Research Article

Predictive Models for Maximum Recommended Therapeutic Dose of Antiretroviral Drugs

Michael Lee Branham,[1] Edward A. Ross,[2] and Thirumala Govender[1]

[1] *School of Pharmacy and Pharmacology, University of KwaZulu-Natal, Durban 4001, South Africa*
[2] *School of Medicine, University of Florida, Gainesville, FL 32601, USA*

Correspondence should be addressed to Michael Lee Branham, malimahweh@gmail.com

A novel method for predicting maximum recommended therapeutic dose (MRTD) is presented using quantitative structure property relationships (QSPRs) and artificial neural networks (ANNs). MRTD data of 31 structurally diverse Antiretroviral drugs (ARVs) were collected from FDA MRTD Database or package inserts. Molecular property descriptors of each compound, that is, molecular mass, aqueous solubility, lipophilicity, biotransformation half life, oxidation half life, and biodegradation probability were calculated from their SMILES codes. A training set ($n = 23$) was used to construct multiple linear regression and back propagation neural network models. The models were validated using an external test set ($n = 8$) which demonstrated that MRTD values may be predicted with reasonable accuracy. Model predictability was described by root mean squared errors (RMSEs), Kendall's correlation coefficients (tau), P-values, and Bland Altman plots for method comparisons. MRTD was predicted by a 6-3-1 neural network model (RMSE = 13.67, tau = 0.643, $P = 0.035$) more accurately than by the multiple linear regression (RMSE = 27.27, tau = 0.714, $P = 0.019$) model. Both models illustrated a moderate correlation between aqueous solubility of antiretroviral drugs and maximum therapeutic dose. MRTD prediction may assist in the design of safer, more effective treatments for HIV infection.

1. Introduction

Acquired immunodeficiency syndrome (AIDS) is a degenerative disease of the immune and central nervous systems caused by the human immunodeficiency virus (HIV). There are an estimated 33.2 million people living with HIV/AIDS globally [1–3]. Of this number, 22.5 million are in Sub-Saharan Africa, which represents 67.8% of the global number [3]. Antiretroviral drugs (ARVs) may be classified as nucleoside reverse transcriptase inhibitors (NRTIs), nucleotide reverse transcriptase inhibitors (NtRTI), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PI), and more recently as fusion or integrase inhibitors [4, 5]. Since most ARVs have low aqueous solubility and poor bioavailability, several alternative drug delivery strategies have been proposed to optimize systemic concentrations [6, 7]. The important biopharmaceutical properties that need to be considered for effective ARV

delivery systems might include solubility, pKa, lipophilicity, permeability, stability in biological fluids, gastrointestinal metabolism, and where possible viral reservoir targeting [6, 8, 9]. To overcome suboptimal biopharmaceutical properties, ARVs are often prescribed at high daily doses which increase the occurrence of adverse side effects and toxicities [10, 11]. Combination therapy, comprising at least three anti-HIV drugs, has become a standard treatment of AIDS [12], but here again the potential for adverse side effects and drug-related noncompliance increases. To address these issues, computational methods have been used to predict dose-limiting toxicities of a few antiretroviral drugs [13, 14] or to optimize ARV formulations [15, 16]. The ability to predict maximum therapeutic dose directly from molecular structure is both clinically and scientifically attractive in terms of treatment management and reducing drug development costs [17]. Unfortunately, such models for drugs used in the treatment of AIDS do not yet exist. Accurate prediction

of the MRTD for antiretroviral type compounds would be particularly useful in formulation studies so that clinically relevant extrapolations on drug dissolution and permeability can be made earlier in the drug development process [17–20]. Several recent studies have been conducted to define a relationship between the dose and physicochemical properties of the drug [20, 21], or to investigate the underlying mechanisms of drug toxicity and bioaccumulation [20, 22]. Still, the prediction of optimal dose continues to challenge pharmaceutical scientists because of its complexity and variability between different organisms. Artificial neural networks (ANNs) have emerged as a powerful tool suitable for processing complex relationships between molecular stimuli and biological system responses [23]. Examples include prediction of warfarin maintenance dose [24], gentamicin steady-state plasma concentrations [25], skin permeability [26], and prediction of HIV drug resistance [27]; supporting data for these and other studies suggest the utility of neural network modeling for predicting maximum therapeutic doses.

We believed that since the MRTD estimates are derived from human data, they would provide a more relevant, accurate, and specific estimate for toxic dose levels compared to risk assessment models based on animal data alone. In this article, we predict the MRTD of antiretroviral drugs from their molecular structures using relevant molecular property descriptors and neural network software as a data mining tool. Predictive performance of the models were evaluated and statistically compared with the results obtained clinically or reported in the literature. The application of predictive models in the design of safe, effective antiretroviral drug delivery systems is discussed.

2. Materials and Methods

2.1. Chemoinformatic Software and Modeling Tools. The physicochemical descriptors, molecular weight (MW), aqueous solubility (ASol), and lipophilicity (AlogP) were determined using ALOGPS 2.1. Virtual Computational Chemistry Laboratory, (http://www.vcclab.org/) [28, 29]. Bioaccumulation descriptors, log biotransformation half life (logBioHL), oxidation half life (OxidHL), and biodegradation probability (P[BD]) were determined using EPI Suite v.410 (http://www.epa.gov/oppt/sf/tools/methods.htm) [30, 31]. All inferential statistics and MRTD data analysis was performed using MedCalc v.12 (MedCalc Software bvba, Belgium). Artificial neural network analysis was performed using Tiberius Data Mining Software v.6.1.9 (Tiberius Data Mining Software Ltd. Pty, UK).

