Application of mechanistic multiparameter optimization and large scale in vitro to in vivo pharmacokinetics correlations to small molecule therapeutic projects

Fabio Broccatelli[1,], Vijayabhaskar Veeravalli[1], Daniel Cashion[1], Javier L. Baylon[1], Franco Lombardo[2] and Lei Jia[1,]**

[1] Bristol-Myers Squibb Company, San Diego, CA 92121

[2] CmaxDMPK LLC, Framingham, MA, 01701 US

**Corresponding authors*

Fabio Broccatelli: Fabio.Broccatelli@bms.com, Lei Jia: leijiachem@gmail.com

Keywords: Multiple Parameter Optimization; Pharmacokinetics; Lead Optimization; Machine Learning.

ABSTRACT

Computational chemistry and machine learning are used in drug discovery to predict target-specific and pharmacokinetic properties of molecules. Multiparameter optimization (MPO) functions are used to summarize multiple properties into a single score, aiding compound prioritization. However, over-reliance on subjective MPO functions risks reinforcing human bias. Mechanistic modeling approaches based on physiological relevance can be adapted to meet different potential key objectives of the project (e.g. minimizing dose, maximizing safety margins and/or minimized drug-drug interaction risk) while retaining the same underlying model structure. The current work incorporates recent approaches to predict *in vivo* PK properties and validates *in vitro* to *in vivo* correlation analysis to support mechanistic PK MPO. Examples of use and impact in small molecule drug discovery projects are provided. Overall the mechanistic MPO identifies 83% of the compounds considered as short-listed for clinical experiments in the top 2nd percentile, and 100% in the top 10th percentile, resulting in a AUC>0.95. In addition, the MPO score successfully recapitulates the chronological progression of the optimization process across different scaffolds. Finally, the MPO scores for compounds characterized in pharmacokinetics experiments are markedly higher compared to the rest of the compounds synthesized, highlighting the potential of this tool to reduce the reliance on *in vivo* testing for compound screening.

INTRODUCTION

Predictive modeling-driven small molecule drug discovery relies on computational chemistry and machine learning (ML) approaches to predict target-specific and pharmacokinetic (PK) properties of new molecular entities (NME).[1] In addition, physiologically based PK (PBPK) approaches are utilized to mechanistically predict absorption, distribution, metabolism, elimination (ADME), therapeutic dose, and drug-drug interaction (DDI) potential of NMEs.[2]

Multiparameter optimization (MPO) functions are used in compound progression and design to summarize multiple in vitro and in silico properties (e.g. potency, clearance, permeability, solubility) into a single score.[3] Without computer-aided analysis, it is practically impossible to readily and accurately balance more than four variables.[4] While the benefits of reducing complexity to inform decision making with MPO functions are evident, over-reliance on subjective MPO functions may ultimately reinforce human bias.[5]

Accurate predictions of in vivo PK and pharmacodynamic (PD) properties are required to inform clinical candidate selection. Although the approaches to predict in vivo PK/PD properties of drug candidates may vary from user to user, there is a consensus in the industry around the use of mechanistic modeling approaches.[6] Unlike empirical MPO functions, the mechanistic approaches are based on physiological relevance, can be tested, and translated across species, and can support estimation of safety margins and drug-drug interaction (DDI) risk. In addition, the interpretability of mechanistic models is a further key differentiation from “black box” methods; drug designers can leverage the model structure and the observed in vitro to in vivo correlations to identify opportunities to design better molecules. In this work, we demonstrate the value of using mechanistic PK models as an MPO scoring function for compound testing and progression [7-9]. We incorporate and extensively validate recent approaches to the prediction of in vivo PK properties, provide adequate examples of in vitro to in vivo correlation (IVIVc)

analysis in support of MPO, and illustrate some examples of use and impact in small molecule drug discovery projects.

METHODS

PKPD Assumptions

The choice of MPO method depends on the known or assumed relationship between PK and PD (PK/PD). Quantitative PK/PD models are valuable during clinical candidate selection for dose projection and DDI assessment, especially for targets and modalities (such as covalent drugs or targeted protein degraders) resulting in delayed responses (indirect response models).[10] In early phases, PK/PD relationship can be simplified at the expense of accuracy to fit into three response categories: C_{max}, C_{min}, and C_{avg} driven efficacy. When PK/PD relationships are simplified as C_{max} driven, the objective of dose optimization is to achieve a transient exposure higher than the minimum efficacious concentration (MEC). In this case, high potency, high bioavailability (F), fast absorption, moderate clearance (CL), and short half-life ($T_{1/2}$) are desirable. However, C_{max} is less commonly associated with on-target activities and often associated with off-target and safety-related concerns. When PK/PD relationships are simplified as C_{min} driven efficacy, the efficacious dose is determined based on the terminal concentration at steady state, matching the MEC. In this case high potency, high F, low CL, and prolonged oral $T_{1/2}$ are desirable, which is a reasonable default assumption for competitive inhibitors. Lastly, with PK/PD relationships simplified as C_{avg} driven efficacy, the efficacious dose is determined based on the average concentration at steady state (that is, exposure over-dosing interval) matching the MEC. For C_{avg} driven efficacy, high potency, high F, and low CL are desirable, while $T_{1/2}$ is believed to be less important. Notably, a C_{avg} that is at least 5-fold higher than the total EC_{50} predicts efficacious dose for over 90% of the 164 drugs studied by Jansson-Löfmark et. Al.[11]

Prediction of PK

All the non-proprietary datasets used for model validation and calibration in this work are available as supporting information. This study used physiologically based methods to predict PK parameters. The hepatic well-stirred model predicts total metabolic CL from *in vitro* measured or *in silico* predicted

$CL_{int,app.}$[12] CL is assumed to be predominantly mediated by hepatic metabolism. The free drug hypothesis is leveraged to derive CL_{int} from $CL_{int,app}$ and fraction unbound in microsomes (equation 1):

$$CL_{int} = \frac{CL_{int,app}}{fu_{mic}} \quad [1]$$

Blood to plasma partitioning is assumed to be unity for all compounds, and hepatic blood clearance is derived using equation 2, where fu_p is the fraction unbound in plasma and Q is the physiological liver blood flow (20.7 mL/min/Kg in human)[12]:

$$CL = \frac{CL_{int} \times fu_p \times Q}{CL_{int} \times fu_p + Q} \quad [2]$$

Volume of distribution at steady state (VD_{ss}) is predicted using the Korzekwa-Nagar model (equation 3), which characterizes distribution in three compartments: plasma, tissue water, and tissue lipids.[13, 14] R1 is the concentration ratio of plasma proteins in tissue water-to-plasma, estimated to be between 0.052 and 0.116. An average value of 0.084 is adopted in this study. V_p and V_t are the volume of plasma (0.043 L/Kg in human) and tissue (0.557 L/Kg in human), respectively. LK_L (product of lipid concentration and association constant for drug binding to lipids) can be predicted based on the fraction unbound in microsome (fu_{mic}) at 1 mg/mL microsomal protein concentration (equation 4)

$$VD_{ss} = V_p + V_t \times fu_p + V_t \times R1 \times (1 - fu_p) + V_t \times fu_p \times LK_L \quad [3]$$

$$V_t \times LK_L = \left(20 \times \frac{1 - fu_{mic}}{fu_{mic}}\right) + 0.76 \quad [4]$$

Equation 3 can be re-arranged to estimate fu_p from in vivo VD_{ss} and measured fu_{mic} as follows:

$$fu_p = \frac{(VD_{ss} - V_p - V_t \times R1) \times fu_{mic}}{fu_{mic} \times V_t - fu_{mic} \times V_t \times R1 + 20 - 19.2 \times fu_{mic}} \quad [5]$$

The use of this back-calculated fu_p in IVIVc is disclosed for the first time in this work.

MRT (mean residence time of a molecule in the body following an IV dose) and its inverse, K_e (rate of elimination), can be estimated from CL and VD_{ss} using equations 6 and 7:

$$MRT = \frac{VD_{SS}}{CL} \quad [6]$$

$$K_e = \frac{1}{MRT} \quad [7]$$

The well-stirred model is also used to estimate the fraction surviving first pass elimination in the liver (F_h) following oral absorption as a fraction of the liver blood flow using the following equation:

$$Fh = 1 - \frac{CL}{Q} \quad [8]$$

This assumption is valid for compounds that are primarily eliminated via hepatic metabolism. Fraction surviving first pass elimination in the gut following oral absorption (F_g) is assumed to be unity due to the lack of predictive in vitro and in silico tools to predict this parameter.

