International Journal of
Molecular Sciences

MDPI

Article

In Silico Prediction of Intestinal Permeability by Hierarchical Support Vector Regression

Ming-Han Lee [1], Giang Huong Ta [1], Ching-Feng Weng [2] and Max K. Leong [1,*]

[1] Department of Chemistry, National Dong Hwa University, Shoufeng, Hualien 974301, Taiwan; 610512018@gms.ndhu.edu.tw (M.-H.L.); 810812203@gms.ndhu.edu.tw (G.H.T.)
[2] Department of Basic Medical Science, Center for Transitional Medicine, Xiamen Medical College, Xiamen 361023, China; cfweng-cfweng@hotmail.com
* Correspondence: leong@gms.ndhu.edu.tw; Tel.: +886-3-890-3609

Received: 8 April 2020; Accepted: 17 May 2020; Published: 19 May 2020

Abstract: The vast majority of marketed drugs are orally administrated. As such, drug absorption is one of the important drug metabolism and pharmacokinetics parameters that should be assessed in the process of drug discovery and development. A nonlinear quantitative structure–activity relationship (QSAR) model was constructed in this investigation using the novel machine learning-based hierarchical support vector regression (HSVR) scheme to render the extremely complicated relationships between descriptors and intestinal permeability that can take place through various passive diffusion and carrier-mediated active transport routes. The predictions by HSVR were found to be in good agreement with the observed values for the molecules in the training set (n = 53, r^2 = 0.93, q^2_{CV} = 0.84, RMSE = 0.17, s = 0.08), test set (n = 13, q^2 = 0.75–0.89, RMSE = 0.26, s = 0.14), and even outlier set (n = 8, q^2 = 0.78–0.92, RMSE = 0.19, s = 0.09). The built HSVR model consistently met the most stringent criteria when subjected to various statistical assessments. A mock test also assured the predictivity of HSVR. Consequently, this HSVR model can be adopted to facilitate drug discovery and development.

Keywords: intestinal permeability; passive diffusion; active transport; in silico; quantitative structure–activity relationship; hierarchical support vector regression

1. Introduction

Oral administration is the predominant route for medication that can be manifested by the fact that ca. 56% of unique drugs approved by FDA in 2018 were orally administrated [1]. Accordingly, drug absorption is one of critical absorption, distribution, metabolism and excretion, and toxicity (ADME/Tox) factors that should be taken into consideration in the process of drug discovery and development as well as practical applications [2]. For instance, curcumin, which is the major constituent of the spice turmeric (*Curcuma longa*), has a great beneficial potential in treating cancer, diabetes, osteoarthritis, antianxiety, and even novel coronavirus disease 2019 (COVID-19) [3,4] and yet its practical clinical applications are very limited mainly due to its poor absorption [5]. Clinically, tuberculosis (TB) is one of the leading causes of death globally, especially for HIV/AIDS patients [6], and the survival of extremely ill TB patients is diminished due to the poor absorption of anti-TB agents [7].

Drug absorption mainly relies on solubility and intestinal permeability [8], which is also termed as intestinal absorption [9], since oral drugs must permeate the gastrointestinal barrier before they can be absorbed by the bodies [9]. In fact, solubility and permeability have been adopted by the biopharmaceutics drug disposition classification system (BDDCS), which suggests that the intestinal permeability rate is closely correlated with the extent of metabolism [10] Nevertheless, intestinal permeability is an extremely complicated process since drugs can pass through the intestinal epithelium

to enter blood vessel by active transport as well as passive diffusion, as illustrated by Figure 3 of Dahlgren and Lennernäs [11]. Mechanistically, the active transport can be mediated by two superfamilies expressed in the intestine, namely the influx transporters of the solute carrier (SLC) family and the efflux transporters of the ATP-binding cassette (ABC) family, whereas the passive diffusion can take place through the transcellular and/or paracellular routes [12].