2.2. MRTD Training and Validation Datasets. The MRTD of 31 structurally diverse antiretroviral drugs were taken from the FDA MRTD database or package inserts. This "clinical MRTD" dataset was randomly split into training and validation subsets as shown in Table 1. Subsequently, each of the calculated descriptors and clinical MRTD values were correlated by multiple linear regression analysis and the results used to identify statistically significant property descriptors. An error back propagation algorithm was used
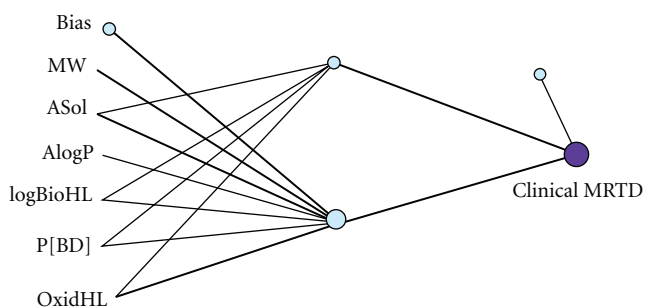

Figure 1: A neural network model was constructed with (2) hidden neuron, (1) output variable that is Clinical MRTD, and (6) input variables that is, OxidHL, P[BD], logBioHL, AlogP, ASol, MW. Total number of patterns (23) were loaded in the data of which 23 were complete and available for training. Two nonlinear neurons were used and the model error minimization was stable for 20 minutes.

for network training, with learning rate set at 0.7. A tangent sigmoid transfer function on the first layer and two neurons with a nonlinear transfer function on the hidden layer were minimalistic structures used to reduce over-fitting. Results of MRTD versus molecular property descriptors with multiple correlation coefficients are listed in Table 2. For the neural network model training, correlation and error statistics are listed in Table 3.

2.3. Model Validation and Statistical Comparisons. The predictability of each of the multiple linear regression (MLR) and neural network (TNN) models was evaluated by a cross-validation procedure [32]. Each model was constructed on the basis of the same training dataset and was subsequently used to predict the excluded test data. Statistical comparisons were performed between the clinical MRTD and predicted MRTD values using the root mean squared error (RMSE), Kendall's correlation coefficient (tau), P-value, and Bland Altman plots for methods comparisons.

3. Results

3.1. Model Predicted versus Actual MRTD Estimates. In this multivariable system, a quantitative relationship between certain molecular property descriptors and maximum recommended therapeutic dose was characterized using two datasets (i.e., MRTD and TEST). A multiple linear regression (Table 2) and a 6-2-1-neural network model (Figure 1) for the prediction ARVs maximum dose were constructed. Table 1 lists summary statistics for the molecular property descriptors and corresponding MRTD values in each dataset. The ARVs in this study, although few in number, covered a broad range (CV equal to 40% or greater) in terms of the physicochemical (MW, ASol, AlogP) and bioaccumulation properties (logBioHL, OxidHL, P[BD]) included. A significant difference in mean ASol between training set (6.1489 g/L) and TEST set (0.436 g/L) was noted although corresponding changes in lipophilic character were not as great. Multiple linear regression analysis was performed with results listed in Table 2. A statistically significant but modest correlation between two of the six descriptors, that is, P[BD]

TABLE 1: Clinical MRTD data consisted of 23 training set compounds (italic) and 8 test set compounds (bold). Mean and standard deviation statistics are shown indicating the distribution and central tendency of each molecular property descriptor.

Drug	MRTD (mg/kg/day)	OxidHL (days)	P[BD] (%)	logBioHL (days)	AlogP (:)	ASol (g/L)	MW (Da)
Acyclovir	13.30	0.135	0.5475	−2.2874	−1.45	8.65	225.21
Ancitabine	20.00	0.139	0.2005	−2.418	−2.62	3.20	225.21
Delaviridine	6.67	0.034	0.0014	−2.3428	2.77	0.086	456.57
Didanosine	6.67	0.140	0.5085	−1.9956	−1.26	6.43	236.23
Famciclovir	25.00	0.051	0.991	−3.6235	0.13	1.32	321.38
Foscarnet	120.00	13.37	0.772	−2.4547	−1.63	16.76	126.01
Indinavir	16.70	0.038	0.0501	−4.6478	3.26	0.048	613.81
Lofexidine	0.040	0.123	0.0792	0.9861	3.31	0.15	259.14
Lamivudine	5.00	0.06	0.0719	−3.9261	−1.29	2.76	229.26
Rimantadine	3.33	0.171	0.4624	0.4539	3.28	0.009	179.31
Ribavirin	200.00	0.264	0.8963	−2.8715	−1.92	33.17	244.21
Valacyclovir	50.00	0.044	0.944	−3.3309	−1.03	1.49	326.41
Zalcitibine	0.0375	0.096	0.0909	−3.5083	−1.29	7.05	211.22
Zanamivir	0.333	0.038	0.8247	−4.4141	−2.29	1.49	332.32
Zidovudine	10.00	0.139	0.0432	−3.0445	−0.1	16.35	267.28
Saquinavir	33.330	0.052	0.9780	−3.9394	4.04	0.002	670.84
Darunavir	53.00	0.094	0.0001	−1.8175	1.76	0.067	547.66
Tipranavir	6.670	0.043	0.0000	−0.0265	5.71	0.0002	602.66
Ritonavir	20.00	0.105	0.9488	−4.61	4.24	0.0012	720.94
Maraviroc	20.00	0.129	0.0700	−0.2247	4.3	0.0106	513.66
Tenofovir	10.00	0.048	0.0016	−3.0098	−1.51	1.87	287.21
Nelfinavir	25.00	0.055	0.6868	−1.6709	6.00	0.0002	567.78
Stavudine	1.333	0.088	0.0768	−3.082	−0.8	40.51	224.21
n	23.00	23.00	23.00	23.00	23.00	23.00	23.00
mean	28.54	0.672	0.402	−2.5133	0.939	6.15	362.72
SD	45.53	2.769	0.394	1.5738	2.809	10.90	178.20
CV%	159.52	412.00	97.00	−62.620	299.04	177.25	48.86
Abacavir	10.00	0.040	0.0229	−2.1757	0.61	1.21	286.38
Emtricitabine	4.00	0.08	0.0566	−3.7784	−0.8	2.00	247.28
Raltegravir	30.07	0.094	0.0006	−2.6157	1.7	0.095	444.47
Nevirapine	3.000	0.167	0.0473	−3.7336	1.75	0.10	266.33
Efavirenz	10.10	0.254	0.0001	0.1111	3.88	0.008	315.67
Fosamprenavir	46.70	0.067	0.0007	−2.4135	0.84	0.068	585.68
Atazanavir	5.35	0.109	0.0202	−3.4392	4.37	0.003	704.96
Lopiravir	53.33	0.103	0.9933	−3.3281	4.07	0.002	614.86
n	8.00	8.00	8.00	8.00	8.00	8.00	8.00
mean	20.32	0.114	0.143	−2.672	2.05	0.436	433.20
SD	20.29	0.067	0.344	1.278	1.878	0.753	180.48
CV%	99.86	58.99	241.300	−47.830	91.50	173.000	41.66