The fraction absorbed in the intestine (F_a) is calculated using the model presented by Yu, which estimates the absorbable dose (D_{abs}) based on the physiological parameters such as intestinal transit time (Ti, 3.3 hr adopted for human) and intestinal surface area (Si, 800cm^2 adopted for human), as well as the compound-specific parameters, including solubility and effective jejunum permeability (P_{eff}) as shown in equation 9.[15, 16] F_a can be calculated using equation 10, which is a ratio of D_{abs}-to-administered dose. The Fa value should be capped to a maximum of 1:

$$D_{abs} = P_{eff} * Sol * Ti * Si \quad [9]$$

$$Fa = \frac{D_{abs}}{Dose} \quad [10]$$

The use of intestinal solubility for the estimation of D_{abs} is a reasonable assumption for neutral, basic, zwitterionic and anionic compounds, for which the stomach solubility do not significantly exceed the intestinal solubility. Weakly basic compounds are typically well solubilized in the stomach due to the low gastric pH and can exhibit a markedly lower solubility in the intestine due to the prevalence of their neutral form. When the rate of precipitation from supersaturated solution is slow for a weakly basic compound, this has the potential to remain in solution during the intestinal transit time. Rate of precipitation is not a simple property to model, so an empirical approach is adopted to model and simplify this phenomenon. A D_{abs} value based on solubility in the stomach is calculated, and the final adopted D_{abs}

value is based on intestinal solubility when this is higher than stomach solubility. The average of D_{abs} based on intestinal and stomach solubilities is adopted in all the other cases to simulate an intermediate level of compound precipitation. A human dose of 210 mg is used in all simulations (3 mg/kg).

The human P_{eff} value is estimated from MDCK apparent permeability (MDCK P_{app}) data based on a linear fitting of the data published by Varma et al (equation 11), resulting in a R^2 0f 0.65 based on Log permeability of 17 marketed drugs (figure 5). [17]

$$P_{eff} = MDCK_{Papp} * 26.3 \quad [11]$$

The literature on predicting rate constant for absorption (k_a) with simple approaches is limited. In this work we leverage the Sinko model[18]:

$$k_a = \frac{2*P_{eff}}{R} \quad [12]$$

In this equation, the R stands for radius of human intestine, which is 1.75 cm. [19] In order to avoid unrealistic k_a values, this is capped at a lower boundary of 0.2 hr and a higher boundary of 1 hr. This model provides an optimistic estimate of k_a given the lack of consideration for dissolution; however, the model could be further evolved by leveraging mechanistic PBPK approaches. [20]

Finally, oral bioavailability (F) can be derived from the following equation:

$$F = F_a * F_h * F_g \quad [13]$$

Oral PK curves are generated using a one-compartment distribution model, chosen for its simplicity.[7] The average, maximum, and minimum concentration of the PK curve at steady state are defined by equations 14, 15, and 16, respectively:

$$Cavg = \frac{AUC}{\tau} \quad [14]$$

$$Cmax = \frac{Dose*k_a*F*\left(\frac{e^{-ke*tmax}}{1-e^{-ke*\tau}} - \frac{e^{-ka*tmax}}{1-e^{-ka*\tau}}\right)}{VD_{ss}*(ka-ke)} \quad [15]$$

$$Cmin = \frac{Dose*ka*F*\left(\frac{e^{-ke*\tau}}{1-e^{-ke*\tau}} - \frac{e^{-ka*\tau}}{1-e^{-ka*\tau}}\right)}{VD_{ss}*(k_a - k_e)} \quad [16]$$

Where τ is the dose interval, t is time, t_{max} is the time corresponding to C_{max}, calculated as:

$$tmax = \frac{\ln(k_a) - \ln(k_e)}{k_a - k_e} \quad [17]$$

AUC is the area under the simulated PK curve and can be calculated by the following equation:

$$AUC = \int_{t=0}^{t_{dose_interval}} \frac{Dose*ka*F*\left(\frac{e^{-ke*t}}{1-e^{-ke*\tau}} - \frac{e^{-ka*t}}{1-e^{-ka*\tau}}\right)}{VD_{ss}*(k_a - k_e)} \quad [18]$$

The coverage scores for average, maximum, and minimum concentrations (C_{avg}, C_{max}, C_{min}), can be derived using the following equations:

$$C_{avg\ score} = \frac{E_{\max} factor * C_{avg}}{MEC} \quad [19]$$

$$C_{max\ score} = \frac{E_{\max} factor * C_{max}}{MEC} \quad [20]$$

$$C_{min\ score} = \frac{E_{\max} factor * C_{min}}{MEC} \quad [21]$$

Here, minimum efficacious concentration (MEC) represents the total in vivo potency, which is estimated from apparent *in vitro* intrinsic potency ($EC_{50,app}$) and binding estimates in potency media (fu_{media}), as well as fu_p. This is calculated using equation 22 assuming same free concentration in plasma and at the site of action.

$$MEC = EC_{50,app} * \frac{fu_{media}}{fu_p} \quad [22]$$

When measurements in the fu_{media} are not available, this can be estimated based on the serum content in the media, fu_p, and by leveraging the dilution formula noted below, where D is the dilution factor for protein content in media-to-plasma (e.g. D=0.1 for a 10-fold dilution) [21]:

$$fu_{media} = \frac{\frac{1}{D}}{(\frac{1}{fu_p}-1)+\frac{1}{D}} \quad [23]$$

The $E_{max\ factor}$ is an adjustment factor to consider the depth of in vitro bioactivity. This is defined as:

$$Emax_{factor} = \begin{cases} 1, (if\ Emax \leq lower\ bound\ Emax) \\ 1 - \sqrt{\frac{Emax - lower\ bound\ Emax}{upper\ bound\ Emax - lower\ bound\ Emax}}, (if\ lower\ bound\ Emax < Emax \leq upper\ bound\ Emax) \\ 0, (Emax > upper\ bound\ Emax) \end{cases} \quad [24]$$

The method may also include an off-target coverage score. For example, a hERG score can be defined as the following:

$$hERG_{score} = \frac{IC_{50}(\text{hERG})}{fu_p * C_{max}} \quad [25]$$

Where IC_{50} hERG is the hERG inhibition potency in a dose response assay.

The off-target score can be combined with the coverage score of choice to provide a surrogate therapeutic index score (TI score). Input values from six experimental assays (fu_{mic}, fu_p, $CL_{int,app}$ in microsomes, kinetic solubility at pH 7.4, potency, and MDCK P_{app}) were derived from experimental assays where possible. If these values were not available, proprietary machine learning models were used instead[3]. Additionally, predicted pK_a values were utilized to estimate solubility at pH 2 (stomach pH) and pH 7.4 (intestinal pH). Solubility at different pH is calculated leveraging Henderson and Hasselbach equation as:

$$Sol_2 = Sol_{7.4} * \frac{FN_{7.4}}{FN_2} \quad [26]$$

$$FN_{7.4} = \frac{1}{1+10^{(pKa,b1-7.4)}+10^{(pKa,b1+pKa,b2-2*7.4)}+10^{(7.4-pKa,a1)}+10^{(2*7.4-pKa,a1-pKa,a2)}} \quad [27]$$

$$FN_2 = \frac{1}{1+10^{(pKa,b1-2)}+10^{(pKa,b1+pKa,b2-2*2)}+10^{(2-pKa,a1)}+10^{(2*2-pKa,a1-pKa,a2)}} \quad [28]$$

Where $pK_{a,a}$ and $pK_{a,b}$ are the most acidic and basic pK_a, respectively; $pK_{a,a2}$ and $pK_{a,b2}$ are the second most acidic and basic pK_a, respectively.[22] When no acidic or basic pK_a are available, use 14 for acidic and 0 for basic.