In addition, the ABC transporters including P-glycoprotein (P-gp, MDR1, *ABCB1*), breast cancer resistance protein (BCRP, *ABCG2*), MRP2 (*ABCC2*), MRP3 (*ABCC3*), MRP4 (*ABCC4*), MRP5 (*ABCC5*), MRP6 (*ABCC6*), MRP7 (*ABCC10*), MRP8 (*ABCC11*), and MRP9 (*ABCC12*) [13], and the SLC transporters involving peptide transporter 1 (PepT1, *SLC15A1*), concentrative nucleoside transporter 1 (CNT1, *SLC28A1*), concentrative nucleoside transporter 2 (CNT2 *SLC28A2*), equilibrative nucleoside transporter (ENT2, *SLC29A2*), organic cation transporters 1 (OCT1, *SLC22A1*), organic cation/carnitine transporter 1 (OCTN1, *SLC22A4*), organic cation/carnitine transporter 2 (OCTN2, *SLC22A5*), monocarboxylate Transporter 1 (MCT1, *SLC16A1*), organic anion transporting polypeptide 2B1 (OATP2B1, *SLC02B1*), serotonin transporter (SERT, *SLC6A4*), and apical sodium-dependent bile acid transporter; (ASBT. *SLC10A2*) [14] can be found in the intestine. Their expression levels can be different in varied segments of intestine [15,16].

Of various in vitro assay systems to measure intestinal permeability, human colorectal adenocarcinoma cells (Caco-2), Madin–Darby canine kidney cells (MDCK), and parallel artificial membrane permeability assay (PAMPA) are commonly used [9], and they can be affected by factors such as cell line types and cultured conditions. The in situ single-pass intestinal perfusion (SPIP) model is the most prevalent assay [17] that normally measures effective permeability (P_{eff}) of the gastrointestinal (GI) tract segments, namely duodenum, jejunum, ileum, and colon, in human, rat, and mouse [18]. The parameter P_{eff}, which is expressed as cm/s, can be calculated by

$$P_{\text{eff}} = \frac{-Q \ln\left(C'_{\text{out}}/C'_{\text{in}}\right)}{A} \quad (1)$$

where Q is the perfusion buffer flow rate; C'_{out} and C'_{in} are the outlet and inlet solute concentrations, respectively; and A represents the surface area within the intestinal segment that can be computed by the radius of the intestinal segment (R) and the length of the perfusion intestinal segment (L) [19],

$$A = 2\pi RL \quad (2)$$

When compared with in vitro assays, in vivo tests provide a closer to real-life environment, but they are costly, time consuming, and sometimes inhumane, and are subjected to discrepancies by a number of factors such as individual differences in intestinal cell surface and epithelial cell integrity [20], especially they are very sensitive to the animal species because of differences in physiological conditions [21]. More importantly, those factors can make substantial contribution to data inhomogeneity that, in turn, can create paramount obstacles to producing a sound quantitative theoretical model based on the data compiled from the public domain since only homogenous data can produce a good in silico model [22].

In silico technologies have been seamlessly integrated into the drug discovery and development and they especially provide valuable advantages in ADME/Tox profiling due to their extremely fast throughput and low cost [23]. As such, it is plausible to expect an in silico model that can predict intestinal permeability is very useful. Nevertheless, no sound quantitative structure–activity relationship (QSAR) model has been published to date despite, even though some qualitative studies have been conducted. The scarcity in QSAR model can be plausibly attributed to the lack of consistent and homogenous data in the public domain and, more importantly, the extremely complex process of intestinal permeability (vide supra) since it can take place through various active transport and passive diffusion routes. More specifically, the SLC transporters can enhance the drug uptake into the intestine and hence increase drug absorption, whereas ABC proteins can elevate drug efflux out

of intestine and therefore reduce drug absorption [24], leading to problematic situations for model development. For instance, the substrates of PepT1 and P-gp, which are two of the most abundant SLC and ABC transporters, respectively, in jejunum [15], can interact with their transporter proteins by hydrogen-bond donor (HBD) [25,26], suggesting that HBD can simultaneously promote and hinder intestinal permeability. As such, traditional or machine learning (ML) modeling schemes are not sophisticated enough to manage such exceedingly nonlinear situations.

