Predicting Volume of Distribution in Humans: Performance of In Silico Methods for a Large Set of Structurally Diverse Clinical Compounds

Neha Murad, Kishore K. Pasikanti, Benjamin D. Madej, Amanda Minnich, Juliet M. McComas, Sabrinia Crouch, Joseph W. Polli, and Andrew D. Weber


Received September 30, 2020; accepted November 3, 2020

ABSTRACT

Volume of distribution at steady state (V_{D,ss}) is one of the key pharmacokinetic parameters estimated during the drug discovery process. Despite considerable efforts to predict V_{D,ss} accuracy and choice of prediction methods remain a challenge, with evaluations constrained to a small set (<150) of compounds. To address these issues, a series of in silico methods for predicting human V_{D,ss} directly from structure were evaluated using a large set of clinical compounds. Machine learning (ML) models were built to predict V_{D,ss} directly and to predict input parameters required for mechanistic and empirical V_{D,ss} predictions. In addition, log D, fraction unbound in plasma (fup), and blood-to-plasma partition ratio (BPR) were measured on 254 compounds to estimate the impact of measured data on predictive performance of mechanistic models. Furthermore, the impact of novel methodologies such as measuring partition (Kp) in adipocytes and myocytes (n = 189) on V_{D,ss} predictions was also investigated. In predicting V_{D,ss} directly from chemical structures, both mechanistic and empirical scaling using a combination of predicted rat and dog V_{D,ss} demonstrated comparable performance (62%-71% within 3-fold). The direct ML model outperformed other in silico methods (75% within 3-fold, r^2 = 0.5, AAFe = 2.2) when built from a larger data set. Scaling to human from predicted V_{D,ss} of either rat or dog yielded poor results (<47% within 3-fold). Measured fup and BPR improved performance of mechanistic V_{D,ss} predictions significantly (81% within 3-fold, r^2 = 0.6, AAFe = 2.0). Adipocyte intracellular Kp showed good correlation to the V_{D,ss} but was limited in estimating the compounds with low V_{D,ss}.

SIGNIFICANCE STATEMENT

This work advances the in silico prediction of V_{D,ss} directly from structure and with the aid of in vitro data. Rigorous and comprehensive evaluation of various methods using a large set of clinical compounds (n = 956) is presented. The scale of techniques evaluated is far beyond any previously presented. The novel data set (n = 254) generated using a single protocol for each in vitro assay reported in this study could further aid in advancing V_{D,ss} prediction methodologies.

Introduction

The current drug discovery path is a sequential, time-consuming process with a high attrition rate (Hinkson et al., 2020). Attrition of small-molecule drug candidates due to poor pharmacokinetic (PK) profiles has diminished significantly in recent years (Waring et al., 2015). This advancement can partly be attributed to the unprecedented emphasis on screening compounds based on PK parameters in the drug discovery phase (Ferreira and Andricopulo, 2019). PK is a well recognized and fundamental property that influences drug concentrations at target, which ultimately determines a drug’s efficacy and safety (Ferreira and Andricopulo, 2019). Volume of distribution at steady state (V_{D,ss}) is a key PK parameter that describes the relationship between drug concentration measured in plasma or blood to the amount of drug in the body at equilibrium (Smith et al., 2015). Estimation of apparent V_{D,ss} is of utmost importance because it influences C_{max} and half-life in plasma and target tissues, which in turn determines dose and dosing regimen in the clinic (del Amo et al., 2013). Toward this end, V_{D,ss} in humans is commonly predicted using preclinical in vivo and in vitro data in conjunction with various allometric scaling methods such as the Oie and Tozer method (Jones et al., 2011). Alternatively, V_{D,ss} can be extrapolated from tissue-to-plasma partition coefficients (Kp) from preclinical species (generally rat) (Nigade et al., 2019). These experiments
are resource-intensive and require the synthesis of compounds; these limitations further hinder the ability to predict human V_D,ss early in drug discovery or during lead optimization. Thus, considerable effort has been undertaken to develop predictive in silico models to accelerate and reduce the cost of drug discovery processes (Wenzel et al., 2019). As V_D,ss is dependent on the tissue partitioning of compounds, numerous studies have focused on developing in silico approaches to predict tissue partitioning based on physicochemical properties such as pKa and log P, plasma protein binding, and blood-to-plasma partition ratio (BPR) (Graham et al., 2012; del Amo et al., 2013). Poulin and Theil were some of the first to propose a mechanistic Kp prediction method (Poulin and Krishnan, 1995; Poulin and Theil, 2002). This method incorporates several important mechanisms, such as albumin binding, neutral lipid, and phospholipid binding. Berezhkovskiy (2004) is another method similar to Poulin and Theil. The Rogers and Rowland method (Rodgers et al., 2005; Rodgers and Rowland, 2006) is by far the most comprehensive Kp prediction method in terms of mechanisms captured. It includes all the mechanisms captured in previous published methods along with the addition of acidic phospholipid and cytosolic ion partitioning. A drawback for the Rogers and Rowland method is that there are two sets of equations based on the dissociation constant or pKa of the compounds, and the cutoff or switch between these equations was set at a pKa of 7. This results in a discontinuous relationship between the dissociation constant and plasma tissue partitioning. Finally, the method is also heavily dependent on accurate pKa predictions. To address these issues, a modified Rogers and Rowland method was developed (Lukacova et al., 2008) that employs a single continuous combined equation for compounds regardless of pKa. Ion partitioning into acidic or basic intracellular compartments (lysosomes and mitochondria) was described by Trapp et al. (2008) and can be used as an aid to Kp prediction method for compounds for which ion trapping is expected. Key mechanisms that play a crucial role in partitioning itself between plasma and the specific organ tissue implemented by each prediction method is summarized in Table 1.