($R_z = 0.427$, $P = 0.035$), ASol ($R_z = 0.476$, $P = 0.014$) and actual MRTD value is noted. MRTD values appear to increase with P[BD], ASol, and OxidHL; but decrease with increasing logBioHL, AlogP, or MW. The multiple correlation coefficient (MCC = 0.7727), residual error (RSD = 33.89), and ANOVA ($P = 0.013$) indicate acceptable predictability of the multiple regression model for ARVs maximum dose. A multiple linear regression equation used for the prediction of MRTD is as follows:

$$-34.3303 - 12.20 * (\text{AlogP}) + 2.1249 * (\text{ASol})$$
$$+15.90 * (\text{logBioHL}) + 0.2159 * (\text{MW}) \qquad (1)$$
$$+5.659 * (\text{OxidHL}) + 46.5133 * (\text{P[BD]}).$$

TABLE 2: The results of the multiple linear regression analysis for antiretroviral drugs versus MRTD. In the training, dataset only two of the six molecular descriptors showed a statistically significant correlation with therapeutic dose, that is, P[BD] ($P = 0.0349$) and ASol ($P = 0.0140$). A multiple regression model equation coefficients are listed with there standard errors, coefficient of determination ($R^2 = 0.5970$), multiple correlation coefficient (MCC = 0.7727) and residual standard deviation (RSD = 33.89).

Multiple linear regression	$n = 23$	$R^2 = 0.5970$	RSD = 33.89	MCC = 0.7727
independent variables	Coefficient	SE	P-value	R_z
Constant	−34.3303			
AlogP	−12.1995	6.6976	0.0873	−0.221
ASol	2.1239	0.7697	0.0140	0.476
logBioHL	15.9000	8.0377	0.0654	−0.071
MW	0.2159	0.1015	0.0494	−0.116
OxidHL	5.6589	2.8478	0.0643	0.448
P[BD]	46.5133	20.1746	0.0349	0.427
ANOVA	F-ratio = 3.9507		$P = 0.013$	

TABLE 3: The results of the neural network analysis for antiretroviral drugs versus MRTDs. All molecular descriptors show weak correlation with MRTD except for ASol, OxidHL, and P[BD]. The 6-2-1 neural network model predicted training set MRTD values with high accuracy ($R^2 = 0.992$, MAX = 13.64, $P < 0.001$). No multicollinearity between independent variables was observed in the training set.

6-2-1 neural network	$n = 23$	$R^2 = 0.992$	MAX = 13.64	RMSE = 5.53
Model versus clinical MRTD	$P < 0.001$		Learning rate = 0.700	
Independent variables	Correlation coefficients		R^2	
AlogP	−0.221		0.049	
ASol	0.476		0.226	
logBioHL	−0.071		0.005	
MW	−0.116		0.013	
OxidHL	0.448		0.201	
P[BD]	0.427		0.182	
ANOVA	F-ratio = 1340.73		$P < 0.001$	