Accordingly, it is extraordinarily difficult, if not entirely infeasible, to develop a robust in silico model to predict the intestinal permeability with the consideration of those critical factors governing the perplexing efflux and influx active transport and passive diffusion mentioned above. Such challenge, nevertheless, can be solved by the novel ML-based hierarchical support vector regression (HSVR) scheme devised by Leong et al. [27] since HSVR can properly depict the complicated and inconstant dependencies of descriptors that can be greatly attributed to the fact that HSVR has the advantageous features of both a local model and a global model, namely larger coverage of applicability domain (AD) and a higher degree of predictivity, respectively. Unlike most predictive models, which are vulnerable to the "mesa effect", i.e., give mediocre performances when applied to extrapolated predictions, HSVR can substantially minimize such performance deterioration, as demonstrated elsewhere [1,28,29], suggesting that HSVR is insusceptible to outliers in contrast to the other predictive models that is of crucial importance to a theoretical model [30]. Herein, this investigation was aimed at developing an accurate, rapid, and predictive in silico model based on the HSVR scheme to predict the intestinal permeability to facilitate drug discovery and development.

2. Results

2.1. Data Partition

Of all molecules selected in this study, 53 and 13 molecules were randomly assigned to the training set and test set, respectively, with a ca. 4:1 ratio. Figure 1 shows the projection of all molecules enrolled in this investigation in chemical space, spanned by the first three principal components (PCs), which characterized 97.7% of the total variable variance. It can be observed that the training samples and test samples showed similar distributions in the chemical space. Furthermore, the high degrees of biological and chemical similarity between both datasets can also be observed in Figure S1, which displays the histograms of log P_{eff}, molecular weight (MW), log D, log P, hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD), and polar surface area (PSA) in density form for the training and test molecules. Thus, the unbiased partition of data samples can be ascertained [31].

It is not trivial to establish the AD of predictive models prior to model development to identify the outliers and exclude them from data collection [32]. There are various versions to define AD [33]. This study adopted the version based on the chemical similarity/dissimilarity using principal component analysis (PCA) to graphically assess the outliers [32]. Those designated outliers, conversely, are very dissimilar from the training samples, as manifested by the fact that they are totally situated outside the perimeter of the training set in the chemical space shown in Figure 1. In fact, the differences between the outliers and the other samples can be realized by the fact that their surface areas are unanimously more than 600 Å^2, whereas surface areas of the others are less than 600 Å^2.

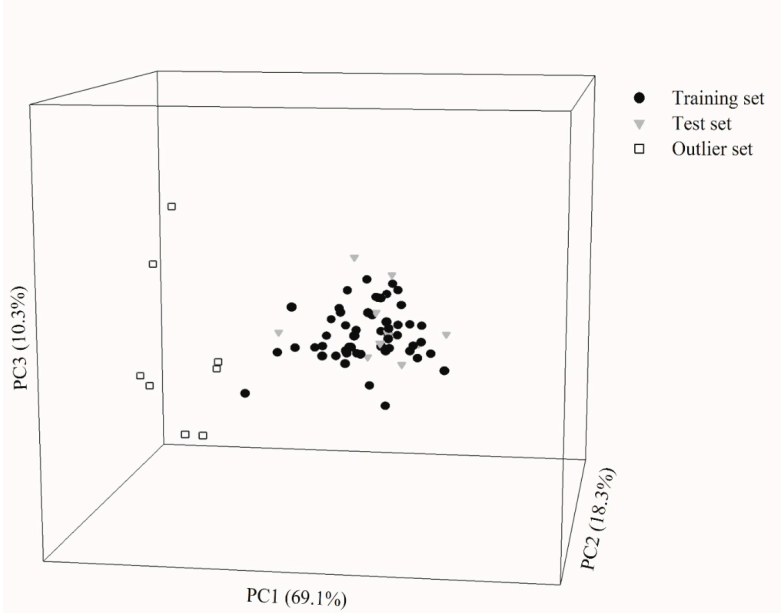

Figure 1. Molecule distribution for the molecules enrolled in this investigation in the training set (solid circle), test set (gray triangle), and outlier set (open square) in the chemical space spanned by three principal components.