Accurately predicting V_D,ss remains a challenge that has not been adequately solved (Smith et al., 2015). Few studies have evaluated the performance of various V_D,ss prediction methods; however, these reports were either in preclinical species (Graham et al., 2012) or used a small set (<150) of clinical compounds (Jones et al., 2011; Korzekwa and Nagar, 2017; Chan et al., 2018; Nigade et al., 2019; Mayumi et al., 2020). Recently, Lombardo et al. (2018) published a manually curated data set of V_D,ss for 1352 drugs after intravenous dosing, which presented an opportunity to evaluate the predictive performance of various V_D,ss methodologies in determining human V_D,ss. Therefore, we investigated the 1) performance of the most common V_D,ss prediction strategies, 2) sensitivity of input parameters that influence V_D,ss predictions, 3) impact of experimental data on mechanistic V_D,ss predictions, and 4) whether novel methodologies such as using adipocyte and myocyte cell partitioning could improve V_D,ss predictions.

### Materials and Methods

#### Experimental Approaches

The V_D,ss prediction strategies investigated are broadly categorized into two approaches based on the starting data for the analysis, which is either fully in silico (e.g., structural) or in vitro (experimental). Based on the compound availability, an initial in vitro experimental data set of 331 compounds (Lombardo et al., 2018) was identified. Predictive performances were assessed using 956 compounds for the in silico and 254 compounds for the in vitro experimental approaches, respectively.

For the in silico approach, V_D,ss was predicted directly from chemical structure using compound Simplified Molecular Input Line Entry System (SMILES) as input by using the following four approaches: 1) mechanistic V_D,ss prediction using predicted physicochemical properties from commercial software (ADMET Predictor 9.0) or 2) using machine learning (ML) models generated by the Accelerating Therapeutics for Opportunities in Medicine (ATOM) consortium, 3) allometric scaling from predicted V_D,ss for preclinical species such as rat and dog ML models, and 4) direct human V_D,ss predictions using an ML model built using clinical compounds (see schematic shown in Fig. 1).

In the Experimental Data approach, two distinct experimental data sets were generated. The first experimental data set included measurement of physicochemical properties under a single protocol for each in vitro experiment, which included log D, fraction unbound in plasma (fup), and BPR for 331 clinical compounds (Lombardo et al., 2018). The above experimental data were used as input parameters individually or in combination to predict mechanistic V_D,ss (Lukacova et al., 2008). In addition, novel experiments were conducted to determine partition of compounds in human adipocytes and myocytes for 200 compounds that were a subset of the 331 compounds selected above. In silico and experimental methodologies are further described in detail below. The percentage of compounds that had accurately predicted V_D,ss within 2-, 3-, or 10-fold; r^2 (Pearson correlation coefficient); and absolute average fold error (AAFE) were used as key criteria for comparison of predictive performance of each method.

#### In Silico Methods

V_D,ss of the clinical compounds data set (Lombardo et al., 2018) was subdivided based on whether experimental data were directly measured (331 compounds) or not (970 compounds). Evaluation of in silico methods was...
performed on both data sets. It is important to note that all the evaluations were performed on a complete hold-out set. For example, when predicting \( V_{Dss} \) for the experimental data set, none of the compounds in the experimental data set were a part of any of the ML model building data sets.

**ADMET Mechanistic \( V_{Dss} \) Prediction.** ADMET Predictor (version 9.0) was used to predict \( pK_a \) (S + Acidic \( pK_a \), S + Basic \( pK_a \), fraction unbound in plasma \( (\text{hum}_\text{fup}\%\), converted to fup), BPR, and log P/D (S + log D, S + log P) from chemical structure. These parameters were subsequently used as input parameters to predict mechanistic \( K_p \) and human \( V_{Dss} \) predictions (Lukacova et al., 2008). Predicted values of the input parameters were limited to typical assay limits for these parameters.

**ATOM Mechanistic, Allometry, and Direct ML Predictions**

**ATOM Mechanistic \( V_{Dss} \) Prediction.** Data sets generated by GlaxoSmithKline (Supplemental Table 1) containing molecular structure information and physicochemical parameters (log D, fup, BPR) were split into train, validation, and test subsets. Model training and evaluation was generally performed as previously described (Minnich et al., 2020). Briefly, a grid search hyperparameter optimization technique was employed to train several machine learning models (neural networks and random forests) with different hyperparameter combinations of learning rates, layer sizes, and number of nodes. Random forest and neural network models were selected by picking the model with the maximum validation set \( R^2 \). Clinical compounds were grouped into two sets. The first set of compounds was the 287 compounds that were selected for experimental measurements. These sets were used in two ways for fitting and prediction. 1) To compare predictive performance of the direct ML models against the other in vitro approaches, models were trained using the 970 human \( V_{Dss} \) of compounds without further experimental measurements. The \( V_{Dss} \)_ML model was then used to predict \( V_{Dss} \) for the 287 compounds with new experimental measurements for comparison with in vitro methods. 2) A very challenging (due to the small size of the training set) external test set was used by inverting the previous approach. Models were developed using 287 compounds with new experimental measurements. Then, the fit model was used to predict \( V_{Dss} \) for the 970 compounds without further experimental measurements. In both approaches, the set of compounds used for model development was further split into training, validation, and internal test sets as previously described.