The accuracy in predicting CL, fu_{mic} and VD_{ss} is quantified as the absolute average fold error:

$$AAFE = 10^{\frac{\Sigma\left|\log_{10}\frac{Measured}{Predicted}\right|}{N}} \quad [29]$$

The area under the receiver operating (ROC) curve (AUC) is used to measure accuracy of ranking and is calculated using the pROC package.[23]

RESULTS AND DISCUSSION

In vitro to *in vivo* correlations of CL

The pre-requisite for the effective use of a mechanistic MPO model is the existence of solid *in vitro*-to-*in vivo* correlations (IVIVc) within the scaffold of interest; IVIVc disconnects for PK properties may point at additional elimination/distribution mechanisms that in turn rely on ancillary *in vitro* and *in silico* models to adequately describe these mechanisms. As a default assumption, discovery projects typically assume phase 1 Cytochrome P450 mediated metabolism as the primary mechanism of elimination, which is experimentally tested with NADPH-supplemented microsomal stability assays. When scaffold IVIVc disconnects are observed between microsomal stability and *in vivo* pre-clinical data, this assumption can be revised leveraging follow-up studies using hepatocytes, liver S9 fractions, bile duct canulated PK, PK in the presence of chemical inhibitors or metabolite identification.

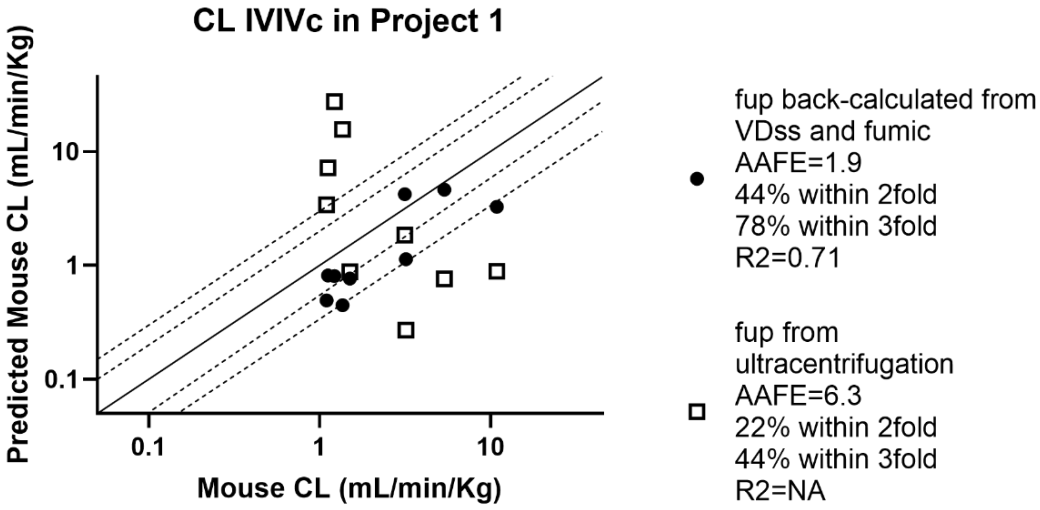
In addition, accuracy of fu_p measurements should also be evaluated. Plasma protein binding assays are prone to provide inaccurate estimates for compounds that have chemical or plasma stability issues, for compounds that are highly non-specifically bound to the assay apparatus, and for compounds that, due to the high extent of binding, fail to reach equilibrium under standard assay conditions.[24] For advanced compounds, customized measurements (e.g. extended incubation time, diluted plasma, bi-directional dialysis) are typically performed to solidify the fu_p estimates. One parsimonious approach to understanding the cause of CL IVIVc disconnects is to leverage the prediction of VD_{ss} from fu_{mic} to back-calculate an *in vivo* fu_p (equation 5). Because VD_{ss} and CL are independent PK parameters, the back-calculated *in vivo* fu_p value can subsequently be used to estimate *in vivo* CL. In project 1 (Figure 1, top graph), characterized by highly bound compounds with varying degrees of plasma instability, the initial method of choice for fu_p measurement was ultracentrifugation,[25] which resulted in poor CL IVIVc correlation. However, when back-calculated *in vivo* fu_p was used instead, the correlation improved significantly (from negative correlation to R^2 of 0.71 for LogCL).

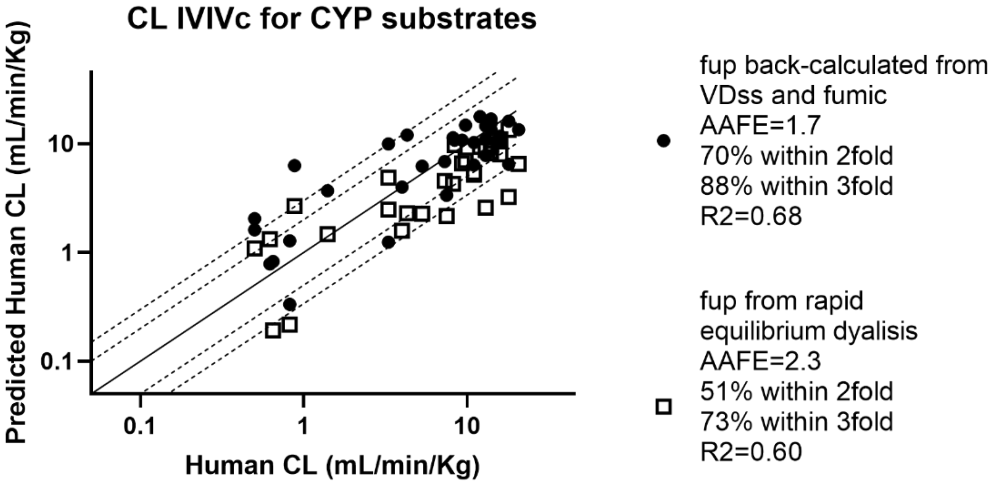
These observations are not unique to Project 1. Jones et al. recently presented a dataset of *in vitro* and *in vivo* data to assess CL predictions based on human hepatocytes.[26] This was followed by a publication from the same group describing a new approach to measuring fu_p, which significantly improved the CL prediction for a set of well-behaved Cytochrome P450 substrates for which CL IVIVc disconnects were previously observed.[27] Figure 1 (middle graph) shows the extended original dataset presented by Jones et al., including 33 Cytochrome P450 substrates with measured CL_{int}; all compounds included in this data exhibit CL_{int} values well within the dynamic range of the assay (>7 ml/min/kg). In this ideal dataset of well-behaved marketed drugs, the use of back-calculated *in vivo* fu_p measurements improved the absolute average fold error from 2.3 to 1.7, and the percent of predictions within 2-fold from 51% to 70%.

Figure 1 (bottom graph) additionally represents a similar analysis across all in-house compounds tested in mouse, rat, dog, or cynomolgus monkey. Since the mechanism of elimination for compounds in this dataset is unknown, property-based filters were adopted to focus on a chemical space in which the microsomal stability assay is likely to be representative of mechanisms of elimination, in agreement with previous reports about the chemical properties of P450 substrates (Caco2 P_{app} >5 x 10-6 cm/s, predicted $LogD_{7.4}$>1, MW between 250 and 750, neutral or basic compounds).[17, 26, 28] In this dataset, inclusive of numerous different scaffolds, a significant improvement in CL prediction accuracy is observed when the back-calculated *in vivo* fu_p was used: the absolute average fold deviation (AAFE) decreased from 4.9 to 2.5, R^2 increased from 0.1 to 0.65, % within 2-fold increased from 23% to 42%.

Based on these findings, it is recommended that fu_p measurements are investigated as a primary source of IVIVc CL disconnects by leveraging equation 5; when the re-analysis offers better correlation for *in vivo* and *in vitro* derived CL within the scaffold, the MPO functions presented in this work can be utilized to rank-order compounds in congeneric series. As previously discussed, fu_p has little influence on the $T_{1/2}$ (that is, MRT) of neutral and basic compounds that are predominantly distributed in lipids; in addition, Liu et al discussed how the fu_p has comparable weightage on total CL and MEC, effectively canceling out

in the C_{avg} based dose projection formula.[29] A notable exception is represented by hydrophilic and negatively charged compounds, for which the primary distribution compartment is not in the lipids. For these molecules, fu_p values have a substantial influence on the predicted MRT, hence when C_{max} or C_{min} driven efficacy is assumed for these molecules, accurate estimates of fu_p might be required. Sensitivity analysis based on equations 2 and 3 can be utilized to assess the impact of fu_p on MRT in a specific congeneric series of molecules.