The results of the neural network model for antiretroviral MRTDs is shown in Table 3. All molecular descriptors show weak correlation with MRTD except ASol, OxidHL, and P[BD]. The 6-2-1 neural network predicted training set MRTD values with high accuracy ($R^2 = 0.992$, MAX = 13.64, $P < 0.001$). No multicollinearity between independent variables was observed in the training set. SPSS code for the 6-2-1 neural network may be executed as follows:

```
COMPUTE Var1 = ((OxidHL *
(−0.543086558961242)) +
3.63976611815824)
 + ((Pr_BD_ * (−5.67413921537257)) +
2.81153598121711)
+ ((LogBioHL * (0.40828801067156)) +
0.747514104338025)
+ ((ALogP * (−0.848110602763694)) +
1.43330691867064)
 + ((Sol * (−2.55399028293542E−02)) +
0.517313285798853)
+ ((MW * (1.40343495376914E−02)) −
5.94322423917294)
+ 3.86733992096867.
```

```
COMPUTE Var2 = ((OxidHL *
(8.23491975851945E−02)) −
0.551904322215974)
+ ((Pr_BD_ * (3.12529587401663)) −
1.54858410557524)
+ ((LogBioHL * (0.431493610228662)) +
0.790000076287146)
+ ((ALogP * (−0.04107624611641)) +
6.94188559367329E−02)
+ ((Sol * (4.12997852313067E−02)) −
0.836531279838639)
+ ((MW * (1.05078568602304E−03)) −
0.444983569959981)
+ 9.28741349960935E−02.
COMPUTE Var3 = −0.535249116383622.
Execute.
COMPUTE Var1 = 0.108990429907616 * Var1.
COMPUTE Var2 = 1.44376333322051 *
(Exp(Var2) − Exp(−Var2)) / (Exp(Var2) +
Exp(−Var2)).
COMPUTE Var3 = Var3.
```

TABLE 4: Test dataset Goodness of fit comparisons. Clinical MRTD values from 8 ARVs were using as and external test dataset the validate model predictability. Model performance is characterized here in terms of root means squared error (RMSE), Kendall's correlation coefficient (tau), and Type II error probability (*P*-value).

Drug	MRTD	MLR	SE	TNN	SE
Abacavir	10.00	−10.6738	427.4060	6.4861	12.3474
Emtricitabine	4.00	−23.9239	779.7441	3.3678	0.3996
Raltegravir	30.07	0.0628	900.4320	19.7700	106.0900
Nevirapine	3.00	−54.1862	3270.2614	−16.5700	382.9849
Efavirenz	10.10	−10.2863	415.6012	−5.7782	252.1172
Fosamprenavir	46.70	44.0515	7.0145	50.7500	16.4025
Atazanavir	5.35	11.4360	37.0393	29.8700	601.2304
Lopiravir	53.33	42.6361	114.3594	64.5900	126.7876
Mean	20.32	−0.1105	RMSE = 27.27	19.06	RMSE = 13.67
Kendall's tau		0.714		0.643	
P-value		0.019		0.035	

```
Execute.

COMPUTE Tiberius_MRTD = (((((Var1 +
Var2 + Var3)/2.0) + 0.5) * 199.9625)
+ 0.0375.

Execute.
```

3.2. Model Validation and Statistical Comparisons. Each of the models was then validated using external TEST MRTD dataset and a cross-validation procedure. Model "goodness of fit" and predictability are summarized in Table 4. RMSE of the 6-2-1 neural network (RMSE = 13.67) was substantially less the multiple linear regression model (RMSE = 27.27). Comparison of model predictivity was confirmed with TNN having maximum squared error (SE = 601.23) compared to MLR (SE = 3270.26). Model Figure 3 correlation with the clinical MRTD values using Kendall's correlation coefficient was, however, greater for the MLR (tau = 0.714, *P* = 0.019) than the resultant 6-2-1 neural network model (tau = 0.643, *P* = 0.035). Bland Altman plots for methods comparison further illustrate the predictive value of the neural network (TNN) and multiple linear regression (MLR) models. The upper and lower limits of agreement for the MLR (58.3, −17.4) and TNN (29.8, −27.3) are illustrated in Figure 2. While it can be seen that all of the differences lie between these limits, MLR model limits of agreement were wider than the TNN model, which were more narrow and nearly symmetrical.

4. Discussion

MRTD values and SMILES codes for antiretroviral drugs were collected from the FDA MRTD database which is a highly reliable source pharmacologic activity based on extensive clinical evidence (http://www.fda.gov/cder/). The ALOGPS+ program was used for the calculation of molecular mass (MW), solubility (ASol), and lipophilicity (AlogP) without alteration. The precision and robustness of these tools are well established in the pharmaceutical and modeling community and its predictive power concerning
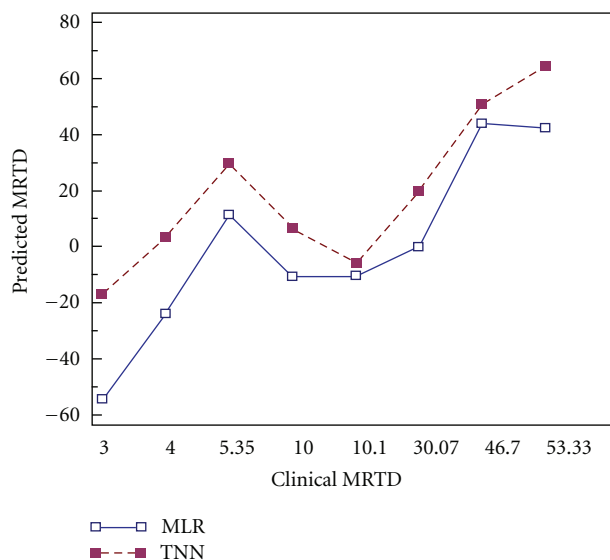

FIGURE 2: Method comparisons graph. MRTD values predicted by multiple linear regression (hollow squares, □) nearly traced those values predicted by the 6-2-1 neural network (solid squares, ■) as shown. Multiple linear regression estimates for MRTD were consistently lower that those predicted by the neural network model.