**Experimental Data**

**Log D.** The chromatographic hydrophobicity index (CHI) (Valkó et al., 1997) values were measured using a reversed phase high-performance liquid chromatography (HPLC) column (50 × 2 mm 3 \( \mu \)M Gemini NX C18; Phenomenex, UK) with fast acetonitrile gradient at starting mobile phase of pH 2, 7.4, and 10.5. CHI values are derived directly from the gradient retention times using calibration.
parameters for standard compounds. The CHI value approximates to the volume percent organic concentration when the compound elutes. CHI is linearly transformed into ChloroformD (Young et al., 2011) by least-squares fitting of experimental CHI values to calculated ChloroformP values for over 20,000 research compounds using the following formula: ChloroformD = 0.0857 × CHI - 2.00.

**Blood-to-Plasma Partition Ratio.** In vitro measurement of blood-to-plasma partition was conducted in human blood (K₂EDTA as anticoagulant) obtained from a commercial source (BioReclamation IVT, Liverpool, NY). Hematicrit (the ratio of volume of red blood cells to total blood) was measured by centrifugation of the whole blood at 3000 rpm for 10 minutes using microhematocrit capillary tubes. Control plasma was prepared from a portion of the whole blood by centrifugation at 3000g for 10 minutes. Both whole blood and control plasma samples were warmed at 37°C in a water bath for 30 minutes. Subsequently, the test compounds (1 μM in the final concentration) and controls were vortexed for 1 minute at 35°C. The resulting mixtures were vortex-mixed, followed by centrifugation at 3000 rpm for 10 minutes. The supernatants were transferred to HPLC autosampler plate. Test compounds and the internal standard (100 ng/ml tolbutamide in acetonitrile) were added to the 0-minute and 240-minute stability wells. Analysis of samples was performed using LC/MS/MS. For all samples, peak area ratios were used to determine percent unbound. Plasma proteins were precipitated with 400 μl of 160 mmol/l of PBS buffer. A sample (160 μl) was spiked into blood and incubated at 37°C (5% CO₂) with shaking at 200 rpm for 60 minutes along with control samples. After incubation for 60 minutes, the incubated whole blood was removed from the water bath, and the plasma was separated by centrifugation at 1000g for 10 minutes. Aliquots of the control plasma were also removed. All plasma samples (50 μl) were treated with 400 μl of ice-cold acetonitrile containing an internal standard (100 ng/ml tolbutamide in acetonitrile). After the removal of protein by centrifugation at 16400 g (3000 rpm) for 10 minutes at 4°C, the supernatants were transferred to HPLC autosampler plate. Test compounds and internal standard solution were measured using liquid chromatography with tandem mass spectrometry (LC/MS/MS). Blood-to-plasma partition was calculated by ratio of mass spectrometric response of compounds in blood samples after 60 minutes of incubation to itself mass spectrometric response in plasma samples.

**Fraction Unbound in Plasma.** In vitro measurement of fup was conducted using a rapid equilibrium dialysis (RED) device. The fup values of test compounds and a positive control (warfarin) were determined at a single time point of 4 hours postincubation. Considering high surface-to-volume ratio of the membrane compartment in a RED device, equilibrium is expected to be achieved within 4 hours of incubation (Waters et al., 2008). Stock solutions of test compounds and warfarin were prepared in DMSO at concentrations of 5 mM and subsequently diluted to a final concentration of 0.5 mM in DMSO/water (1:1, v/v). Incubation mixtures were prepared by diluting the stock solution into human plasma obtained from a commercial source (BioReclamation IVT). Final concentrations of compounds in incubation mixture were 5 μM. Human plasma was prewarmed in a water bath at 37°C prior to the experiment. In total, 400 μl of the stopping solution (100 ng/ml tolbutamide in acetonitrile) was added to a 96-well deep well sample collection plate on ice. In a RED device, 500 μl of PBS was added to the white chambers (receiver side), and aliquots (300 μl) of each incubation mixture were spiked into the red wells (donor side). A sample (40 μl) of the incubation mixture was transferred into the 0-minute wells on the sample collection plate. The device and remaining spiked plasma samples were incubated at 37°C for 4 hours with shaking at 150 rpm. After the incubation period, 40 μl of the remaining spiked plasma was transferred to the sample collection plate. All samples in the RED device were mixed by pipetting prior to aliquoting (40 μl) from each donor well into a well containing 160 μl of PBS buffer. A sample (160 μl) of each receiver well was aliquoted into a tube containing 40 μl of blank plasma. PBS (160 μl) was added to the 0-minute and 240-minute stability wells. Analysis of samples was performed using LC/MS/MS. For all samples, peak area ratios were used to determine percent unbound. Plasma proteins were precipitated with 400 μl of acetonitrile containing 100 ng/ml tolbutamide as a mass spectral internal standard. The resulting mixtures were vortex-mixed, followed by centrifugation for 15 minutes at >3500 rpm/min. A sample (100 μl) of the supernatant/well was transferred to a clean 96-well plate containing 100 μl of ultrapure water/well. The plate was vortexed for 1 minute at >1700 rpm/min. Aliquots (4 μl) of the resulting supernatant were injected onto the LC/MS/MS system to obtain peak area ratios for each compound to determine fraction unbound in plasma. Equilibrium dialysis method for measuring fup is amenable to automation and is generally accepted as the gold standard (Trainor, 2007).

