Biol. Pharm. Bull.

Regular Article

Machine Learning Prediction of the Three Main Input Parameters of a Simplified Physiologically Based Pharmacokinetic Model Subsequently Used to Generate Time-Dependent Plasma Concentration Data in Humans after Oral Doses of 212 Disparate Chemicals

Yusuke Kamiya,^a Kentaro Handa,^b Tomonori Miura,^a Junya Ohori,^b Airi Kato,^a Makiko Shimizu,^a Masato Kitajima,^b and Hiroshi Yamazaki^{a,*}

^a*Showa Pharmaceutical University, Machida, Tokyo 194-8543, Japan, and* ^b*Fujitsu, Nakahara-ku, Kawasaki 211-8588, Japan*

***Corresponding author:**

Hiroshi Yamazaki, Ph.D.

Laboratory of Drug Metabolism and Pharmacokinetics, Showa Pharmaceutical University, 3-3165 Higashi-tamagawa Gakuen, Machida, Tokyo 194-8543, Japan.
E-mail address: hyamazak@ac.shoyaku.ac.jp

Abstract

Physiologically based pharmacokinetic (PBPK) modeling has the potential to play significant roles in estimating internal chemical exposures. The three major PBPK model input parameters (i.e., absorption rate constants, volumes of the systemic circulation, and hepatic intrinsic clearances) were generated *in silico* for 212 chemicals using machine learning algorithms. These input parameters were calculated based on sets of between 17 and 65 chemical properties that were generated by *in silico* prediction tools before being processed by machine learning algorithms. The resulting simplified PBPK models were used to estimate plasma concentrations after virtual oral administrations in humans. The estimated absorption rate constants, volumes of the systemic circulation, and hepatic intrinsic clearance values for the 212 test compounds determined traditionally (i.e., based on fitting to measured concentration profiles) and newly estimated had correlation coefficients of 0.65, 0.68, and 0.77 ($p < 0.01$, $n = 212$), respectively. When human plasma concentrations were modeled using traditionally determined input parameters and again using *in silico* estimated input parameters, the two sets of maximum plasma concentrations ($r = 0.85$, $p < 0.01$, $n = 212$) and areas under the curve ($r = 0.80$, $p < 0.01$, $n = 212$) were correlated. Virtual chemical exposure levels in liver and kidney were also estimated using these simplified PBPK models along with human plasma levels. These results indicate that the PBPK model input parameters for humans of a diverse set of compounds can be reliability estimated using chemical descriptors calculated using *in silico* tools.

Keywords: PBPK modeling; pharmacokinetics; general chemicals, food components

Introduction

Therapeutic drug monitoring of steady-state plasma concentrations at designated intervals in individual patients for specific medicines, such as antipsychotics drugs, is an accepted clinical practice to maintain drug concentrations in the target ranges.^{1,2)} Such monitoring methods could be supported by pharmacokinetic modeling and simulations to address critical questions related to patient care.³⁾ Full physiologically based pharmacokinetic (PBPK) models⁴⁾ can predict drug monitoring results in patients⁵⁻⁷⁾ based on the physiological and anatomical properties of the relevant organ and the chemical properties of the modeled substance.^{8,9)} Simplified PBPK models, which are easier to handle than full models, have been developed for use by a wide range of paramedical staff for predicting drug monitoring results and have been applied in some nonfatal overdose cases.¹⁰⁻¹⁴⁾ Such models can predict plasma concentration curves and can quickly determine whether time-dependent treatments with gastric lavage and/or activated charcoal are likely to be beneficial in real-time emergency situations. Rapid treatment with gastric lavage is generally indicated in emergency rooms if it can be delivered within 60 minutes of ingestion¹⁵⁾; however, this does not take into account the specific pharmacokinetics of the ingested substance, and in some cases, gastric lavage might be beneficial even if delivered outside the 60-min window.