CL IVIVc in Project 1
Predicted Mouse CL (mL/min/Kg)
Mouse CL (mL/min/Kg)
0.1
1
10
fup back-calculated from VDss and fumic
AAFE=1.9
44% within 2fold
78% within 3fold
R2=0.71
fup from ultracentrifugation
AAFE=6.3
22% within 2fold
44% within 3fold
R2=NA


CL IVIVc for CYP substrates
Predicted Human CL (mL/min/Kg)
Human CL (mL/min/Kg)
0.1
1
10
fup back-calculated from VDss and fumic
AAFE=1.7
70% within 2fold
88% within 3fold
R2=0.68
fup from rapid equilibrium dyalisis
AAFE=2.3
51% within 2fold
73% within 3fold
R2=0.60

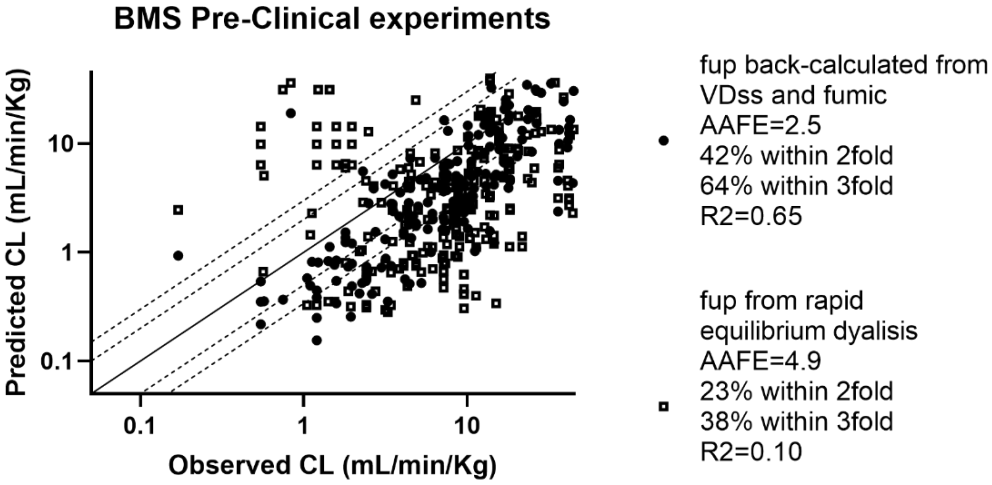

BMS Pre-Clinical experiments
Predicted CL (mL/min/Kg)
Observed CL (mL/min/Kg)
0.1
1
10
fup back-calculated from VDss and fumic
AAFE=2.5
42% within 2fold
64% within 3fold
R2=0.65
fup from rapid equilibrium dyalisis
AAFE=4.9
23% within 2fold
38% within 3fold
R2=0.10

Figure 1. Clearance predictions across 3 different datasets. Empty squares represent predictions generated leveraging in vitro plasma protein binding data. Full circles represent predictions generated leveraging plasma protein binding data back-calculated from in vivo volume of distribution at steady state leveraging the Korzekwa-Nagar method. Accuracy for the different approaches is quantified as absolute average fold error as well as percentage of predictions within two and three-fold.

In vitro to *in vivo* correlations of VDss and microsomal binding predictions

Lombardo et al. demonstrated that the mechanistic model described by Oie-Tozer is superior to allometric approaches in predicting human VDss across a large dataset of marketed drugs.[30] The Oie-Tozer model leverages similar assumptions and structure as the recent model presented by Korzekwa and Nagar, with the former describing tissue affinity in terms of fraction unbound in tissues (fut) and the latter using concentration and affinity to lipids (LKL).[14] Berellini and Lombardo used a large set of marketed drugs with experimentally determined pKa and lipophilicity to derive a prediction of fut and VDss based on physico-chemical parameters.[31] The accuracy of human VDss predictions based on physicochemical parameters is superior or comparable to the accuracy of predictions resulting from any individual pre-clinical species. While predictions based on LogD7.4 and pKa are of great value in the optic of model-driven drug discovery, these assays are not always available (especially pKa) during early phases for many compounds. Reliance on predicted values or surrogate chromatographic assays might produce significant deviation from the ideal values.[32]

In this work, the same dataset presented by Berellini et al. is used to compare the accuracy of the Oie-Tozer (state of the art) and more recent Korzekwa-Nagar models. In the context of early dose projection, the value proposition for the Korzekwa-Nagar model is the ability to predict VDss from fumic values, a parameter that in lead optimization is frequently measured in parallel with microsomal CLint shortly after synthesis. Experimental VDss in preclinical species can additionally be used to estimate in vivo pre-

clinical LKL, and subsequentially human VDss. Additionally, a prediction of LKL and VDss based on physicochemical properties is developed using the same dataset (Set 1) and following the same method as described by Berellini et al:

$$LK_L = 10^{0.3*LogD_{7.4}+1.2*FI_p-0.7*Logfu_p-0.48} \quad [30]$$

Where FI_p is the fraction positively ionized, a value that can be derived from equation 27 by setting the $pK_{a,a1}$ and $pK_{a,a2}$ values to 14.

Finally, a recent dataset with available experimental fu_{mic}, predicted pK_a, and chromatographic $LogD_{7.4}$ is used in combination with the dataset presented by Obach et al as further validation (Set 2).[33] Results from this analysis highlighted similar accuracy between the two models; predictions based on in vitro data inputs emerge as a valid alternative to predictions based on early rodent PK screening for estimating human VD_{ss}. Minor differences in the compound count reflects the exclusion from the Oie-Tozer predictions of compounds with aberrant fu_t as described by Lombardo et al. These results are summarized in Figure 2.

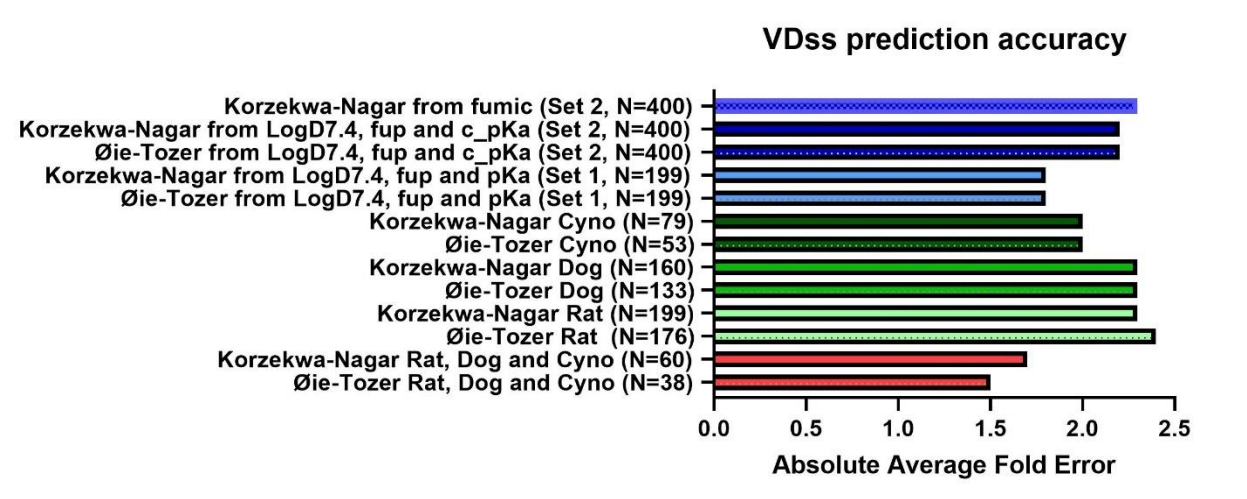


Figure 2. Predictions of Volume of distribution at steady state across different datasets and methods