**Adipocyte and Myocyte Partition.** Intracellular partition of compounds in adipocytes and myocytes was determined using a protocol described previously (Treyer et al., 2018). Primary human adipocytes and myocytes were obtained from commercial sources (Lonza, MD). The test compounds and controls at a final concentration of 0.5 μM were incubated with fully differentiated myocytes and adipocytes plated in culture in triplicate at 37°C (5% CO₂) with shaking at 100 rpm for 45 minutes. After the end of the incubation, the medium was transferred to a stop solution containing acetonitrile and internal standard (100 ng/ml tolbutamide in acetonitrile). The cell layer was washed with 200 μl of cold Hanks’ buffered salt solution and extracted with stop solution (100 ng/ml tolbutamide in acetonitrile). Both the intracellular and extracellular compound concentrations were analyzed using LC/MS/MS. The cell protein concentration was determined by the bicinchoninic acid assay. Intracellular drug accumulation (Kp) was calculated from the peak area ratios of the analyte to internal standard in the medium, cells, and protein concentration from the following equation: Protein content was quantified using the bicinchoninic acid assay in representative wells to calculate the cellular volume (Vc), assuming 6.5 μg/mg protein (Treyer et al., 2018). Amount of drug in the cells (Acell) was estimated using peak area ratio and volume of cell lysate (area ratio × volume of cell lysate). Cmedium refers to corrected medium concentration. Intracellular accumulation was determined using cell lysate concentration × volume of cell lysate (150 μl). Subsequently, the Kpinter, cell is calculated accounting from protein binding in plasma.

\[
Kp_{\text{inter, cell}} = \frac{A_{\text{cell}}}{C_{\text{medium}}} \\
Kp_{\text{fat}} = fup \times Kp_{\text{adipocyte, inter, cell}} \\
Kp_{\text{muscle}} = fup \times Kp_{\text{myocyte, inter, cell}}
\]

**Predictions Based on Experimental Data**

**Mechanistic Models for Kp Prediction.** Experimental data (log D, fup, BPR) were used as input parameters individually or in combination to predict Kp (Lukacova et al., 2008) and subsequently were used to calculate VDss using the following relationship:

\[
VD_{\text{ss}} = V_{p} + \frac{E_{p}}{P} + \sum_{i=1}^{n} Kp_{i} V_{f_i}
\]

where \(V_{p}\) is the volume of plasma; \(V_{f_i}\) is the volume of erythrocytes (\(V_{\text{blood}} - V_{p}\)); \(E_{p}\) is the erythrocyte-to-plasma ratio, which is derived by the equation \(BPR + \text{hematocrit} - 1\)/hematocrit; and \(Kp_{i}\) and \(V_{f_i}\) are the tissue partition ratio and volume, respectively, for the \(i\)th tissue (Nigade et al., 2019).

**Tissue-Level Kp Prediction.** We used five strategies for predicting VDss using adipocytes and myocyte Kp values:

1. **Adipocyte-only method:** Adipocyte Kp values were used to calculate partitioning into fat (\(Kp_{\text{fat}}\)). Kp for other organs was assumed to be 1 to predict VDss using the following equation:

\[
VD_{\text{ss}} = V_{p} + (Kp_{\text{fat}} \times V_{\text{fat}}) + 1 \times V_{\text{rest of the body}}
\]

2. **Myocyte-only method:** Myocyte Kp values were used to calculate partitioning into muscle tissue (\(Kp_{\text{muscle}}\)), and Kp for other organs was assumed to be 1 to predict VDss using the following equation:

\[
VD_{\text{ss}} = V_{p} + (Kp_{\text{muscle}} \times V_{\text{muscle}}) + 1 \times V_{\text{rest of the body}}
\]

3. **Combined method:** Both adipocyte and myocyte Kp values were used to calculate fat and muscle volumes, respectively. Kp for all nonfat and muscles organs was assumed to be 1 to predict VDss.

\[
VD_{\text{ss}} = V_{p} + (Kp_{\text{fat}} \times V_{\text{fat}}) + (Kp_{\text{muscle}} \times V_{\text{muscle}}) + (1 \times V_{\text{rest of the body}})
\]

4. **Average method:** Average of adipocyte and myocyte Kp values were used as Kp for all nonfat and muscle tissues. Both adipocyte and myocyte Kp values were used to calculate fat and muscle volumes, respectively, to predict VDss.