2.2. SVRE

Various SVR models were built based on various descriptor combinations and runtime parameters, and three SVR models, denoted as SVR A, SVR B, and SVR C, were selected to construct the SVR ensemble, which, in turn, was subjected to regression by another SVR to generate the HSVR model. The optimal runtime parameters of SVR A, SVR B, SVR C, and HSVR, are listed in Table S2.

These three SVR models, which unanimously selected four descriptors with different combinations (Table 1), were adopted based on their individual performances on the molecules and statistical assessments in the training set and test set. Table S1 lists their predictive log P_{eff} values. Tables 2 and 3 summarize the corresponding statistical assessments of these three SVR models in the training set and test set, respectively.

Table 1. Descriptors selected as the input of SVR models in the ensemble, the correlation coefficient (r) with log P_{eff}, and their descriptions.

Descriptor	SVR A	SVR B	SVR C	r	Description
μ	X †	X	X	−0.16	Dipole moment of molecule
Log D	X	X		0.23	Logarithm of the n-octanol–water partition coefficient at PH 6.5
Log P			X	0.22	Logarithm of the n-octanol–water partition coefficient
HBD		X	X	−0.07	Hydrogen-bond donor
n_{N+O}	X			−0.29	Number of nitrogen and oxygen
Shadow-v	X			0.23	Ratio of largest to smallest dimension
MR		X	X	−0.12	Sum of molar refractivity of substituents

† Selected.

Table 2. Statistic evaluations, namely correlation coefficient (r^2), maximal absolute residual (Δ_{Max}), mean absolute error (MAE), standard deviation (s), RMSE, leave one out cross-validation correlation coefficient (q^2_{CV}), and average correlation coefficient of Y-scrambling ($<r_s^2>$) evaluated by SVR A, SVR B, SVR C, and HSVR in the training set.

	SVR A	SVR B	SVR C	HSVR
r^2	0.62	0.83	0.79	0.93
Δ_{Max}	1.18	0.91	1.42	0.50
MAE	0.33	0.25	0.15	0.16
s	0.26	0.15	0.27	0.08
RMSE	0.42	0.29	0.31	0.17
q^2_{CV}	0.12	0.02	0.07	0.84
$\langle r_s^2 \rangle$	0.02	0.02	0.02	0.02

Table 3. Statistic evaluations, correlation coefficients q^2, q^2_{F1}, q^2_{F2}, and q^2_{F3}, concordance correlation coefficient (CCC), maximal absolute residual (Δ_{Max}), mean absolute error (MAE), standard deviation (s), and RMSE evaluated by SVR A, SVR B, SVR C, and HSVR in the test set.

	SVR A	SVR B	SVR C	HSVR
q^2	0.40	0.67	0.73	0.81
q^2_{F1}	0.15	0.55	0.66	0.75
q^2_{F2}	0.15	0.55	0.66	0.75
q^2_{F3}	0.50	0.73	0.79	0.85
CCC	0.55	0.82	0.86	0.89
Δ_{Max}	0.99	0.68	0.64	0.47
MAE	0.39	0.26	0.24	0.22
s	0.29	0.24	0.20	0.14
RMSE	0.47	0.35	0.30	0.26

Figure 2 displays scatter the plot of observed versus the predicted log P_{eff} values by SVR A, SVR B, SVR C, and HSVR for the molecules in the training set. The predictions by SVR A, SVR B, and SVR C are generally in good agreement with the observed values for most of the molecules in the training set, as depicted by their small MAE and s values (Table 2). Furthermore, Figure 2 shows that most of the points predicted by SVR B mostly lie on or are closer to the regression line when compared with SVR A and SVR C. As such, SVR B generated the lowest Δ_{Max} (0.91), MAE (0.25), s (0.15), and RMSE (0.29) and the largest r^2 parameter (0.83), suggesting that SVR B executed better than SVR A and SVR C for the molecules in the training set. Nevertheless, SVR B produced not only the lowest q^2_{CV} value (0.02) but also the largest difference between r^2 and q^2_{CV} (0.81) when subjected to the leave-one-out cross-validation (Table 2), signifying its high level of overtraining that, in turn, can severely limit its practical application. SVR A, SVR B, and SVR C unanimously gave rise to the miniature $<r_s^2>$ values of 0.02 (Table 2) when subjected to the Y-scrambling, and their almost zero values of $<r_s^2>$ apparently depict that there is little chance correlation associated with those SVR models [34].