Equation 30 and equation 4 can additionally be combined to predict fumic from physico-chemical properties (equation 31). Winiwarter et al. reviewed the accuracy of different methods for predicting

fumic from physicochemical properties and found that all the models reviewed demonstrated higher accuracy for compounds with LogD7.4<3, while predictions for more highly bound/lipophilic compounds were generally less accurate.[34] In this work, the Turner model, which previously demonstrated state-of-the-art performance, is selected as a comparison for fumic predictions based on the combination of equations 30 and 4.[34] This is shown in Figure 3:

$$fu_{mic} = \frac{20}{0.56*(10^{0.3*\,LogD_{7.4}+1.2*FI_p-0.7*Logfu_p-0.48})/V_t+19.2} \quad [31]$$

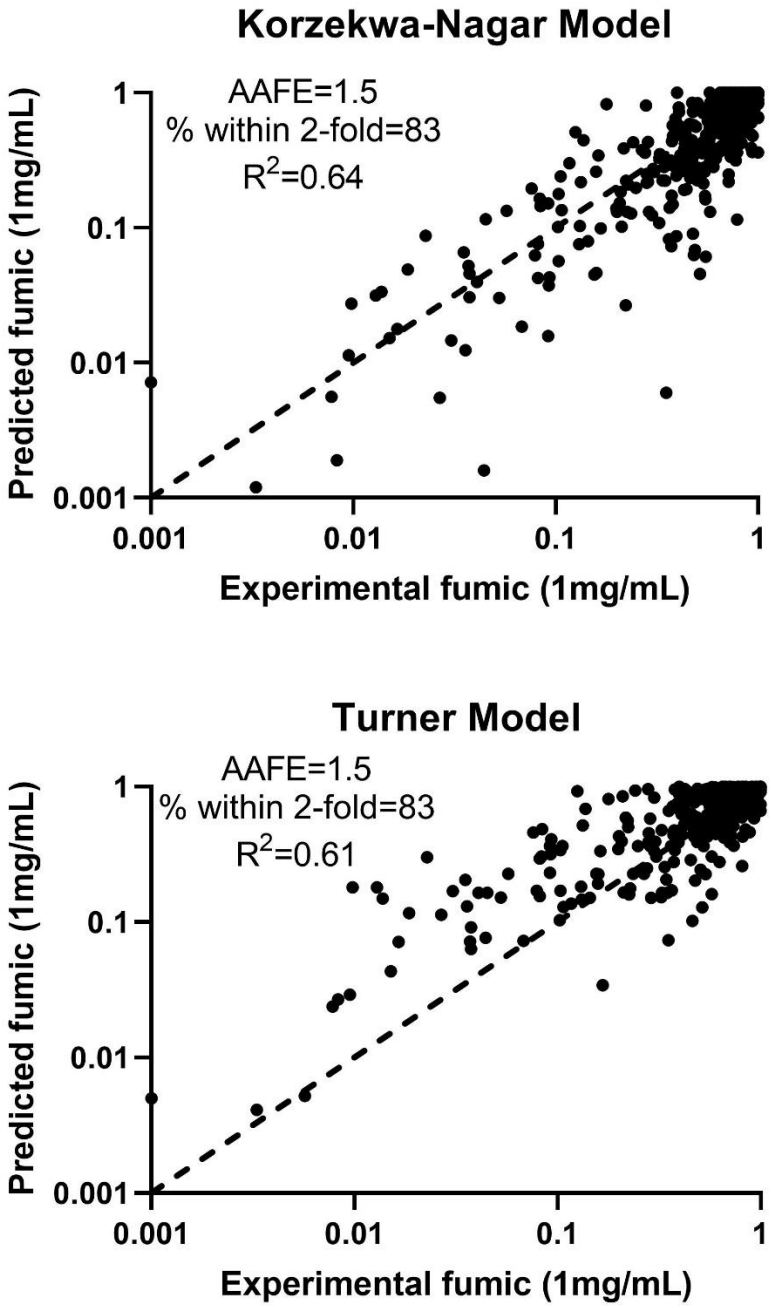


Figure 3. Predictions of microsomal binding based on physico-chemical properties.

The predictions based on the equation 31 (presented in this work) demonstrates comparable performance with state-of-the-art models, while retaining self-consistence with VDss estimate and physico-chemical properties. While proprietary machine learning models were used in-house for fu_{mic}

prediction due to their superior accuracy in the scaffold of interest, equation 31 provides an alternative approach to the same goal when robust fu_{mic} models are not available.

***In vitro* to *in vivo* correlations of MRT**

Cecere et al. previously characterized accuracy in MRT predictions from in vitro properties leveraging lipophilic metabolic efficiency (LipMetE).[35] In this manuscript, "Set 2" is further refined to only include compounds that are primarily eliminated via metabolism (<20% excreted unchanged) and in a physicochemical space in which transporters are expected to have minimal influence on total CL (neutral or basic, P_{app} in MDCK>$5x10^{-6}$ cm/s and $LogD_{7.4}$>1).[17, 26, 28] For the prediction of VD_{ss}, the Korzekwa-Nagar model based on physicochemical data is used when $LogD_{7.4}$<3 (due to the high fu_{mic} values in this space making the estimate of LK_L from microsomes less quantitative); predictions based on the same approach as well as fu_{mic} were averaged for the remaining compounds. Figure 4 summarizes the comparative performance of the methods presented in this work and the method presented by Cecere et al. While quantitative predictions of MRT remain challenging, the methods presented in this work demonstrate good ability to differentiate shorter vs longer MRT compounds.

The predicted MRT shows a reasonable rank correlation with experimental MRT (Spearman R = 0.53), while LipMetE exhibited a weaker correlation (Spearman R = 0.19) . In addition, employing the LipMetE approach revealed approximately 30% of high MRT (MRT >8 h) and low MRT (MRT<4 h) compounds being misrepresented in the lower and higher bins for LipMetE, respectively (Figure 4). However, a significant improvement was observed when utilizing the predicted MRT from the Korzekwa-Nagar model, with approximately 13% and 20% of high MRT and low MRT compounds being incorrectly categorized into lower and higher bins for predicted MRT, respectively.

An additional benefit of using the Korzekwa-Nagar approach in conjunction with the well stirred model to predict MRT stems from the similarity in the input parameters for the two methods (fu_p, fu_{mic}). The

error propagation emerging from experimental inaccuracies of fu_p and fu_{mic} measurements is reduced in the MRT estimate, compared to estimates of VD_{ss} and CL. For example, a compound with $CL_{int,app}$ of 50 mL/min/kg and fu_{mic} equal to 0.9, variations in fu_p of up to 5-fold from the "true value" of 0.1, results in an approximately 5-fold variability in CL and/or VD_{ss}, while such error imparts <1.5-fold variability to the calculated MRT.

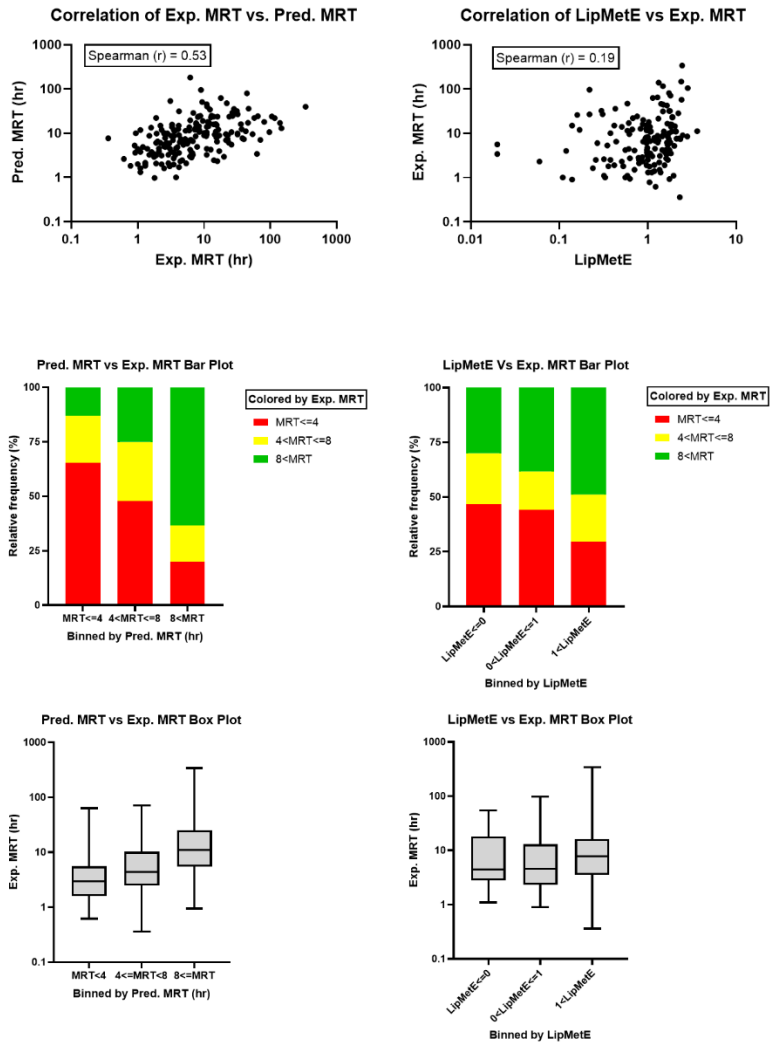


Figure 4. Comparison between intravenous mean residence time predictions in human from the combination of well-stirred hepatic model and Korzekwa-Nagar model, and lipophilic metabolic efficiency.