\[
VD_{\text{ss}} = V_{p} + (Kp_{\text{fat}} \times V_{\text{fat}}) + (Kp_{\text{muscle}} \times V_{\text{muscle}}) + (Kp_{\text{fat}} \times V_{\text{rest of the body}}) + (Kp_{\text{muscle}} \times V_{\text{rest of the body}})
\]

5. **Separate method:** Mechanistic Kp (Lukacova et al., 2008) calculations were used for nonfat or nonmuscle organs. Both adipocyte and myocyte
Kp values were used to calculate fat and muscle volumes, respectively. Both of the volumes were subsequently added to predict VD,ss as follows:

$$VD_{ss} = V_p + (K_{p_{fat}} \times V_{fat}) + (K_{p_{muscle}} \times V_{muscle}) + VD_{ss\text{, tissues not considered}}$$

**Results**

As summarized in Fig. 1, we investigated the performance of the most common VD,ss prediction strategies, sensitivity of input parameters that influence VD,ss predictions, impact of experimental data on mechanistic VD,ss predictions, and whether adipocyte and myocyte cell partitioning could improve predictive performance by using a large compound data set. An in silico–only approach was applied using a set of 956 compounds (the ATOM in silico set) related to the Lombardo intravenous dosing drug set (n = 1352 drugs) in which VD,ss values were reported (Lombardo et al., 2018). A separate set of compounds, the ATOM experimental set (n = 254 compounds), had additional in vitro data collected under uniform experimental conditions (see Materials and Methods; Supplemental Table 2) and was used as a comparator against the purely in silico methods. Although the ATOM experimental data set was selected based on the compound availability from an initial set of 331 drugs, it represented chemical diversity of the clinical data set (Supplemental Fig. 1).

The comparative assessments of various in silico approaches evaluated to predict human VD,ss for two discrete sets of compounds are summarized in Fig. 2 and Table 2. Details of ATOM ML models used to predict input parameters for mechanistic VD,ss predictions are shown in (Supplemental Table 1). Model/featurization combination that resulted in the best models varied by data sets. MOE or graph convolution featurization with random forest or neural network models

<table>
<thead>
<tr>
<th>Method Description</th>
<th>Input Parameters</th>
<th>n</th>
<th>Within 2-Fold</th>
<th>Within 3-Fold</th>
<th>Within 10-Fold</th>
<th>$r^2$</th>
<th>AAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADMET mechanistic</td>
<td>Log D, fup, BPR (predicted using ADMET Predictor models)</td>
<td>956</td>
<td>47</td>
<td>65</td>
<td>90</td>
<td>0.25</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>287</td>
<td>51</td>
<td>71</td>
<td>92</td>
<td>0.35</td>
<td>2.4</td>
</tr>
<tr>
<td>ATOM mechanistic</td>
<td>Log D, fup, BPR (predicted using ATOM ML models)</td>
<td>936</td>
<td>45</td>
<td>62</td>
<td>87</td>
<td>0.23</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>285</td>
<td>50</td>
<td>66</td>
<td>92</td>
<td>0.38</td>
<td>2.7</td>
</tr>
<tr>
<td>Allometry (rat and dog)</td>
<td>Rat $V_{D,ss}$, dog $V_{D,ss}$ (predicted using ATOM ML models)</td>
<td>956</td>
<td>45</td>
<td>66</td>
<td>93</td>
<td>0.28</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>283</td>
<td>45</td>
<td>69</td>
<td>96</td>
<td>0.37</td>
<td>2.5</td>
</tr>
<tr>
<td>Allometry (rat)</td>
<td>Rat $V_{D,ss}$, rat fup, fup (human) (predicted using ATOM ML models)</td>
<td>956</td>
<td>30</td>
<td>47</td>
<td>79</td>
<td>0.02</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>283</td>
<td>23</td>
<td>40</td>
<td>76</td>
<td>0.0</td>
<td>4.9</td>
</tr>
<tr>
<td>Allometry (dog)</td>
<td>Dog $V_{D,ss}$, dog fup, fup (human) (predicted using ATOM ML models)</td>
<td>956</td>
<td>28</td>
<td>43</td>
<td>75</td>
<td>0.05</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>283</td>
<td>23</td>
<td>38</td>
<td>77</td>
<td>0.12</td>
<td>4.7</td>
</tr>
<tr>
<td>Direct ML model for human $V_{D,ss}$</td>
<td>SMILES/MOE descriptors</td>
<td>956</td>
<td>36</td>
<td>55</td>
<td>88</td>
<td>0.14</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>285</td>
<td>58</td>
<td>75</td>
<td>98</td>
<td>0.52</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Fig. 2. Summary of model performance of in silico $V_{D,ss}$ prediction methodologies: (A) ATOM in silico set (n = 956 compounds) and (B) ATOM experimental set (n = 254 compounds).
most frequently outperformed other featureization and models investigated in this study. Relative to other in silico methods, mechanistic \( V_{D,ss} \) predictions (both by ATOM and ADMET ML models) and two-species allometry demonstrated superior predictive performance, with 62%–71% of compounds within 3-fold of observed \( V_{D,ss} \) for both data sets (Table 2). In contrast, scaling from single species using allometric methods performed poorly, with only 38%–47% of compounds within 3-fold (Table 2). Trends in predictive performance (such as percentage within 2-, 3-, and 10-fold; AAFE; and Pearson’s \( r^2 \)) across various in silico models were comparable using either the smaller or larger data sets (Table 2, 287 and 956 compounds), with an exception for direct ML model. Predictive performance of the direct ML model to predict \( V_{D,ss} \) increased significantly when the ML model was built using a larger data set (Fig. 3; Table 2). The percentage of compounds within 2-, 3-, and 10-fold increased to 58%, 75%, and 98% from 36%, 55%, and 88%, respectively (Fig. 2; Table 2). The scatter plots of direct ML model predictions are shown in Fig. 3. Additional scatter plots of predicted \( V_{D,ss} \) compared with reported (Lombardo et al., 2018) values across both data sets and various in silico methods are presented in Supplemental Fig. 2.