Simplified PBPK models are useful not only in the above-mentioned fields of therapeutic drug monitoring and drug poisoning, but also during drug discovery and chemical risk assessment.¹⁶⁾ Although the toxicological profiles of a variety of chemicals to which humans are exposed have been investigated in rodents, new alternative methods to evaluate compound safety are being developed worldwide using *in vitro* or *in silico* approaches.^{17,18)} PBPK modeling, which is known as PBK modeling in the European Union,¹⁹⁾ has the potential to

play a significant role in reducing and replacing the use of animals for estimating toxicokinetics or internal exposures.²⁰⁾

The input parameters for PBPK models are generally based on *in vivo* data; however, to investigate the feasibility of making such modeling more accessible in real-time, we generated PBPK model input parameters *in silico* for a broad range of chemicals using a machine learning algorithm without reference to *in vivo* studies.^{21,22)} The values of the three major PBPK model input parameters, i.e., the absorption rate constant (k_a), the volume of the systemic circulation (V_1), and the hepatic intrinsic clearance ($CL_{h,int}$), for this selection of chemicals were previously generated *in silico* for rats.²¹⁾ Therefore, the aim of the present study was to establish *in silico* models for predicting k_a , V_1 , and $CL_{h,int}$ values (for use in human PBPK models) based on a number of chemical descriptors. These estimated input parameters for human PBPK models were then used to generate the plasma and tissue concentrations of 212 disparate chemicals after virtual oral administrations of 1.0-mg/kg doses.

Materials and methods

The pharmacokinetic data sets to which machine learning was applied in this study covered 212 chemicals; these were mainly taken from our literature survey of reported blood concentrations versus time data for oral administration in humans (180 chemicals) and included 46 medicines from a new Japanese drug database (**Supplemental Table S1**). The details for setting up simplified PBPK models consisting of gut, liver, kidney, and central compartments have been described previously^{10,11,13)} and are briefly outlined in the **Supplemental materials and methods**. The necessary input parameters, i.e., k_a , V_1 , and

$CL_{h,int}$, for PBPK models are conventionally computed to achieve the best fit to reported human plasma concentrations in a similar manner to that done previously for rats (which included 11 of the current test chemicals) and humanized-liver mice (which included 21 of the current test chemicals).²¹⁾ Human input parameters for these 32 chemicals were scaled up from the previous animal PBPK model parameters as previously described.²³⁻²⁸⁾ The criterion for acceptable predictive abilities that was previously applied for rat PBPK models was a threefold error for reported/observed values for plasma C_{max} and AUC after virtual administrations of oral doses of 1.0 mg/kg. Human values for k_a , V_1 , and $CL_{h,int}$ for another six chemicals were estimated in this study by employing a scale-up strategy to humans from preliminary humanized-liver mice data shown in **Supplemental Fig. S1**, as previously described.^{25,26)}

For each substance, the chemical descriptor sets of 1710 parameters describing chemical structural and physicochemical properties were obtained using open-source *in silico* programs (such as Rdkit and Mordred) as described previously.²¹⁾ Before the prediction process, some chemical descriptors were auto-scaled to account for missing data by supplementation with median values. To establish prediction systems based on the *in silico*-derived values, machine learning algorithms were applied that utilized ridge regression²⁹⁾ for k_a and V_1 and LightGBM³⁰⁾ for $CL_{h,int}$. Estimation models for k_a , V_1 , and $CL_{h,int}$ were extensively validated with a modeling approach designed to integrate nested cross-validation.²¹⁾ Average absolute fold errors for k_a , V_1 , and $CL_{h,int}$ values were calculated as described previously.³¹⁾

Results and Discussion

In the medical field, applications of machine learning to drug discovery or repositioning have been focused.³²⁾ Automated covariate selection in pharmacometrics model building has been reported.³³⁾ On the other hand, humans are generally exposed to unexpected general chemicals or food components from the environments. To illustrate their broad diversity in chemical properties, the current 212 compounds (**Supplemental Table S1**) underwent projection onto our previously proposed two-dimensional chemical space consisting of 25 partitions^{16,21,34-37)} (**Figure 1**). The newly tested 104 chemicals among the total of 212 compounds also demonstrated broad diversity. The relationships between the traditionally determined human PBPK model input values (hereafter referred to as *in vivo*-derived) and the values estimated using our proposed new computational approach (*in silico*-derived) for the 212 chemicals included in the current study (**Supplemental Tables S1 and S2**) are shown in **Figure 2**. The correlation coefficients of 0.65, 0.68, and 0.77 ($p < 0.01$, $n = 212$), respectively, for the k_a , V_1 , and $CL_{h,int}$ values, were statistically significant for the current 212 chemicals. Average absolute fold errors for the k_a , V_1 , and $CL_{h,int}$ values were 1.89, 2.02, and 3.28, respectively. These results indicated that the k_a , V_1 , and $CL_{h,int}$ values in humans for a diverse set of chemicals could be reliably estimated using between 17 and 65 chemical descriptors (**Supplemental Table S2**) that were calculated using *in silico* tools.