In vitro to *in vivo* correlation of oral absorption

Oral absorption is largely a function of compound's physico-chemical properties, passive permeability, and formulation; however, the majority of these parameters are unknown during early stages of research, which is a limitation for IVIVc based on pre-clinical data as well as for preliminary human dose projections; consequently, a qualitative assessment of F is a reasonable expectation before candidate nomination. Figure 5 presents the predictivity of the absorption models discussed in this study, using data published by Varma et al.[17] The average error in predicting F_a for a set of 46 oral drugs is 0.15 (15% of % Fa). Notably, some compounds with low molecular weight, which have been previously associated with paracellular transport (absorption between tight junctions), are included in this set and incorrectly predicted as having low F_a (such as cimetidine, ranitidine, and methyldopa).[36] Although this study does not explore thorough differentiation of absorption mechanisms, we attribute these discrepancies to the inability of the in vitro systems to account for paracellular transport accurately.

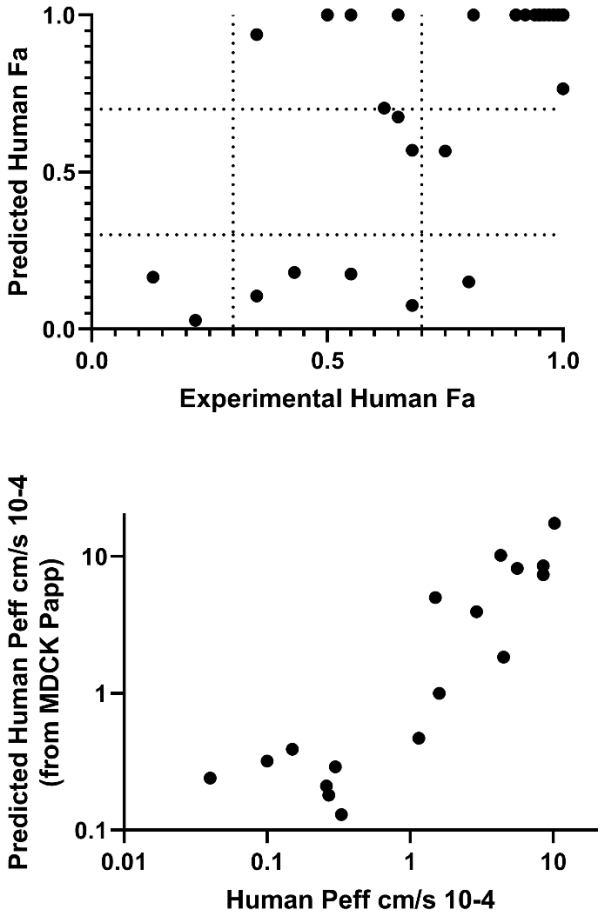


Figure 5. Prediction of fraction absorbed orally in human and human jejunum effective permeability from in vitro parameters.

Application of PK MPO function in small molecule drug discovery projects

This work provides examples of two small molecule projects, one of which assumes Cmin driven efficacy (Project 1) and one assuming Cavg driven efficacy (Project 2). Compounds are divided in different cohorts (early stage, middle stage and late stage) based on registration dates. Because both project teams started from sub-optimal chemical matter and progressed to clinical candidates, the average MPO score is expected to improve over time to reflect the progress made during the lead optimization campaign.

Project 1 involved 571 compounds for which lack of selectivity did not limit progression; when different entries are associated with racemic mixture and enantiomeric pure compounds, only racemates are included in the analysis. Twenty-four compounds were tested in rodent (4% of total) and four molecules in this project were short-listed as potential clinical candidates (0.7% of total). Figure 6 illustrates the MPO score for compounds in different cohorts. The team working on this project focused on two scores: Cmin score and a TI score (estimating hERG margins based on Cmin score and Cmax hERG score).

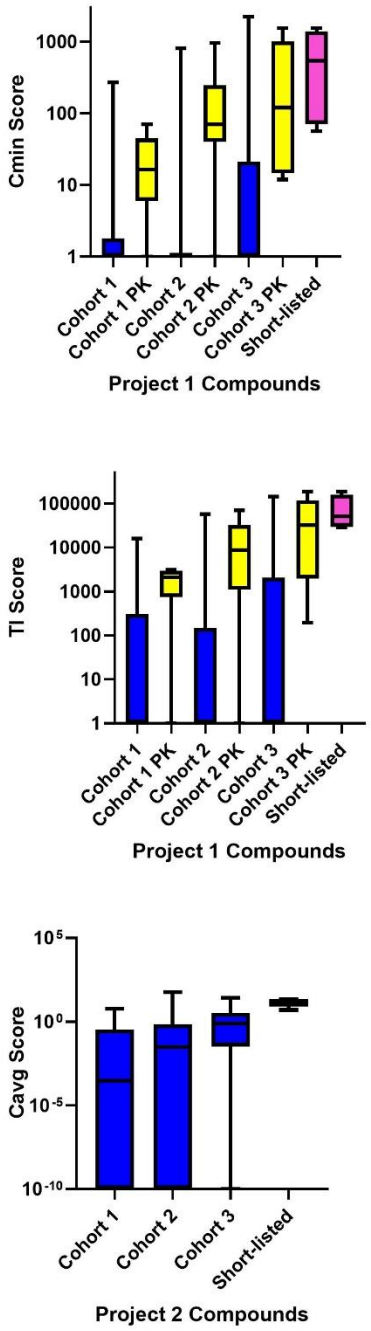


Figure 6. Application of the mechanistic pharmacokinetic multi-parameter optimization score in two internal projects. The different cohorts of compounds synthesized are based on the registration date; within the same cohort, the suffix PK indicates compounds tested in vivo in rodent, while the groups

lacking the PK suffix are compounds that have not been tested in vivo. Short-listed compounds were considered as potential clinical candidates during the final stages of the lead optimization.

The PK MPO successfully parallels the chronological progression of the chemical series and the better quality of the compounds selected for in vivo studies, as well as those considered for clinical experiments. The group of 15 top ranked molecules based on the TI Score includes the four short-listed compounds, and 8 compounds that had been characterized in PK studies; overall the short-listed compounds were found in the top 2nd percentile of the total distribution based on the TI score (AUC=0.99). The prediction of compounds characterized in vivo based on the TI score results in an AUC=0.86. The PK MPO score was developed in the context of project 1 and has been increasingly adopted to identify compounds as candidates for in vivo PK and PK/PD experiments. The chemistry team became familiar with this method while approaching candidate selection. Although the method was not used to identify additional candidates this was leveraged during subsequent efforts to identify back-up compounds, leading to a significant reduction of *in vivo* testing.

