

Drug group	Drug pair set	NRMSE (different versions vs. obs)			Tanimoto score
		V ₁	V2	V3	
Group I	A-1	$0.14*$	$0.13*$	0.49	0.50
	$A-2$	0.26	$0.19*$	0.50	0.50
	$B-1$	0.44	0.49	0.49	0.52
	$B-2$	0.67	0.49	$0.07*$	0.52
	$C-1$	0.64	0.34	0.31	0.57
	$C-2$	0.68	0.67	$0.13*$	0.57
	$D-1$	$0.14*$	$0.16*$	0.48	0.63
	$D-2$	0.61	$0.16*$	$0.19*$	0.63
	$E-1$	0.22	0.32	0.35	0.65
	$E-2$	0.43	0.56	0.35	0.65
	$F-1$	0.24	$0.19*$	0.33	0.69
	$F-2$	0.88	0.28	0.27	0.69
Group II	$G-1$	0.94	0.94	$0.04*$	0.74
	$G-2$	0.58	0.56	$0.14*$	0.74
	$H-1$	0.56	0.56	0.52	0.78
	$H-2$	0.44	0.08	$0.02*$	0.78
	$I-1$	0.49	0.36	$0.04*$	0.82
	$I-2$	0.34	0.38	0.21	0.82
	$J-1$	0.43	0.57	$0.14*$	0.84
	$J-2$	0.45	$0.13*$	$0.12*$	0.84
	$K-1$	0.69	0.67	$0.06*$	0.88
	$K-2$	0.20	$0.15*$	0.38	0.88
Group III	$L-1$	0.34	$0.06*$	$0.12*$	0.93
	$L-2$	0.22	$0.19*$	$0.17*$	0.93
	$M-1$	$0.08*$	$0.10*$	0.66	0.95
	$M-2$	$0.13*$	$0.15*$	0.40	0.95

The normalized root mean square errors (NRMSEs) of the target and template drugs, which are adopted from Table [3](#page-9-0), measure the quality of the ADME prediction using ADMET Predictor and/or the inherent diference between the two software platforms. The Tanimoto scores in the last column come from Table [1](#page-4-0)

Obs, the drug pharmacokinetic parameter reported by clinical research; V1, version 1; V2, version 2; V3, version 3

5 Conclusions

In this work, we have introduced and tested a novel computational protocol to develop an in silico PBPK model for a compound lacking measured ADME/pharmacokinetic properties and pharmacokinetic profles. The general idea is to choose a proper PBPK model as the template when the

corresponding compound, the template drug, is structurally similar to the target drug. For the target drug, we calculated the ADME properties using ADMET Predictor of SimulationPlus Inc. We have developed an overall guidance using this method to build PBPK models for an arbitrary drug. First, the structural similarity between the template and target drug is very important; thus, template drugs that have the highest structural similarity to the target drug should be considered frst; second, once the target drug is selected, the ADME parameter substitution protocol is selected based on the Tanimoto score between the target and template drugs. If TS is \leq 0.7, V2 or V3 protocol is recommended; if TS is > 0.7 but ≤ 0.9 , V3 protocol is suggested. If TS is > 0.9 , V2 is recommended. Following this guidance, $> 60\%$ (16 out of 26) of the PBPK models have satisfactory performance. It is emphasized that this method relies greatly on the collaboration between SimCYP and ADMET Predictor as well as the prediction accuracy of ADMET Predictor. The NRMSE values of the template and target drugs can guide us to select proper substitution protocols. If the NRMSE values are small, one can select a protocol with many ADME parameters being substituted, such as V3; however, if the NRMSE values are large, adopting V2 or V1 protocols can minimize the error due to the poor "collaboration" between the two software platforms. Unfortunately, the NRMSE value of the target drug is unknown in practice. A tool which can predict this NRMSE parameter is thus needed to further improve this method. While future experimental work is defnitely needed to further improve the model performance, our novel approach proposed in this work can help identify drug candidates with favorable pharmacokinetic profles, reducing experimental cost and providing insight into drug discovery and development.

Supplementary Information The online version contains supplementary material available at<https://doi.org/10.1007/s13318-022-00758-9>.

Declarations

Conflict of interest The authors declare no competing interests.

Funding The authors gratefully acknowledge the funding support from the National Science Foundation (1955260) and National Institutes of Health of the USA (R01GM079383, R21GM097617 and P30DA035778).

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material All data have been provided in the manuscript and supplemental information.

Code availability Not applicable.

Author contributions JZ, JW and SL designed the research. JZ and BJ performed the research and analyzed the data, all authors wrote the manuscript.