the input parameters in question are published on the company website (http://www.vcclab.com/) and cited in the manuscript [28, 29, 33]. Independent validation of the parameters calculated with ALOGPS has been published [33, 34] and the estimates were found to be appropriately accurate for the use intended. The bioaccumulation input parameters, oxidation half life (oxidHL), log biotransformation half life (logBioHL), and biodegradation probability (P[BD]) were calculated using the EPI Suite software which is publically available and has been validated in hundreds of modeling experiments for the estimated parameters. EPI Suite software is continuously updated and the predictive power of its latest version is always available at the website (http://www.epa.gov/oppt/sf/tools/methods.htm). None of
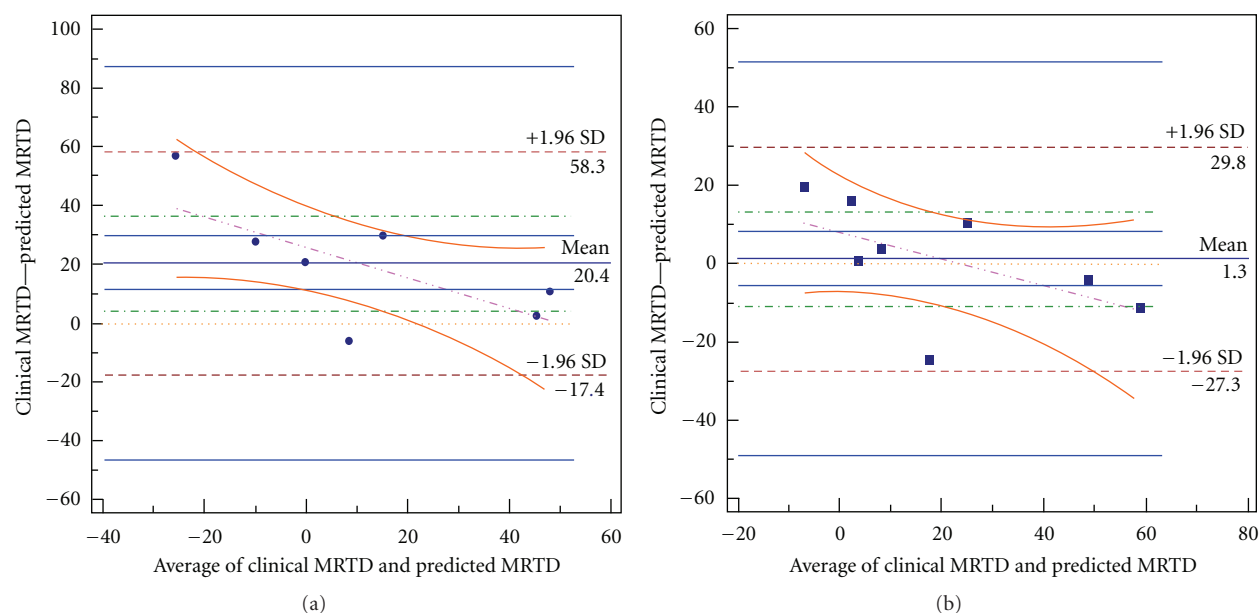
(a)

(b)

FIGURE 3: Bland Altman plot for method comparisons. Bland Altman plots are shown for (a) multiple linear regression predicted versus clinical MRTD, and (b) neural network model predicted versus clinical MRTD. Horizontal lines are drawn at the mean difference, and at the limits of agreement. All predicted values were within limits of agreement for both models, although these limits were more narrow for the neural network model estimates. The plots are useful revealing the relationship between the differences and the averages, the slight deviation from symmetry in the multiple regression indicates some systematic bias but no possible outliers were identified.

the parameters were log-transformations of the software output but instead used as direct input values for the training set. Although some of the predictor values ranged over 3 orders of magnitude, this did not appear to adversely affect comparisons of the MLR or ANN methods. Furthermore, because of the highly diverse nature of antiretroviral compounds in terms of physicochemical descriptors and split-data validation procedure with randomly selected training and test sets instead of cross-validating scheme. It is our intention to adapt the method as new ARV drugs become listed in evolving FDA database.

We began our study looking at dose-related adverse effects of commercial antiretroviral drugs or new ARVs in development. Although the appearance of serious long-term metabolic complications, such as cardiovascular disturbances [35], hyperlipidemia [35, 36], and diabetes, have been extensively reported [35, 37], few reports on computational models to predict these dose-limiting toxicities [38, 39] can be found in the literature. A mathematical model to predict the optimal dosing regimen for AIDS therapy has been reported [40], but in this case, CD4+ cell counts and knowledge of the adherence interval of individual patients is required to adjust the dose. While the model was effective at reducing dose-limiting toxicities in an AIDS patient population, it cannot be applied to nonapproved ARV formulations or to drugs in development. Since the overwhelming majority of anti-HIV drugs demonstrate efficacy over a small range of treatment doses, MRTD predicting models would be beneficial to the drug delivery scientist who needs to design experiments based on the most effective therapeutic dose of the medication [41]. The MRTD is empirically derived

from human clinical trials, thus it is a direct measure of the threshold for dose-related adverse effects in humans. Prediction of the MRTD from molecular structure is of importance for both new and existing drugs which may require modifications to improve their aqueous solubility or bioavailability *in vivo*. ARVs fit this description also and are excellent subject molecules for predictive modeling to estimate maximum dose. Unfortunately, the number of antiretroviral drugs with established MRTD is yet small, so the development of models specifically of ARVs is both tedious and rare.