In the medical field, direct estimation of drug concentrations of the typical single key medicine like cyclosporin A in body fluids has been reportedly proposed.³⁸⁾ On the other hand, the three input PBPK parameters were used to generate time-series kinetic data in this study because direct estimation for a variety of chemicals is not realistic in non-medical research field. The human plasma concentrations of the current 212 chemicals were generated using PBPK models using two sets of input parameters: either traditional *in vivo*-derived values^{16,36)}

or *in silico*-derived values. When the traditionally determined PBPK model input values for the 212 chemicals were replaced with the *in silico*-estimated values of k_a , V_1 , and $CL_{h,int}$ (as shown in **Figure 2** and **Supplemental Table S2**), the output C_{max} values in human plasma generated using the two sets of input parameters were well correlated (correlation coefficient, $r = 0.85$, $p < 0.01$, **Figure 3A** and **Supplemental Table S3**). Moreover, there was a good correlation between the AUC values in human plasma of the 212 chemicals obtained using PBPK models (**Figure 3B** and **Supplemental Table S3**) with empirically determined and *in silico*-estimated input parameters ($r = 0.80$, $p < 0.01$). In this way, the applicability to human pharmacokinetics of these estimated input parameters for PBPK modeling was demonstrated.

In rats, the proposed PBPK models were able to simulate concentrations in rat livers and kidneys in a similar manner as rat plasma; the observed hepatic concentrations were confirmed with limited numbers of chemicals.^{26,28} The current human PBPK models were also able to output the virtual concentrations in human tissue in a similar manner. The concentrations of 212 chemicals in human liver and kidney generated by PBPK models were also evaluated using *in vivo*- and *in silico*-derived sets of input parameters. When the traditionally determined PBPK model input values for 212 chemicals were replaced with the *in silico*-estimated values of k_a , V_1 , and $CL_{h,int}$ (as shown in **Figure 2** and **Supplemental Table S2**), the output C_{max} and AUC values in human liver using the two sets of input parameters were well correlated ($r = 0.89$ and 0.83 for C_{max} and AUC, respectively, $p < 0.01$, **Supplemental Figure S2A and S2B** and **Supplemental Table S3**). There were also good correlations for C_{max} and AUC values in human kidney of the 212 chemicals between those obtained using PBPK models (**Supplemental Figure S2C and S2D** and **Supplemental Table S3**) with empirically determined and *in silico*-derived input parameters, i.e., $r = 0.87$ and 0.84 ($p < 0.01$) for C_{max} and AUC, respectively.

Recently we developed a computational method to predict the apparent membrane permeability values (both influx and efflux) as determined in an *in vitro* pH-dependent monolayer cell system for ~200 diverse compounds. These membrane permeability values are essentially related to absorption processes *in vivo*, and the prediction method primarily used trivariate linear regression with an additional machine learning approach.³⁷⁾ This trivariate regression analysis used three common physicochemical properties (reflecting the experimental apical and basal pH conditions used in the assays and the molecular weight of the substance) in combination to predict the bidirectional transport parameters. The prediction accuracy was then enhanced by applying a light gradient boosting machine learning system to estimate influx and efflux apparent permeability values by incorporating 17 and 19 less common *in silico* chemical descriptors, respectively.³⁷⁾ Although the selected chemical descriptors should fundamentally reflect and be related to chemical solubility/dissolution and/or intestinal permeability in relation to the absorption rate constants in the current study, the physicochemical properties relevant to the overall absorption of compounds may contribute in a diverse manner to these machine learning approaches. Despite the previous chemical descriptors for rat $CL_{h,int}$ model applying a light gradient boosting machine learning system²¹⁾ could be commonly applicable to the current human $CL_{h,int}$ model, human k_a and V_1 models utilized ridge regression with chemical descriptors were differently set up from the previous rat k_a and V_1 models (data not shown).