The predictions by SVR A, SVR B, and SVR C in the test set are in modest agreement with the experimental values, as shown in Figure 3, which displays scatter the plot of observed versus the predicted log P_{eff} values by SVR A, SVR B, SVR C, and HSVR for the molecules in the test set. Nevertheless, the mean absolute errors calculated by SVR A, SVR B, and SVR C increase from 0.33, 0.25, and 0.15 in the training set to 0.39, 0.26, and 0.24 in the test set, respectively (Table 3). The other statistical assessments also indicate the performance deteriorations of these three models in the SVRE from the training set to the test set (Tables 2 and 3).

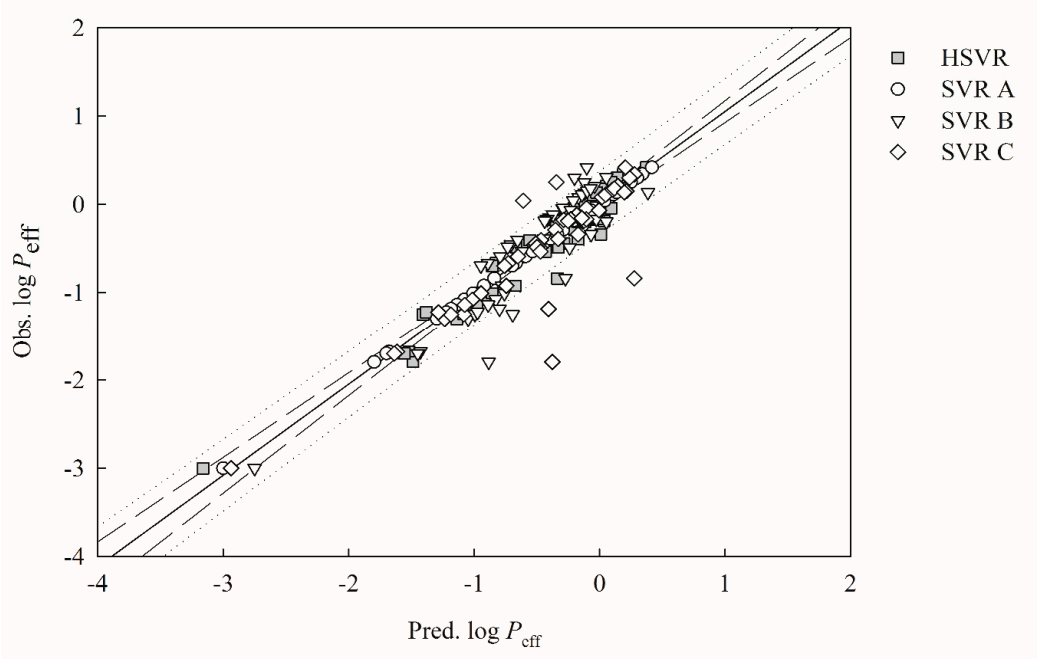

Figure 2. Observed log P_{eff} vs. the log P_{eff} predicted by SVR A (open circle), SVR B (open triangle), HSVR SVR C (open diamond), and HSVR (gray square) for the molecules in the training set. The solid line, dashed lines, and dotted lines correspond to the HSVR regression of the data, 95% confidence intervals for the HSVR regression, and 95% confidence interval for the prediction, respectively.