Experimentally measured log D, \( f_{up} \), and BPR in vitro assays for 254 compounds are summarized in Supplemental Table 2. Although 331 compounds were originally included, some of the compounds showed analytical or recovery issues in different assays and were removed from the data sets. Figure 4 and Table 3 summarize predictive performance of various combinations of experimental data (Supplemental Table 2) as input parameters. Scatter/kernel density estimation plots of mechanistic \( V_{D,ss} \) predictions using various combinations of experimental data (\( f_{up} \), BPR, and log D) as input parameters are shown in Supplemental Fig. 4. The highest percentage of compounds within 3-fold of prediction error was observed when experimentally determined \( f_{up} \) and BPR were used as input parameters, with 81% of the compounds within 3-fold of Lombardo reference values; a good correlation between predicted and observed values (\( r^2 = 0.58 \)) was seen.

Correlation between observed and predicted \( V_{D,ss} \) for 254 compounds using experimental \( f_{up} \) and BPR data as input parameters is shown in Fig. 5. Among the experimental parameters investigated, \( V_{D,ss} \) predictions were sensitive to BPR. \( V_{D,ss} \) predictions within 3-fold dropped to 73% from 81%, and \( r^2 \) reduced from 0.58 to 0.42 when only \( f_{up} \) was used instead of \( f_{up} \) and BPR. In absence of experimental data, assuming BPR as 1 could be recommended, as better performance was observed when the BPR value was assumed to be 1 instead of inputting ML-predicted values (Table 3); 63% of the compounds were predicted within 2-fold when BPR was assumed to be 1, compared with 56% when BPR was predicted from ML models in combination with measured \( f_{up} \). This
significantly different using either adipocyte or myocyte partitioning, percentage of compounds within 3-fold. Among all the methods in which lower AAFEs were observed demonstrated the highest there was a good correlation between AAFE and percentage of compounds within 3-fold. (Table 4).

Across all the prediction methods evaluated using different data sets, there was a good correlation between AAFE and percentage of compounds within 3-fold. Among all the methods investigated, mechanistic VD,ss predictions utilizing measured fup and BPR as input parameters demonstrated superior performance, with lowest AAFE, highest $r^2$, and percentage of compounds within 3-fold.

Discussion

Mechanistic VD,ss Predictions. Kp calculations use physiologic parameters of the tissue and physicochemical properties of the drug to ascertain how compounds partition themselves between plasma and tissue. Based on preliminary evaluations and other reports in the literature (Graham et al., 2012), the Lukacova method (Lukacova et al., 2008) was used as a method of choice for mechanistic VD,ss predictions. Key prerequisite input parameters to predict mechanistic VD,ss are pKa, log D, log P, fup, and BPR. Therefore, estimating these input parameters either by in silico methods or by experimental measurements, and impact of measured parameters on mechanistic VD,ss predictions have been explored.

Mechanistic VD,ss predictions using input parameters predicted by either ATOM ML models or ADMET Predictor demonstrated similar performance across data sets (Table 2). Therefore, either of the two ML models set (ATOM or ADMET Predictor) can be used to predict mechanistic VD,ss in silico. It is important to note that ML models for BPR (ATOM ML or ADMET Predictor (from user manual)) were built using very small data sets (Supplemental Table 1), and predictive performances of ML models to predict BPR are questionable. When predicted BPR values were replaced with experimental data, significant improvement in mechanistic VD,ss predictive performance was observed; $r^2$ increased from 0.38 to 0.51 and percentage within 3-fold increased from 66% to 79%, highlighting the sensitivity of VD,ss predictions to BPR values (Fig. 4; Table 3). As BPR is a key parameter, particularly for calculation of intracellular acidic phospholipid binding of strongly basic drugs, it could be anticipated to improve the predictions. However, impact of BPR measurement was not definitely demonstrated in literature until recently (Yau et al., 2020). The current evaluations (Table 3) clearly demonstrate the importance of measuring BPR in predicting VD,ss, and the need to fill the existing gaps in BPR data sets used to build predictive ML models. It is noteworthy that with only two in vitro measurements (fup and BPR), 81% of compounds are within 3-fold of observed VD,ss (Table 3), with AAFE of 2.0.