In conclusion, the concentrations of 212 chemicals in plasma, liver, and kidney after virtual oral doses in humans could be reliably estimated using simplified PBPK models with input parameters calculated from sets of between 17 and 65 chemical properties that were generated using *in silico* prediction tools. Apparently, an increasing number of datasets for non-medical chemicals with pharmacokinetic data is one of determinant factors for updating the established *in silico* models for three input human PBPK parameter values. It should be noted

that the virtual plasma $C_{\max}$ values of chemicals could represent an inverse marker for rat hematotoxicity (as reflected in the lowest-observed-effect levels),³⁶⁾ and the *in silico* predicted values for permeability across intestinal cells could represent a putative indicator for hepatotoxicity (as reflected in the no-observed-effect levels).³⁵⁾ On the other hand, areas under the curve of unmetabolized (remaining) dimethylaniline derivatives estimated using rat pharmacokinetic models showed an apparently positive correlation with the reported lowest-observed-effect levels for haematotoxicity of these chemicals such as 3,5-dimethylaniline and 2,6-dimethylaniline.³⁹⁾ Consequently, the above-described computational methods represent new alternative approaches that could contribute to chemical safety screening. This approach to human PBPK modeling, with the availability of *in vitro* and *in silico* alternatives for generating pharmacokinetic properties, contributes to the effectiveness of computational toxicology for assessing the potential risk of industrial chemicals and/or food components. The simplified PBPK models described here, which are easy to implement, could be applied for predicting a variety of drug exposures by a wide range of paramedical staff, industry researchers, and regulatory authorities.

Acknowledgments

The authors thank Drs. Kimito Funatsu, Fumiaki Shono, Masaya Fujii, Jun Tomizawa, Wataru Kobari, Tasuku Sato, Masayoshi Utsumi, Shohei Otsuka, and Norie Murayama for their assistance and David Smallbones for copyediting a draft of this article. This work was supported in part by the METI Artificial Intelligence-based Substance Hazard Integrated Prediction System Project, Japan. T.M. was partly supported by the Japan Society for the Promotion of Science Grant-in-Aid for Young Scientists 202021210.

Conflict of interest

The authors declare no conflict of interest.

Supplementary Materials

The online version of this article contains Supplementary Materials.