The molecular descriptors used in this study were selected to represent physicochemical (MW, AlogP, ASol) and bioaccumulation (OxidHL, P[BD], logBioHL) property influences therapeutic dose. Although molecular weight does not strongly correlate with toxicity of most compounds, the larger the molecular size of a compound, the smaller its membrane permeability and diffusion coefficient become [34]. Therefore, compounds with higher weights are less likely to be absorbed, which limits their systemic toxicity. Results of this study predict no correlation between molecular weight and maximum ARV dose. In contrast, bioactivity and drug toxicity almost always increase with increasing lipophilicity. This is due in part to the fact that lipophilic molecules tend to cross cellular membranes more readily increasing exposure and residence in the body. In addition, lipophilic drugs are characterized by increased plasma protein binding, thus an assessment of lipophilicity is almost always included in the physicochemical evaluation of a drug because of its close association with pharmacologic, permeability and potential bioaccumulation [42, 43]. Our

results here did not indicate AlogP or lipophilic character as having any influence on ARV dose. The aqueous solubility of a compound significantly affects its absorption and distribution characteristics. Typically, low solubility goes along with a poor absorption and, therefore, its systemic toxicity reduced, however, local irritability may develop and/or reduced elimination rates, both are characteristic of drugs with low aqueous solubility. For highly potent drugs, increasing solubility usually enhances the elimination rate and lowers systemic half life, [44]. This means that for ARVs with low aqueous solubility, some correlation with maximum therapeutic dose (as we has shown) is likely to be observed.

Any chemical (even water) can produce toxic side effects in the body if allowed to accumulate to sufficiently large concentrations. While much of the effort in bioaccumulation modeling [43, 45] has been initiated by scientists estimating the equilibrium distribution of chemicals between organisms and their environments (e.g., fish-water, plants-soil), effective physicochemical property estimation routines (i.e., PERs) have resulted from these studies that may be applied to similar biodistribution problems in pharmaceutical research. For example, oxidation half life (OxidHL) is an estimate of the molecules ability to form stable hydroxyl radicals or to interact with such moiety under ambient conditions. Formation of hydroxyl-radicals is often associated with dose-limiting toxicities. Although in our studies, oxidation half life was not statistically significant in the prediction of MRTD. Another bioaccumulation descriptor, biotransformation half life (logBioHL), is the (linearized) fraction of drug mass in the whole body that has been metabolized per day. Our estimates do not account for the formation of specific metabolites which may be toxic nor do they identify specific pathways in the process (i.e., phase I redox reactions or phase II type conjugation reactions). Consequently, logBioHL was not correlated with MRTD in the present study and was a weaker bioaccumulation property descriptor in comparison to oxidation half life or biodegradation probability P[BD]. The probability of biodegradation attempts to combine both oxidative and biotransformation susceptibility of the structure to give an estimate of overall persistence. P[BD] estimates are based upon molecular fragments [46–48] and in our investigation showed only moderate correlation with MRTD in both multiple linear regression and neural network models. Each of the models was evaluated for predictive accuracy and by statistical comparison. In terms of predictability, root mean squared errors were larger for the multiple linear regression than the neural network model. Some advantages of ANN over MLR models were illustrated in the current study.

Artificial neural networks (ANNs) are biologically inspired data-mining algorithms which work by detecting the patterns and relationships in data. We used the back propagation rule in which the neural network is trained to map a set of input data by iterative adjustment of the weights. A tangent sigmoid transfer function on the first layer and two neurons with a nonlinear transfer function on the hidden layer were minimalistic structures used to reduce overfitting. Our training processes for TNN were allowed to run until no change in RMSE was observed for 20 minutes, at which point the model was saved. This learning method is commonly used for neural network predictive models given dose-response type data. However, ANNs have several limitations, a major theoretical concern is the "black box" nature of the output, that is, conclusions are generated without mechanistic explanations. ANNs also are limited by the quality of their data and may need to be retrained periodically if its performance changes over time. This is not necessarily counterproductive, since it indicates robustness in the model which adapts to changes in the predictive criteria. Real-time monitoring of the training process is also important since overtraining can easily occur, especially when the datasets are small in size. This is may be one of the unique advantages of real-time visualization of the data-mining process allowing the investigator to make "intermediate evaluations" of model predictability and then continue training until the reliability and accuracy required of the predictions are met.

In conclusion, antiretroviral drugs are a chemically diverse class of compounds in terms of both physicochemical properties and bioaccumulation potential. However, commercial ARVs may be categorized for predictive modeling purposes into two groups based on aqueous solubility and lipophilic character, in which hydrophilic compounds may be administered at higher doses (MRTD) and prediction of their MRTD value may be possible using simple multiple linear regression models. In contrast, the prediction of MRTD values for antiretrovirals with poorer aqueous solubility would be the most effective when the neural network approach is used and when both physicochemical and bioaccumulation property descriptors are available for training. With regard to future studies, ANN represents a promising tool for predicting maximum therapeutic dose, especially for antiretroviral drugs with narrow therapeutic index in the treatment of AIDS.

Abbreviations

AlogP: Calculated octanaol/water partition coefficient
ASol: Aqueous solubility
logBioHL: Biotransformation half life
CV: Coefficient of variation
R^2:   Coefficient of determination
MLR: Multiple linear regression
MRTD: Maximum recommended therapeutic dose
MAX: maximum absolute error
MW: Molecular weight
MCC: Multiple correlation coefficient
OxidHL: Oxidation half life
P[BD]:   Biodegradation probability
TNN: Tiberius Neural Network
QSPR: Quantitative Structure Property Relationships
RSD: Residual standard deviation
RMSE: Root mean squared error
SMILES: Simplified molecular input line entry system
SE: Squared error
SD: Standard deviation
R_z: Zero order correlation coefficient.