Project 2 is a legacy campaign for which fewer details are available on PK and PK/PD experiments due to different informatics systems for compound experiment tracking. A total of 415 compounds are available in this set, of which 2 were short-listed as clinical candidates (0.4% of total). PK/PD experiments run at the early stages of the project indicated AUC driven efficacy, consequently, compounds were ranked based on their C_{avg} score. Similar to what was observed in Project 1, the score successfully recapitulates compound progression and identifies the short-listed lead compounds with a large margin compared to the distribution of the compounds synthesized in the campaign (Figure 6): one lead is in the 1st percentile (5th out of 415), while the other is in the 9th percentile (39th out of 415). Overall, the prediction of short-listed compounds based on the C_{avg} score results in an AUC=0.95. The mechanistic PK MPO can be compared to more traditional linear scoring functions to gain insights on its differentiation and usefulness.[3] In project 1, the linear MPO score used by the same team identifies the 4

short-listed compounds in top 30, but none of them ranks in the top 15. In project 2, both short-listed compounds are predicted between the 3rd and the 8th percentile. In short, the two methods offer orthogonal and complementary approaches to score compounds, the observed differences can inform additional strategies to compound optimization that may have not been initially envisioned by the team.

After the success observed in Projects 1 and 2, the mechanistic PK MPO method was applied to six additional ongoing internal projects, consistently identifying the top lead compounds within the top 10th percentile of the scoring molecules. Currently, the PK MPO score is primarily used to rank-order synthesized compounds and inform in vitro and in vivo testing. Efforts to leverage the mechanistic PK MPO score during compound design are promising, but it is too early to draw meaningful conclusions.

CONCLUSION

In this study, we introduce a methodology based on mechanistic PK science to rank compounds in small molecule therapeutic projects. The equations, methods, and datasets used in the mechanistic score are all available either in the this manuscript or in the supporting information. The aim of this work is to produce high-throughput calculations to be incorporated in standard data mining software used by medicinal chemists for SAR; therefore, while we strived to leverage state of the art PBPK approaches for PK parameter prediction, no attempt was made to incorporate slower methods based on ordinary differential equations to describe the distribution of the test compounds across different organs.

Our retrospective IVIVc VDss analysis demonstrates that the Korzekwa-Nagar method exhibits accuracy comparable to the state of the art Oie-Tozer model.[30] In addition, we further expand the scope of the Korzekwa-Nagar to derive *in vivo* distribution and microsome binding predictions from physio-chemical properties (pK_a, fu_p, and $LogD_{7.4}$). Interestingly, predictions of microsome binding, indirectly obtained from the Korzekwa-Nagar method, perform similarly to previously published state-of-the-art methods for the same endpoint. Furthermore, we demonstrate how the *in vivo* fu_p back-calculated using

the Korzekwa-Nagar method can be leveraged to understand IVIVc CL disconnects across multiple datasets. These findings highlight plasma protein binding assays as an underappreciated potential source of IVIVc disconnects, in agreement with previous observations from DMPK scientists in academia and industry.[24-27, 33, 37-39] The Korzekwa-Nagar model combined with the well-stirred model allows for prediction of MRT based on three readily available experimental parameters (fu_{mic}, fu_p, and CL_{int}). While quantitative predictions of MRT based on *in vitro* data are still an active area of research, the categorical accuracy obtained using the approach presented in this work is a substantial improvement compared to the alternative method LipMetE.

The data shared regarding the use of mechanistic PK MPO scores in two small molecule drug discovery projects highlights the usefulness of these scoring functions in prioritizing compounds for *in vitro* and *in vivo* testing. In both Project 1 and 2, the mechanistic PK MPO approach recognizes the superior profile of the short-listed clinical candidates compared to the rest of the molecules synthesized during the project's lifetime (AUC between 0.95 and 0.99); all molecules short-listed for clinical experiments were found in the top 10th percentile. In addition, 5 out of 6 short-listed compounds were in the top 2nd percentile. Similar to short-listed compounds, molecules tested in PK experiments are associated with significantly higher scores compared to molecules that did not progress towards *in vivo* testing (AUC=0.86). Finally, in both Project 1 and 2, the mechanistic PK MPO scores increase chronologically, highlighting its value in tracking compound quality progression.

The combination of mechanistic PK MPO and IVIVc approaches presented in this work has the potential to significantly accelerate small molecule drug discovery campaigns and reduce *in vitro* and *in vivo* testing, as observed internally at during the Project 1 back-up effort.

ASSOCIATED CONTENT

Supporting Information. The following files are available free of charge.

Excel File with datasets used in this study.

ACKNOWLEDGEMENT

We thank Christoph Zapf, Natalie Hosea, and Stephen Johnson for spending time reading manuscript and providing suggestions. Thanks to scientists at BMS who are involved in projects to generate data used for our analysis.

REFERENCES

(1) Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R. K. Artificial intelligence in drug discovery and development. Drug Discov Today 2021, 26 (1), 80-93. DOI: 10.1016/j.drudis.2020.10.010 From NLM Medline.

(2) Jones, H. M.; Chen, Y.; Gibson, C.; Heimbach, T.; Parrott, N.; Peters, S. A.; Snoeys, J.; Upreti, V. V.; Zheng, M.; Hall, S. D. Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective. Clin Pharmacol Ther 2015, 97 (3), 247-262. DOI: 10.1002/cpt.37 From NLM Medline.

(3) Jia, L.; Weiss, D.; Shields, B.; Claus, B.; Shanmugasundaram, V.; Johnson, S.; Riggs, J.; Zapf, C. 2023. DOI: 10.26434/chemrxiv-2023-4hn4s-v2.

(4) Halford, G. S.; Baker, R.; McCredden, J. E.; Bain, J. D. How Many Variables Can Humans Process? Psychological Science 2005, 16 (1), 70-76. DOI: 10.1111/j.0956-7976.2005.00782.x.

(5) Pennington, L. D.; Muegge, I. Holistic drug design for multiparameter optimization in modern small molecule drug discovery. Bioorganic & Medicinal Chemistry Letters 2021, 41, 128003. DOI: https://doi.org/10.1016/j.bmcl.2021.128003.

(6) Peters, S. A.; Petersson, C.; Blaukat, A.; Halle, J. P.; Dolgos, H. Prediction of active human dose: learnings from 20 years of Merck KGaA experience, illustrated by case studies. Drug Discov Today 2020, 25 (5), 909-919. DOI: 10.1016/j.drudis.2020.01.002.

(7) Page, K. M. Validation of Early Human Dose Prediction: A Key Metric for Compound Progression in Drug Discovery. Mol Pharm 2016, 13 (2), 609-620. DOI: 10.1021/acs.molpharmaceut.5b00840.

(8) D.F. McGinnity, J. C., R.P. Austin, R.J. Riley. <EarlyHumanDoseProjectionMcGinnity.pdf>. Current Drug Metabolism 2007, 8, 7.

(9) Tautermann, C. S.; Borghardt, J. M.; Pfau, R.; Zentgraf, M.; Weskamp, N.; Sauer, A. Towards holistic Compound Quality Scores: Extending ligand efficiency indices with compound pharmacokinetic characteristics. Drug Discov Today 2023, 28 (11), 103758. DOI: 10.1016/j.drudis.2023.103758 From NLM Medline.

(10) Bartlett, D. W.; Gilbert, A. M. Translational PK-PD for targeted protein degradation. Chem Soc Rev 2022, 51 (9), 3477-3486. DOI: 10.1039/d2cs00114d From NLM Medline.

(11) Jansson-Löfmark, R.; Hjorth, S.; Gabrielsson, J. Does In Vitro Potency Predict Clinically Efficacious Concentrations? Clin Pharmacol Ther 2020, 108 (2), 298-305. DOI: 10.1002/cpt.1846 From NLM.

(12) Yang, J.; Jamei, M.; Yeo, K. R.; Rostami-Hodjegan, A.; Tucker, G. T. Misuse of the well-stirred model of hepatic drug clearance. Drug Metab Dispos 2007, 35 (3), 501-502. DOI: 10.1124/dmd.106.013359.

(13) Hsu, F.; Chen, Y. C.; Broccatelli, F. Evaluation of Tissue Binding in Three Tissues across Five Species and Prediction of Volume of Distribution from Plasma Protein and Tissue Binding with an Existing Model. Drug Metab Dispos 2021, 49 (4), 330-336. DOI: 10.1124/dmd.120.000337.

(14) Korzekwa, K.; Nagar, S. Drug Distribution Part 2. Predicting Volume of Distribution from Plasma Protein Binding and Membrane Partitioning. Pharm Res 2017, 34 (3), 544-551. DOI: 10.1007/s11095-016-2086-y.