References

- 1) Ereshefsky L. Pharmacokinetics and drug interactions: update for new antipsychotics. *J Clin Psychiatry*, **57 Suppl 11**, 12-25 (1996).
- 2) Zhou SF. Polymorphism of human cytochrome P450 2D6 and its clinical significance: part II. *Clin Pharmacokinet*, **48**, 761-804 (2009).
- 3) Huang SM, Rowland M. The role of physiologically based pharmacokinetic modeling in regulatory review. *Clin Pharmacol Ther*, **91**, 542-549 (2012).
- 4) Jamei M, Marciniak S, Feng K, Barnett A, Tucker G, Rostami-Hodjegan A. The Simcyp population-based ADME simulator. *Expert Opin Drug Metab Toxicol*, **5**, 211-223 (2009).
- 5) Ota M, Shimizu M, Kamiya Y, Emoto C, Fukuda T, Yamazaki H. Adult and infant pharmacokinetic profiling of dihydrocodeine using physiologically based pharmacokinetic modeling. *Biopharm Drug Dispos*, **40**, 350-357 (2019).
- 6) Notsu Y, Shimizu M, Sasaki T, Nakano A, Ota M, Yoshida S, Yamazaki H. Simple pharmacokinetic models accounting for drug monitoring results of atomoxetine and its 4-hydroxylated metabolites in Japanese pediatric patients genotyped for cytochrome P450 2D6. *Drug Metab Pharmacokinet*, **35**, 191-200 (2020).
- 7) Emoto C, Shimizu M, Tanaka T, Yamazaki H. Feasibility of physiologically based pharmacokinetic simulations for assessing pediatric patients after accidental drug ingestion: A case study of a 1.4-year-old girl who ingested alprazolam. *Drug Metab Pharmacokinet*, **39**, 100394 (2021).
- 8) Wambaugh JF, Hughes MF, Ring CL, MacMillan DK, Ford J, Fennell TR, Black SR, Snyder RW, Sipes NS, Wetmore BA, Westerhout J, Setzer RW, Pearce RG, Simmons JE, Thomas RS. Evaluating In Vitro-In Vivo Extrapolation of Toxicokinetics. *Toxicol Sci*, **163**, 152-169 (2018).
- 9) Sayre RR, Wambaugh JF, Grulke CM. Database of pharmacokinetic time-series data and parameters for 144 environmental chemicals. *Sci Data*, **7**, 122 (2020).
- 10) Adachi K, Beppu S, Nishiyama K, Shimizu M, Yamazaki H. Pharmacokinetics of duloxetine self-administered in overdose with quetiapine and other antipsychotic drugs in a Japanese patient admitted to hospital. *J Pharm Health Care Sci*, **7**, 6 (2021).
- 11) Adachi K, Beppu S, Terashima M, Fukuda T, Tomizawa J, Shimizu M, Yamazaki H. Pharmacokinetics of caffeine self-administered in overdose in a Japanese patient admitted to hospital. *J Pharm Health Care Sci*, **7**, 36 (2021).
- 12) Adachi K, Beppu S, Terashima M, Kobari W, Shimizu M, Yamazaki H. Pharmacokinetic modeling of over-the-counter drug diphenhydramine self-administered in overdoses in Japanese patients admitted to hospital. *J Pharm Health Care Sci*, **7**, 32 (2021).

- 13) Adachi K, Sugitani Y, Unita R, Yoshida K, Beppu S, Terashima M, Fujii M, Shimizu M, Yamazaki H. Pharmacokinetics of loxoprofen in a self-administered overdose in a Japanese patient admitted to hospital. *J Pharm Health Care Sci*, **7**, 33 (2021).
- 14) Adachi K, Tuchiya J, Beppu S, Nishiyama K, Shimizu M, Yamazaki H. Pharmacokinetics of anticoagulant edoxaban in overdose in a Japanese patient transported to hospital. *J Pharm Health Care Sci*, **6**, 20 (2020).
- 15) Vale JA. Position statement: gastric lavage. American Academy of Clinical Toxicology; European Association of Poisons Centres and Clinical Toxicologists. *J Toxicol Clin Toxicol*, **35**, 711-719 (1997).
- 16) Kamiya Y, Otsuka S, Miura T, Takaku H, Yamada R, Nakazato M, Nakamura H, Mizuno S, Shono F, Funatsu K, Yamazaki H. Plasma and hepatic concentrations of chemicals after virtual oral administrations extrapolated using rat plasma data and simple physiologically based pharmacokinetic models. *Chem Res Toxicol*, **32**, 211-218 (2019).
- 17) Rovida C, Barton-Maclaren T, Benfenati E, Caloni F, Chandrasekera PC, Chesne C, Cronin MTD, De Knecht J, Dietrich DR, Escher SE, Fitzpatrick S, Flannery B, Herzler M, Hougaard Bennekou S, Hubesch B, Kamp H, Kisitu J, Kleinstreuer N, Kovarich S, Leist M, Maertens A, Nugent K, Pallocca G, Pastor M, Patlewicz G, Pavan M, Presgrave O, Smirnova L, Schwarz M, Yamada T, Hartung T. Internationalization of read-across as a validated new approach method (NAM) for regulatory toxicology. *ALTEX*, **37**, 579-606 (2020).
- 18) Costin GE, Hill E, Brown J, Clippinger AJ. Qualification of a non-animal vaginal irritation method admitted as nonclinical assessment model (NAM) in the Incubator Phase of the United States Food and Drug Administration (US FDA) Medical Devices Development Tool (MDDT). *Toxicol In Vitro*, **62**, 104680 (2020).
- 19) Paini A, Leonard JA, Joossens E, Bessems JGM, Desalegn A, Dorne JL, Gosling JP, Heringa MB, Klaric M, Kliment T, Kramer NI, Loizou G, Louise J, Lumen A, Madden JC, Patterson EA, Proença S, Punt A, Setzer RW, Suciú N, Troutman J, Yoon M, Worth A, Tan YM. Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making. *Computational Toxicology*, **9**, 61-72 (2019).
- 20) Bell SM, Chang X, Wambaugh JF, Allen DG, Bartels M, Brouwer KLR, Casey WM, Choksi N, Ferguson SS, Fraczkievicz G, Jarabek AM, Ke A, Lumen A, Lynn SG, Paini A, Price PS, Ring C, Simon TW, Sipes NS, Sprankle CS, Strickland J, Troutman J, Wetmore BA, Kleinstreuer NC. In vitro to in vivo extrapolation for high throughput prioritization and decision making. *Toxicol In Vitro*, **47**, 213-227 (2018).
- 21) Kamiya Y, Handa K, Miura T, Yanagi M, Shigeta K, Hina S, Shimizu M, Kitajima M, Shono F, Funatsu K, Yamazaki H. *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**, 507-513 (2021).