References

[1] K. A. Pkowitz, "AIDS: the first 20 years," *New England Journal of Medicine*, vol. 344, no. 23, pp. 1764–1772, 2001.

[2] P. Piot, M. Bartos, P. D. Ghys, and N. Walker, "The global impact of HIV/AIDS," *Nature*, vol. 410, no. 6831, pp. 968–973, 2001.

[3] K. H. Mayer and K. K. Venkatesh, "Antiretroviral therapy as HIV prevention: status and prospects," *American Journal of Public Health*, vol. 100, no. 10, pp. 1867–1876, 2010.

[4] E. de Clercq, "Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV," *International Journal of Antimicrobial Agents*, vol. 33, no. 4, pp. 307–320, 2009.

[5] E. de Clercq, "Antiretroviral drugs," *Current Opinion in Pharmacology*, vol. 10, no. 5, pp. 507–515, 2010.

[6] G. C. Williams and P. J. Sinko, "Oral absorption of the HIV protease inhibitors: a current update," *Advanced Drug Delivery Reviews*, vol. 39, no. 1–3, pp. 211–238, 1999.

[7] E. Ojewole, I. Mackraj, P. Naidoo, and T. Govender, "Exploring the use of novel drug delivery systems for antiretroviral drugs," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 70, no. 3, pp. 697–710, 2008.

[8] C. Y. Wu and L. Z. Benet, "Predicting drug disposition via application of BCS: transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system," *Pharmaceutical Research*, vol. 22, no. 1, pp. 11–23, 2005.

[9] M. A. Cosenza, M. L. Zhao, and S. C. Lee, "HIV-1 expression protects macrophages and microglia from apoptotic death," *Neuropathology and Applied Neurobiology*, vol. 30, no. 5, pp. 478–490, 2004.

[10] C. M. Bates, "HIV medicine: drug side effects and interactions," *Postgraduate Medical Journal*, vol. 72, no. 843, pp. 30–36, 1996.

[11] D. D. Richman, D. M. Margolis, M. Delaney, W. C. Greene, D. Hazuda, and R. J. Pomerantz, "The challenge of finding a cure for HIV infection," *Science*, vol. 323, no. 5919, pp. 1304–1307, 2009.

[12] C. Flexner, "HIV drug development: the next 25 years," *Nature Reviews Drug Discovery*, vol. 6, no. 12, pp. 959–966, 2007.

[13] R. D. Benz, "Toxicological and clinical computational analysis and the US FDA/CDER," *Expert Opinion on Drug Metabolism and Toxicology*, vol. 3, no. 1, pp. 109–124, 2007.

[14] P. Palumbo, H. Wu, E. Chadwick et al., "Virologic response to potent antiretroviral therapy and modeling of HIV dynamics in early pediatric infection," *Journal of Infectious Diseases*, vol. 196, no. 1, pp. 23–29, 2007.

[15] D. Woolfson, M. L. Umrethia, V. L. Kett, and R. K. Malcolm, "Freeze-dried, mucoadhesive system for vaginal delivery of the HIV microbicide, dapivirine: optimisation by an ANN," *International Journal of Pharmaceutics*, vol. 388, no. 1-2, pp. 136–143, 2010.

[16] N. Subramanian, A. Yajnik, and R. S. Murthy, "Artificial neural network as an alternative to multiple regression analysis in optimizing formulation parameters of cytarabine liposomes," *AAPS PharmSciTech*, vol. 5, no. 1, p. E4, 2004.

[17] J. F. Contrera, E. J. Matthews, N. L. Kruhlak, and R. D. Benz, "Assessment of the health effects of chemicals in humans: I. QSAR estimation of the maximum recommended therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data," *Current Drug Discovery Technologies*, vol. 1, no. 1, pp. 61–76, 2004.

[18] J. Röling, H. Schmid, M. Fischereder, R. Draenert, and F. D. Goebel, "HIV-associated renal diseases and highly active antiretroviral therapy-induced nephropathy," *Clinical Infectious Diseases*, vol. 42, no. 10, pp. 1488–1495, 2006.

[19] Y. Sun, Y. Peng, Y. Chen, and A. J. Shukla, "Application of artificial neural networks in the design of controlled release drug delivery systems," *Advanced Drug Delivery Reviews*, vol. 55, no. 9, pp. 1201–1215, 2003.

[20] E. J. Matthews, N. L. Kruhlak, R. D. Benz, and J. F. Contrera, "Assessment of the health effects of chemicals in humans: I. QSAR estimation of the maximum recommended therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data," *Current Drug Discovery Technologies*, vol. 1, no. 1, pp. 61–76, 2004.

[21] J. F. Blake, "Chemoinformatics—predicting the physicochemical properties of "drug-like" molecules," *Current Opinion in Biotechnology*, vol. 11, no. 1, pp. 104–107, 2000.

[22] D. Butina, M. D. Segall, and K. Frankcombe, "Predicting ADME properties in silico: methods and models," *Drug Discovery Today*, vol. 7, no. 11, pp. S83–S88, 2002.

[23] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22, no. 5, pp. 717–727, 2000.

[24] I. Solomon, N. Maharshak, G. Chechik et al., "Applying an artificial neural network to warfarin maintenance dose prediction," *Israel Medical Association Journal*, vol. 6, no. 12, pp. 732–735, 2004.

[25] B. W. Corrigan, P. R. Mayo, and F. Jamali, "Application of a neural network for gentamicin concentration prediction in a general hospital population," *Therapeutic Drug Monitoring*, vol. 19, no. 1, pp. 25–28, 1997.