(15) Yu, L. X. An integrated model for determining causes of poor oral drug absorption. Pharmaceutical Research 1999, 16 (12), 5.

(16) Ding, X.; Rose, J. P.; Van Gelder, J. Developability assessment of clinical drug products with maximum absorbable doses. Int J Pharm 2012, 427 (2), 260-269. DOI: 10.1016/j.ijpharm.2012.02.003.

(17) Varma, M. V.; Gardner, I.; Steyn, S. J.; Nkansah, P.; Rotter, C. J.; Whitney-Pickett, C.; Zhang, H.; Di, L.; Cram, M.; Fenner, K. S.; et al. pH-Dependent solubility and permeability criteria for provisional biopharmaceutics classification (BCS and BDDCS) in early drug discovery. Mol Pharm 2012, 9 (5), 1199-1212. DOI: 10.1021/mp2004912.

(18) Sinko, P. J.; Leesman, G. D.; Amidon, G. L. Predicting fraction dose absorbed in humans using a macroscopic mass balance approach. Pharm Res 1991, 8 (8), 979-988. DOI: 10.1023/a:1015892621261 From NLM.

(19) Price, E.; Kalvass, J. C.; DeGoey, D.; Hosmane, B.; Doktor, S.; Desino, K. Global Analysis of Models for Predicting Human Absorption: QSAR, In Vitro, and Preclinical Models. J Med Chem 2021, 64 (13), 9389-9403. DOI: 10.1021/acs.jmedchem.1c00669.

(20) Agoram, B.; Woltosz, W. S.; Bolger, M. B. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Advanced Drug Delivery Reviews 2001, 50, S41-S67. DOI: https://doi.org/10.1016/S0169-409X(01)00179-X.

(21) Kalvass, J. C.; Phipps, C.; Jenkins, G. J.; Stuart, P.; Zhang, X.; Heinle, L.; Nijsen, M.; Fischer, V. Mathematical and Experimental Validation of Flux Dialysis Method: An Improved Approach to Measure Unbound Fraction for Compounds with High Protein Binding and Other Challenging Properties. Drug Metab Dispos 2018, 46 (4), 458-469. DOI: 10.1124/dmd.117.078915.

(22) Butcher, G.; Comer, J.; Avdeef, A. pKa-critical Interpretations of Solubility–pH Profiles: PG 300995 and NSC-639829 Case Studies. Admet & Dmpk 2015, 3 (2). DOI: 10.5599/admet.3.2.182.

(23) Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011, 12 (1), 77. DOI: 10.1186/1471-2105-12-77.

(24) Di, L.; Breen, C.; Chambers, R.; Eckley, S. T.; Fricke, R.; Ghosh, A.; Harradine, P.; Kalvass, J. C.; Ho, S.; Lee, C. A.; et al. Industry Perspective on Contemporary Protein-Binding Methodologies: Considerations for Regulatory Drug-Drug Interaction and Related Guidelines on Highly Bound Drugs. J Pharm Sci 2017, 106 (12), 3442-3452. DOI: 10.1016/j.xphs.2017.09.005.

(25) Di, L. An update on the importance of plasma protein binding in drug discovery and development. Expert Opin Drug Discov 2021, 16 (12), 1453-1465. DOI: 10.1080/17460441.2021.1961741.

(26) Jones, R. S.; Leung, C.; Chang, J. H.; Brown, S.; Liu, N.; Yan, Z.; Kenny, J. R.; Broccatelli, F. Application of empirical scalars to enable early prediction of human hepatic clearance using IVIVE in drug discovery: an evaluation of 173 drugs. Drug Metab Dispos 2022. DOI: 10.1124/dmd.121.000784.

(27) Yan, Z.; Ma, L.; Huang, J.; Carione, P.; Kenny, J. R.; Hop, C.; Wright, M. New Methodology for Determining Plasma Protein Binding Kinetics Using an Enzyme Reporter Assay Coupling with High-Resolution Mass Spectrometry. Anal Chem 2023, 95 (8), 4086-4094. DOI: 10.1021/acs.analchem.2c04864.

(28) Varma, M. V.; Steyn, S. J.; Allerton, C.; El-Kattan, A. F. Predicting Clearance Mechanism in Drug Discovery: Extended Clearance Classification System (ECCS). Pharm Res 2015, 32 (12), 3785-3802. DOI: 10.1007/s11095-015-1749-4.

(29) Liu, X.; Wright, M.; Hop, C. E. Rational use of plasma protein and tissue binding data in drug design. J Med Chem 2014, 57 (20), 8238-8248. DOI: 10.1021/jm5007935.

(30) Lombardo, F.; Waters, N. J.; Argikar, U. A.; Dennehy, M. K.; Zhan, J.; Gunduz, M.; Harriman, S. P.; Berellini, G.; Rajlic, I. L.; Obach, R. S. Comprehensive assessment of human pharmacokinetic prediction based on in vivo animal pharmacokinetic data, part 1: volume of distribution at steady state. J Clin Pharmacol 2013, 53 (2), 167-177. DOI: 10.1177/0091270012440281.

(31) Berellini, G.; Lombardo, F. An Accurate In Vitro Prediction of Human VDss Based on the Oie-Tozer Equation and Primary Physicochemical Descriptors. 3. Analysis and Assessment of Predictivity on a Large Dataset. Drug Metab Dispos 2019, 47 (12), 1380-1387. DOI: 10.1124/dmd.119.088914.

(32) Chan, R.; De Bruyn, T.; Wright, M.; Broccatelli, F. Comparing Mechanistic and Preclinical Predictions of Volume of Distribution on a Large Set of Drugs. Pharm Res 2018, 35 (4), 87. DOI: 10.1007/s11095-018-2360-2.

(33) Tess, D.; Chang, G. C.; Keefer, C.; Carlo, A.; Jones, R.; Di, L. In Vitro-In Vivo Extrapolation and Scaling Factors for Clearance of Human and Preclinical Species with Liver Microsomes and Hepatocytes. Aaps j 2023, 25 (3), 40. DOI: 10.1208/s12248-023-00800-x From NLM.

(34) Winiwarter, S.; Chang, G.; Desai, P.; Menzel, K.; Faller, B.; Arimoto, R.; Keefer, C.; Broccatell, F. Prediction of Fraction Unbound in Microsomal and Hepatocyte Incubations: A Comparison of Methods across Industry Datasets. Mol Pharm 2019, 16 (9), 4077-4085. DOI: 10.1021/acs.molpharmaceut.9b00525.

(35) Cecere, G.; Guasch, L.; Olivares-Morales, A. M.; Umehara, K.; Stepan, A. F. LipMetE (Lipophilic Metabolism Efficiency) as a Simple Guide for Half-Life and Dosing Regimen Prediction of Oral Drugs. ACS Medicinal Chemistry Letters 2022, 13 (9), 1444-1451. DOI: 10.1021/acsmedchemlett.2c00183.

(36) Bourdet, D. L.; Pollack, G. M.; Thakker, D. R. Intestinal Absorptive Transport of the Hydrophilic Cation Ranitidine: A Kinetic Modeling Approach to Elucidate the Role of Uptake and Efflux Transporters and Paracellular vs. Transcellular Transport in Caco-2 Cells. Pharmaceutical Research 2006, 23 (6), 1178-1187. DOI: 10.1007/s11095-006-0204-y.

(37) Francis, L. J.; Houston, J. B.; Hallifax, D. Impact of Plasma Protein Binding in Drug Clearance Prediction: A Data Base Analysis of Published Studies and Implications for In Vitro-In Vivo Extrapolation. Drug Metab Dispos 2021, 49 (3), 188-201. DOI: 10.1124/dmd.120.000294.

(38) Baker, M.; Parton, T. Kinetic determinants of hepatic clearance: plasma protein binding and hepatic uptake. Xenobiotica 2007, 37 (10-11), 1110-1134. DOI: 10.1080/00498250701658296.

(39) Chen, Y. C.; Kenny, J. R.; Wright, M.; Hop, C.; Yan, Z. Improving Confidence in the Determination of Free Fraction for Highly Bound Drugs Using Bidirectional Equilibrium Dialysis. J Pharm Sci 2019, 108 (3), 1296-1302. DOI: 10.1016/j.xphs.2018.10.011.