References

- 1. Eddershaw PJ, Beresford AP, Bayliss MK. ADME/PK as part of a rational approach to drug discovery. Drug Discov Today. 2000;5(9):409–14.
- 2. Lu J, Goldsmith MR, Grulke CM, Chang DT, Brooks RD, Leonard JA, Phillips MB, Hypes ED, Fair MJ, Tornero-Velez R, Johnson J, Dary CC, Tan YM Developing a physiologically-based pharmacokinetic model knowledgebase in support of provisional model construction. PLoS Comput Biol. 2016;12(2):e1004495.
- 3. Zhuang X, Lu C. PBPK modeling and simulation in drug research and development. Acta Pharm Sin B. 2016;6(5):430–40.
- 4. Lin L, Wong H. Predicting oral drug absorption: mini review on physiologically-based pharmacokinetic models. Pharmaceutics. 2017;9(4):41.
- 5. Turner JV, Maddalena DJ, Agatonovic-Kustrin S. Bioavailability prediction based on molecular structure for a diverse series of drugs. Pharm Res. 2004;21(1):68–82.
- 6. Wang J, Krudy G, Xie XQ, Wu C, Holland G. Genetic algorithmoptimized QSPR models for bioavailability, protein binding, and urinary excretion. J Chem Inf Model. 2006;46(6):2674–83.
- Daga PR, Bolger MB, Haworth IS, Clark RD, Martin EJ. Physiologically based pharmacokinetic modeling in lead optimization. 1. Evaluation and adaptation of GastroPlus to predict bioavailability of medchem series. Mol Pharm. 2018;15(3):821–30.
- 8. Daga PR, Bolger MB, Haworth IS, Clark RD, Martin EJ. Physiologically based pharmacokinetic modeling in lead optimization. 2. Rational bioavailability design by global sensitivity analysis to identify properties afecting bioavailability. Mol Pharm. 2018;15(3):831–9.
- 9. Matsumura N, Hayashi S, Akiyama Y, Ono A, Funaki S, Tamura N, Kimoto T, Jiko M, Haruna Y, Sarashina A, Ishida M, Nishiyama K, Fushimi M, Kojima Y, Yoneda K, Nakanishi M, Kim S, Fujita T, Sugano K. Prediction characteristics of oral absorption simulation software evaluated using structurally diverse low-solubility drugs. J Pharm Sci. 2020;109(3):1403–16.
- 10. Jamei M, Marciniak S, Feng K, Barnett A, TuckerRostami-Hodjegan GA. The Simcyp population-based ADME simulator. Expert Opin Drug Metab Toxicol. 2009;5(2):211–23.
- 11. Hassan SF, Rashid U, Ansari FL, Ul-Haq Z. Bioisosteric approach in designing new monastrol derivatives: an investigation on their ADMET prediction using in silico derived parameters. J Mol Graph Model. 2013;45:202–10.
- 12. Hecht D, Fogel GB. Computational intelligence methods for ADMET prediction. Front Drug Des Discov. 2009;4:351–77.
- 13. Toropov AA, Toropova AP, Mukhamedzhanoval DV, Gutman I. Simplifed molecular input line entry system (SMILES) as an alternative for constructing quantitative structure-property relationships (QSPR). Indian J. Chem. 2005;44:1545–1552.
- 14. <https://www.drugbank.ca/>. Accessed 14 Feb 2022.
- 15. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform. 2015;7(1):20.
- 16. Sharma A, Lal SP. Tanimoto based similarity measure for intrusion detection system. J Inf Secur. 2011;2(04):195.
- 17. Compound Similarity. <https://chemminetools.ucr.edu/similarity>. Accessed 14 Feb 2022.
- 18. Agoram B, Woltosz WS, Bolger MB. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Adv Drug Deliv Rev. 2001;50(Suppl 1):S41-67.
- 19. Franceschi L, Faggiani A, Furlanut M. A simple method to monitor serum concentrations of fluoxetine and its major metabolite for pharmacokinetic studies. J Pharm Biomed Anal. 2009;49(2):554–7.
- 20. D'Souza DL, Dimmitt DC, Robbins DK, Nezamis J, Simms L, Koch KM. Effect of alosetron on the pharmacokinetics of fluoxetine. J Clin Pharmacol. 2001;41(4):455–8.
- 21. WebPlotDigitizer. <https://automeris.io/WebPlotDigitizer/>. Accessed 14 Feb 2022.
- 22. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci. 2003;22(1):69–77.
- 23. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? GMDD. 2014;7(1):1525–34.
- 24. Laizure SC, Meibohm B, Nelson K, Chen F, Hu ZY, Parker RB. Comparison of cafeine disposition following administration by oral solution (energy drink) and inspired powder (AeroShot) in human subjects. Br J Clin Pharmacol. 2017;83(12):2687–94.