[26] C. W. Lim, S. Fujiwara, F. Yamashita, and M. Hashida, "Prediction of human skin permeability using a combination of molecular orbital calculations and artificial neural network," *Biological and Pharmaceutical Bulletin*, vol. 25, no. 3, pp. 361–366, 2002.

[27] S. Drăghici and R. B. Potter, "Predicting HIV drug resistance with neural networks," *Bioinformatics*, vol. 19, no. 1, pp. 98–107, 2003.

[28] I. V. Tetko, "The WWW as a tool to obtain molecular parameters," *Mini Reviews in Medicinal Chemistry*, vol. 3, no. 8, pp. 809–820, 2003.

[29] I. V. Tetko, V. Y. Tanchuk, and A. E. P. Villa, "Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices," *Journal of Chemical Information and Computer Sciences*, vol. 41, no. 3–6, pp. 1407–1421, 2001.

[30] A. Arnot, W. Meylan, J. Tunkel et al., "A quantitative structure-activity relationship for predicting metabolic biotransformation rates for organic chemicals in fish," *Environmental Toxicology and Chemistry*, vol. 28, no. 6, pp. 1168–1177, 2009.

[31] A. Arnot and F. A. P. C. Gobas, "A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in fish," *Environmental Reviews*, vol. 14, no. 4, pp. 257–297, 2006.

[32] C. Cobelli, E. R. Carson, L. Finkelstein, and M. S. Leaning, "Validation of simple and complex models in physiology and medicine," *The American Journal of Physiology*, vol. 246, no. 2, pp. R259–266, 1984.

[33] R. Mannhold, G. I. Poda, C. Ostermann, and I. V. Tetko, "Calculation of molecular lipophilicity: state-of-the-art and

comparison of logP methods on more than 96,000 compounds," *Journal of Pharmaceutical Sciences*, vol. 98, no. 3, pp. 861–893, 2009.

[34] M. Yazdanian, S. L. Glynn, J. L. Wright, and A. Hawi, "Correlating partitioning and Caco-2 cell permeability of structurally diverse small molecular weight compounds," *Pharmaceutical Research*, vol. 15, no. 9, pp. 1490–1494, 1998.

[35] R. G. Jain, E. S. Furfine, L. Pedneault, A. J. White, and J. M. Lenhard, "Metabolic complications associated with antiretroviral therapy," *Antiviral Research*, vol. 51, no. 3, pp. 151–177, 2001.

[36] S. K. Gan, K. Samaras, A. Carr, and D. Chisholm, "Antiretroviral therapy, insulin resistance and lipodystrophy," *Diabetes, Obesity and Metabolism*, vol. 3, no. 2, pp. 67–71, 2001.

[37] M. J. Glesby, "Toxicities and adverse findings associated with protease inhibitors," in *Protease Inhibitors in AIDS Therapy*, R. C. Ogden and C. W. Flexner, Eds., pp. 237–256, Marcel Dekker, Inc., New York, NY, USA, 2001.

[38] P. A. Pham and C. Flexner, "Emerging antiretroviral drug interactions," *Journal of Antimicrobial Chemotherapy*, vol. 66, no. 2, pp. 235–239, 2011.

[39] D. Wang, B. Larder, A. Revell et al., "A comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy," *Artificial Intelligence in Medicine*, vol. 47, no. 1, pp. 63–74, 2009.

[40] O. Krakovska and L. M. Wahl, "Optimal drug treatment regimens for HIV depend on adherence," *Journal of Theoretical Biology*, vol. 246, no. 3, pp. 499–509, 2007.

[41] S. Hongzong, Y. Shuping, Z. Kejun, F. Aiping, D. Yun-Bo, and H. Zhide, "Quantitative structure activity relationship study on EC50 of anti-HIV drugs," *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 1, pp. 15–24, 2008.

[42] B. Testa, P. Crivori, M. Reist, and P. A. Carrupt, "The influence of lipophilicity on the pharmacokinetic behavior of drugs: concepts and examples," *Perspectives in Drug Discovery and Design*, vol. 19, pp. 179–211, 2000.

[43] J. W. Nichols, M. Bonnell, S. D. Dimitrov, B. I. Escher, X. Han, and N. I. Kramer, "Bioaccumulation assessment using predictive approaches," *Integrated Environmental Assessment and Management*, vol. 5, no. 4, pp. 577–597, 2009.

[44] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Advanced Drug Delivery Reviews*, vol. 23, no. 1–3, pp. 3–25, 1997.

[45] W. B. Neely, D. R. Branson, and G. E. Blau, "Partition coefficient to measure bioconcentration potential of organic chemicals in fish," *Environmental Science and Technology*, vol. 8, no. 13, pp. 1113–1115, 1974.

[46] S. Dimitrov, T. Pavlov, D. Nedelcheva et al., "A kinetic model for predicting biodegradation," *SAR and QSAR in Environmental Research*, vol. 18, no. 5-6, pp. 443–457, 2007.

[47] S. D. Dimitrov, N. C. Dimitrova, J. D. Walker, G. D. Veith, and O. G. Mekenyan, "Predicting BCFs of highly hydrophobic chemicals: effects of molecular size," *Pure and Applied Chemistry*, vol. 74, no. 10, pp. 1823–1830, 2002.

[48] P. Howard, W. Meylan, D. Aronson et al., "A new biodegradation prediction model specific to petroleum hydrocarbons," *Environmental Toxicology and Chemistry*, vol. 24, no. 8, pp. 1847–1860, 2005.