- 22) Yanagi M, Kamiya Y, Murayama N, Banju K, Shimizu M, Yamazaki H. Metabolic profiles for the pyrrolizidine alkaloid neopetasitenine and its metabolite petasitenine in humans extrapolated from rat in vivo and in vitro data sets using a simplified physiologically based pharmacokinetic model. *J Toxicol Sci*, **46**, 391-399 (2021).
- 23) Miura T, Uehara S, Mizuno S, Yoshizawa M, Murayama N, Kamiya Y, Shimizu M, Suemizu H, Yamazaki H. Steady-State Human Pharmacokinetics of Monobutyl Phthalate Predicted by Physiologically Based Pharmacokinetic Modeling Using Single-Dose Data from Humanized-Liver Mice Orally Administered with Dibutyl Phthalate. *Chem Res Toxicol*, **32**, 333-340 (2019).
- 24) Shimizu M, Suemizu H, Mitsui M, Shibata N, Guengerich FP, Yamazaki H. Metabolic profiles of pomalidomide in human plasma simulated with pharmacokinetic data in control and humanized-liver mice. *Xenobiotica*, **47**, 844-848 (2017).
- 25) Miura T, Kamiya Y, Uehara S, Murayama N, Shimizu M, Suemizu H, Yamazaki H. Hepatotoxicological potential of *p*-toluic acid in humanized-liver mice investigated using simplified physiologically based pharmacokinetic models. *Xenobiotica*, **51**, 636-642 (2021).
- 26) Miura T, Uehara S, Shigeta K, Yoshizawa M, Kamiya Y, Murayama N, Shimizu M, Suemizu H, Yamazaki H. Metabolic Profiles of Tetrabromobisphenol A in Humans Extrapolated from Humanized-Liver Mouse Data Using a Simplified Physiologically Based Pharmacokinetic Model. *Chem Res Toxicol*, **34**, 522-528 (2021).
- 27) Shimizu M, Suemizu H, Mizuno S, Kusama T, Miura T, Uehara S, Yamazaki H. Human plasma concentrations of trimethylamine N-oxide extrapolated using pharmacokinetic modeling based on metabolic profiles of deuterium-labeled trimethylamine in humanized-liver mice. *J Toxicol Sci*, **43**, 387-393 (2018).
- 28) Miura T, Shimizu M, Uehara S, Yoshizawa M, Nakano A, Yanagi M, Kamiya Y, Murayama N, Suemizu H, Yamazaki H. Different Hepatic Concentrations of Bromobenzene, 1,2-Dibromobenzene, and 1,4-Dibromobenzene in Humanized-Liver Mice Predicted Using Simplified Physiologically Based Pharmacokinetic Models as Putative Markers of Toxicological Potential. *Chem Res Toxicol*, **33**, 3048-3053 (2020).
- 29) Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, **12**, 55-67 (1970).
- 30) Zhang J, Mucs D, Norinder U, Svensson F. LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity-Application to the Tox21 and Mutagenicity Data Sets. *J Chem Inf Model*, **59**, 4150-4158 (2019).
- 31) Obach RS, Baxter JG, Liston TE, Silber BM, Jones BC, MacIntyre F, Rance DJ, Wastall P. The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data. *J Pharmacol Exp Ther*, **283**, 46-58 (1997).
- 32) Talevi A, Morales JF, Hather G, Podichetty JT, Kim S, Bloomingdale PC, Kim S, Burton J, Brown JD, Winterstein AG, Schmidt S, White JK, Conrado DJ. Machine