Because it can impact both the pharmacokinetics and pharmacodynamics of a drug, fup is measured routinely in drug discovery (Smith et al., 2010). On the other hand, BPR of compounds in the early discovery phase is relatively less routinely measured and might lead to missed opportunities not only in predicting VD,ss (as observed in this study) but also in predicting the impact on overall pharmacokinetics of a compound (Kalamaridis and DiLoreto, 2014). Comparable predictive performance was noted by Chan et al. (2018) using a smaller data set of 152 clinical compounds. They demonstrated that mechanistic VD,ss predictions were accurate or superior to empirical approaches based on the extrapolation of VD,ss from preclinical species (Chan et al., 2018). In addition to superior performance of mechanistic VD,ss prediction methods (using either ML-predicted or experimental input parameters),
a mechanistic approach uniquely offers the ability to calculate partitioning (Kp) of compounds into various tissues.

Allometric Scaling. Traditionally, prediction of human V_{Dss} has relied on scaling of V_{Dss} obtained from preclinical species using allometric equations (Jones et al., 2011). Although allometry has some limitations in predicting distribution of highly protein-bound drugs, it has been a valuable technique to predict human PK parameters to determine first-time-in-human dose (Choi et al., 2019). To leverage existing data from animal studies during early drug discovery, use of ML-predicted V_{Dss} employing allometric scaling from preclinical species was explored. Although there continue to be translational questions about interspecies scaling, it was hypothesized that deployment of this technique could allow for much wider chemical space coverage relative to human V_{Dss} trained models, as well as to provide insight into mechanisms not captured by mechanistic models such as transporter-driven tissue uptake. Although ML models to predict V_{Dss} and fup values in preclinical species have demonstrated good performance (Supplemental Table 1), single-species scaling performed poorly in predicting human V_{Dss} (Table 2, <50% were within 3-fold). This poor performance could be due to magnification of errors in predictions of V_{Dss} and/or fup values in addition to limitations of single-species scaling. Several studies have shown that plasma protein binding corrections significantly enhanced predictive performance of allometric scaling from preclinical V_{Dss} (Zou et al., 2012). As the V_{Dss} predictions are inversely proportional to fup in preclinical species (see Materials and Methods for equations), errors in the predictions of fup values will have a significant impact on V_{Dss} predictions. Therefore, we investigated V_{Dss} comparisons without fup corrections. Direct correlation of predicted dog V_{Dss} (without fup corrections) with human V_{Dss} demonstrated improved performance, with 48%, 65%, and 97% of compounds within 2-, 3-, and 10-fold of observed human V_{Dss}, respectively, when compared with fup accounting for the difference between dog and human (23%, 37%, and 75%, Table 2). This supports that the poor predictive accuracy of the dog fup model magnified the prediction errors. However, similar improved performance or correlations were not observed in the case extrapolating from rat V_{Dss} predictions. In contrast, human V_{Dss} scaled using both rat and dog by the Wajima method demonstrated predictive performance similar to mechanistic models (Table 2). Although overall predictive performance is not significantly different between the two methods, it is noteworthy that mechanistic models were relatively better at predicting anionic compounds within 2-fold compared with the Wajima method (Supplemental Fig. 7). V_{Dss} predictions classified by ionization class across various methods can be found in Supplemental Fig. 6.

Direct ML Models. Previously, we observed that the data set size has a direct impact on model predictivity for several pharmacokinetic related data sets (Minnich et al., 2020). As anticipated, ML models built using smaller data sets, such as that for BPR, showed lower model performance statistics compared with models built using a larger data set (Supplemental Table 1). Furthermore, the direct ML model built on a larger data set (using 970 clinical compounds) outperformed other in silico methods, including the mechanistic V_{Dss} method (Table 2). When utilizing direct ML models built on a larger data set, 75% of compounds (Table 2) were predicted within 3-fold of observed V_{Dss}, with excellent correlation (Fig. 3B). It is important to highlight that the clinical data set is highly diverse across physicochemical, in vitro ADME, and in vivo PK properties (Lombardo et al., 2018). Models built on diverse data sets of chemical space have a greater applicability domain and generalizability (Simeon et al., 2019). Therefore, direct ML predictions of V_{Dss} might be the most computationally efficient and predictive way to process in silico predictions of V_{Dss} for de novo compounds. One limitation of the current model is the relatively small training set, possibly restricting the application of the model to certain chemotypes. In such cases, models that are limited to structurally related analogs may prove more predictive than global models built on a diverse set of compounds (Simeon et al., 2019). Despite some differences in hyperparameters and data set splits used relative to our study, Simeon et al. (2019) demonstrated similar predictive performance for a direct ML model built using a data set of 941 compounds. These independent
studies provide promising evidence of improved performances of direct ML models with enhanced data sets of clinical compounds.