- 25. Qiu F, Wang G, Zhao Y, Sun H, Mao GAJ, Sun J. Efect of danshen extract on pharmacokinetics of theophylline in healthy volunteers. Br J Clin Pharmacol. 2008;65(2):270–4.
- 26. Sugimoto K, Ohmori M, Tsuruoka S, Nishiki K, Kawaguchi A, Harada K, Arakawa M, Sakamoto K, Masada M, Miyamori I, Fujimura A. Diferent efects of St John's wort on the pharmacokinetics of simvastatin and pravastatin. Clin Pharmacol Ther. 2001;70(6):518–24.
- 27. Chalon SA, Desager JP, Desante KA, Frye RF, Witcher J, Long AJ, Sauer JM, Golnez JL, Smith BP, Thomasson HR, Horsmans Y. Efect of hepatic impairment on the pharmacokinetics of atomoxetine and its metabolites. Clin Pharmacol Ther. 2003;73(3):178–91.
- 28. Nelson E, Powell JR, Conrad K, Likes K, Byers J, Baker S, Perrier D. Phenobarbital pharmacokinetics and bioavailability in adults. J Clin Pharmacol. 1982;22(2–3):141–8.
- 29. Wedlund PJ, Aslanian WS, Jacqz E, McAllister CB, Branch RA, Wilkinson GR. Phenotypic differences in mephenytoin pharmacokinetics in normal subjects. J Pharmacol Exp Ther. 1985;234(3):662–9.
- 30. Greenblatt DJ, Harmatz JS, von Moltke LL, Wright CE, Durol AL, Harrel-Joseph LM, Shader RI. Comparative kinetics and response to the benzodiazepine agonists triazolam and zolpidem: evaluation of sex-dependent diferences. J Pharmacol Exp Ther. 2000;293(2):435–43.
- 31. Otani K, Yasui N, Furukori H, Kaneko S, Tasaki H, Ohkubo T, Nagasaki T, Sugawara K, Hayashi K. Relationship between single oral dose pharmacokinetics of alprazolam and triazolam. Int Clin Psychopharmacol. 1997;12(3):153–7.
- 32. Link B, Haschke M, Grignaschi N, Bodmer M, Aschmann YZ, Wenk M, Krähenbühl S. Pharmacokinetics of intravenous and oral midazolam in plasma and saliva in humans: usefulness of saliva as matrix for CYP3A phenotyping. Br J Clin Pharmacol. 2008;66(4):473–84.
- 33. Chen X, Jacobs G, de Kam M, Jaeger J, Lappalainen J, Maruf P, Smith MA, Cross AJ, Cohen A, Van Gerven J. The central nervous system effects of the partial GABA-A α 2, 3-selective receptor modulator AZD7325 in comparison with lorazepam in healthy males. Br J Clin Pharmacol. 2014;78(6):1298–314.
- 34. East T, Dye D. Determination of dextromethorphan and metabolites in human plasma and urine by high-performance liquid chromatography with fluorescence detection. J Chromatogr. 1985;338(1):99–112.
- 35. Connarn JN, Flowers S, Kelly M, Luo R, Ward KM, Harrington G, Moncion I, Kamali M, McInnis M, Feng MR, Ellingrod V, Babiskin A, Zhang X, Sun D. Pharmacokinetics and pharmacogenomics of bupropion in three diferent formulations with diferent release kinetics in healthy human volunteers. AAPS J. 2017;19(5):1513–22.
- 36. Pringle TH, Francis RJ, East PB, Shanks RG. Pharmacodynamic and pharmacokinetic studies on bufuralol in man. Br J Clin Pharmacol. 1986;22(5):527–34.
- 37. Kolb KW, Garnett WR, Small RE, Vetrovec GW, Kline BJ, Fox T. Efect of cimetidine on quinidine clearance. Ther Drug Monit. 1984;6(3):306–12.
- 38. Ciraulo DA, Barnhill JG, Jafe JH. Clinical pharmacokinetics of imipramine and desipramine in alcoholics and normal volunteers. Clin Pharmacol Ther. 1988;43(5):509–18.
- 39. Madani S, Barilla D, Cramer J, Wang Y, Paul C. Efect of terbinafine on the pharmacokinetics and pharmacodynamics of desipramine in healthy volunteers identified as cytochrome P450 2D6 (CYP2D6) extensive metabolizers. J Clin Pharmacol. 2002;42(11):1211–8.

Authors and Afliations

Jingchen Zhai1 [·](http://orcid.org/0000-0003-2691-8867) Beihong Ji1 [·](http://orcid.org/0000-0003-0387-4056) Shuhan Liu1 [·](http://orcid.org/0000-0002-6627-4843) Yuzhao Zhang1 · Lianjin Cai¹ · Junmei Wang[1](http://orcid.org/0000-0002-9607-8229)

Jingchen Zhai jiz183@pitt.edu

Beihong Ji bej22@pitt.edu

Shuhan Liu shl169@pitt.edu Yuzhao Zhang yuz170@pitt.edu

¹ Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening Center, School of Pharmacy, University of Pittsburgh, 06407 Salk Hall, Terrace Street, Pittsburgh, PA 15261, USA