Learning in Drug Discovery and Development Part 1: A Primer. *CPT Pharmacometrics Syst Pharmacol*, **9**, 129-142 (2020).

- 33) Yamashita F, Fujita A, Sasa Y, Higuchi Y, Tsuda M, Hashida M. An Evolutionary Search Algorithm for Covariate Models in Population Pharmacokinetic Analysis. *J Pharm Sci*, **106**, 2407-2411 (2017).
- 34) Yamazaki H, Kamiya Y. Extrapolation of Hepatic Concentrations of Industrial Chemicals Using Pharmacokinetic Models to Predict Hepatotoxicity. *Toxicol Res*, **35**, 295-301 (2019).
- 35) Kamiya Y, Takaku H, Yamada R, Akase C, Abe Y, Sekiguchi Y, Murayama N, Shimizu M, Kitajima M, Shono F, Funatsu K, Yamazaki H. Determination and prediction of permeability across intestinal epithelial cell monolayer of a diverse range of industrial chemicals/drugs for estimation of oral absorption as a putative marker of hepatotoxicity. *Toxicol Rep*, **7**, 149-154 (2020).
- 36) Kamiya Y, Otsuka S, Miura T, Yoshizawa M, Nakano A, Iwasaki M, Kobayashi Y, Shimizu M, Kitajima M, Shono F, Funatsu K, Yamazaki H. Physiologically Based Pharmacokinetic Models Predicting Renal and Hepatic Concentrations of Industrial Chemicals after Virtual Oral Doses in Rats. *Chem Res Toxicol*, **33**, 1736-1751 (2020).
- 37) Kamiya Y, Omura A, Hayasaka R, Saito R, Sano I, Handa K, Ohori J, Kitajima M, Shono F, Funatsu K, Yamazaki H. Prediction of permeability across intestinal cell monolayers for 219 disparate chemicals using in vitro experimental coefficients in a pH gradient system and in silico analyses by trivariate linear regressions and machine learning. *Biochem Pharmacol*, **192**, 114749 (2021).
- 38) Ogami C, Tsuji Y, Seki H, Kawano H, To H, Matsumoto Y, Hosono H. An artificial neural network-pharmacokinetic model and its interpretation using Shapley additive explanations. *CPT Pharmacometrics Syst Pharmacol*, **10**, 760-768 (2021).
- 39) Miura T, Kamiya Y, Murayama N, Shimizu M, Yamazaki H. Differences in pharmacokinetics and haematotoxicities of aniline and its dimethyl derivatives orally administered in rats. *Biol Pharm Bull*, in press (2021).

Figure 1. Coordinate values of 212 compounds in a two-dimensional plane with 25 partitions illustrating variety in the chemical space. The values for 104 newly evaluated compounds (gray circles) and 108 previously evaluated compounds (open circles) are shown. The proposed evaluation of the chemical space^{16,21,34-37)} was conducted for the 108 previously examined chemicals in studies on intestinal permeability^{35,37)} and rat pharmacokinetics.²¹⁾

Figure 2. Correlations between determined (*in vivo*-derived) and estimated (*in silico*-derived) PBPK model input parameter values k_a (A), V_1 (B), and $\text{CL}_{h,\text{int}}$ (C) for 212 chemicals. Solid and dotted/dashed lines indicate equivalence and two- and tenfold ranges, respectively.

Figure 3. Correlations of $C_{\max}$ (A) and AUC (B) values for 212 chemicals in human plasma generated by PBPK models using traditionally determined (*in vivo*-derived) and currently estimated (*in silico*-derived) values of the three PBPK model input parameters. Solid and dotted/dashed lines indicate equivalence and two- and tenfold ranges, respectively.