**Predictions Using Adipocyte and Myocyte Cell Partitioning.** Muscle and fat are tissues with larger physiologic volumes (60% of tissue volume), and distribution of compounds to these tissues have a major impact on the V_D,ss of compounds in human (Davies and Morris, 1993). Björkman (2002) evaluated relative contributions of various tissue partition coefficients (Kp, tissues) in predicting V_D,ss in rat and observed an excellent linear correlation (>0.99) between V_D,ss when calculated using only Kp values from muscle and fat. In this study, we hypothesized that intracellular partitioning of compounds into human adipocytes and myocytes in vitro could be used as a surrogate to determine fat and muscle Kp values and subsequently be used to estimate human V_D,ss. In addition, measuring Kp values directly in human cells could improve translation to human tissues. Higher predictive performance was observed, but only when one of the adipocyte partition or myocyte partition values was included to predict V_D,ss (Table 4). Adipocyte and myocyte partition values and predicted V_D,ss were highly correlated (r^2 > 0.7), suggesting that measurement of partition in only one cell type is adequate. Between the two measurements, adipocyte partition (Kp_fat only) showed better performance, particularly with respect to the percentage of compounds within 2-fold compared with myocyte partition (Kp muscle only). Combination of both adipocyte and myocyte partition in various combinations did not provide significant improvement in V_D,ss predictions (Table 4). Although, Kp_fat showed good correlation to human V_D,ss, it failed to predict compounds with low V_D,ss (<1 l/kg) because of volume contributions from other tissues (assumption of Kp = 1) (Supplemental Fig. 5a). Surprisingly, predictive performance was lower when fat and muscle volumes were predicted using both adipocyte and myocyte measured data, and the volume of the remaining tissues was predicted using mechanistic Kp prediction method. Only 56% of the compounds were within 3-fold compared with 63% when Kp was assumed to be 1 for other tissues (Table 4). However, it improved prediction of compounds with low V_D,ss. Measured adipocyte and myocyte partition data provided in Supplemental Table 3 enable further exploration of V_D,ss prediction methods.

**Conclusions**

One of the purposes of comparing various in silico V_D,ss prediction methods was to establish the best in silico approaches to predict V_D,ss for de novo compounds. Based on the extensive comparisons of results across the in silico methods (Table 2), we conclude that 1) the mechanistic V_D,ss prediction methods using a combination of ML models for predicting physicochemical properties paired with mechanistic equations for Kp or 2) the Wajima method employing predicted rat and dog V_D,ss are our recommended in silico approaches to predict human V_D,ss. If a larger training data set of chemically diverse V_D,ss experimental values is available, then direct ML predictions of V_D,ss might be the most computationally efficient and predictive way to process in silico predictions of V_D,ss for de novo compounds. Once these de novo compounds have been synthesized in discovery, it is most useful to externally measure BPR and fup to get to a more accurate estimation of human V_D,ss. Based on our analysis, BPR is the most sensitive physicochemical property to determine V_D,ss in silico. Further, we investigated the utility of adipocyte and myocyte partitioning in predicting V_D,ss. If fat or muscle partition coefficients are being considered as part of the model, adipocyte Kp measurements may provide more predictive power than either myocyte Kp alone or adipocyte and myocyte combined. In summary, the scale of prediction strategies evaluated and size of data sets used in this study are novel and significantly larger than those presented in the literature thus far. In

**TABLE 4**

Performance of V_D,ss prediction methods utilizing adipocyte and myocyte Kp experimental data

<table>
<thead>
<tr>
<th>Method Description</th>
<th>Input Parameters</th>
<th>n</th>
<th>Within 2-Fold</th>
<th>Within 3-Fold</th>
<th>Within 10-Fold</th>
<th>r^2</th>
<th>AAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kp fat only</td>
<td>Kp adipocyte, fup and BPR</td>
<td>189</td>
<td>54</td>
<td>68</td>
<td>97</td>
<td>0.42</td>
<td>2.4</td>
</tr>
<tr>
<td>Kp muscle only</td>
<td>Kp myocyte, fup and BPR</td>
<td>189</td>
<td>41</td>
<td>65</td>
<td>94</td>
<td>0.43</td>
<td>2.6</td>
</tr>
<tr>
<td>Kp fat and muscle</td>
<td>Kp adipocyte, Kp myocyte, fup and BPR</td>
<td>189</td>
<td>36</td>
<td>63</td>
<td>92</td>
<td>0.46</td>
<td>2.9</td>
</tr>
<tr>
<td>Kp separate with exp. logD</td>
<td>Kp adipocyte, Kp myocyte, fup and BPR</td>
<td>189</td>
<td>31</td>
<td>46</td>
<td>83</td>
<td>0.46</td>
<td>3.9</td>
</tr>
<tr>
<td>Lukacova with exp. fup, BPR</td>
<td>Experimental fup and BPR. Other input parameters were predicted by ATOM ML models</td>
<td>189</td>
<td>60</td>
<td>76</td>
<td>95</td>
<td>0.50</td>
<td>2.2</td>
</tr>
</tbody>
</table>
addition, we investigated novel methodologies such as adipocyte and myocyte partitioning in predicting Ves. Finally, we have provided several novel in vitro data sets (e.g., BPR, adipocyte Kp, myocyte Kp) generated using a single protocol for 254 clinical compounds that will enable the research community to further enhance Ves prediction methods. This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes. The authors declare no competing financial interest.

Authorship Contributions
Participated in research design: Polli, Pasikanti, Weber, Murad, Crouch.
Conducted experiments: McComas, Pasikanti.
Performed data analysis: Murad, Madej, Pasikanti, Minnich.
Wrote or contributed to the writing of the manuscript: Pasikanti, Murad, Polli.

References

Address correspondence to: Kishore K. Pasikanti, GlaxoSmithKline (GSK), 1250 S. Collegeville Road, Collegeville, Pennsylvania 19426-0989. E-mail: kishore.k.pasikanti@gsk.com