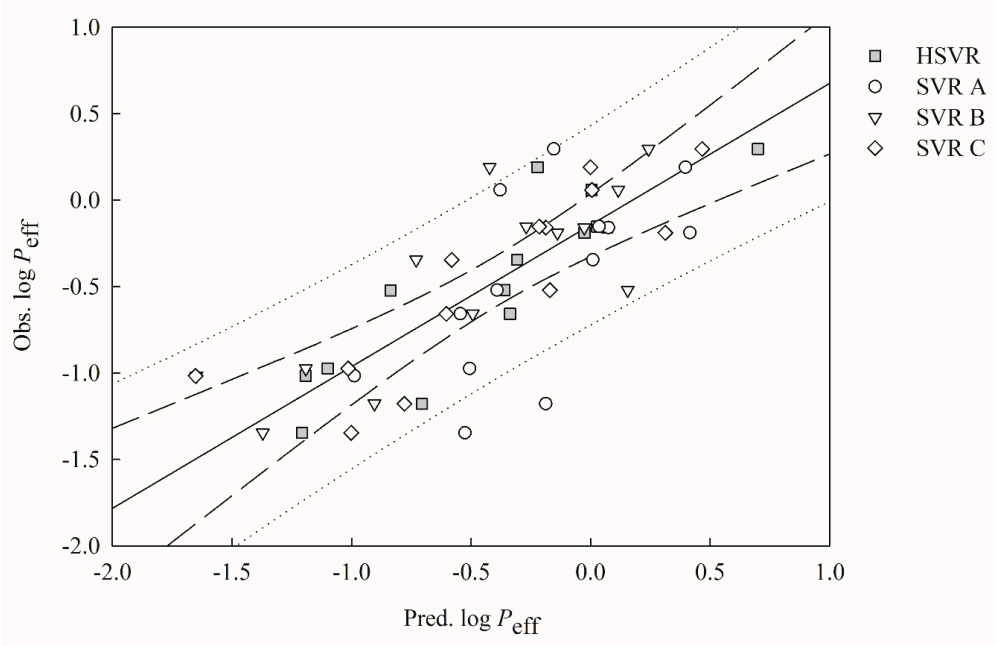

Figure 3. Observed log P_{eff} vs. the log P_{eff} predicted by SVR A (open circle), SVR B (open triangle), HSVR SVR C (open diamond), and HSVR (gray square) for the molecules in the test set. The solid line, dashed lines, and dotted lines correspond to the HSVR regression of the data, 95% confidence intervals for the HSVR regression, and 95% confidence interval for the prediction, respectively.

Furthermore, SVR A, SVR B, and SVR C yielded q^2 values of 0.40, 0.67, and 0.73 in the test set, which are smaller than their r^2 counterparts in the training set. Most notably, the difference between r^2 and q^2 evaluated by SVR A was 0.22, indicating the overtraining nature of SVR A that is also confirmed by its q^2_{F1}, q^2_{F2}, q^2_{F3}, and *CCC*.

Significant performance decreases can be observed when those SVR models in the ensemble were applied to the molecules in the outlier set, as depicted by the statistical parameters listed in Table 4. SVR A, SVR B, and SVR C, for instance, gave rise to q^2_{F1} values of −0.10, 0.04, and 0.47, respectively, which differ greatly from their r^2 values in the training set (Table 2). In addition, they showed larger distances between points and the regression line in the outlier set when compared with their counterparts in the training set, as displayed in Figure 4. Such substantial variations in performance can be realized by the fact most predictive models are vulnerable to the "mesa effect", which leads to a predictive model executing poorly when applied to extrapolated predictions [35].

Table 4. Statistic evaluations, correlation coefficients q^2, q^2_{F1}, q^2_{F2}, and q^2_{F3}, concordance correlation coefficient (*CCC*), maximal absolute residual (Δ_{Max}), mean absolute error (MAE), standard deviation (*s*), and RMSE evaluated by SVR A, SVR B, SVR C, and HSVR in the outlier set.

	SVR A	SVR B	SVR C	HSVR
q^2	0.34	0.63	0.72	0.83
q^2_{F1}	−0.10	0.04	0.47	0.78
q^2_{F2}	−0.11	0.04	0.47	0.78
q^2_{F3}	0.58	0.64	0.80	0.92
CCC	0.34	0.40	0.65	0.89
Δ_{Max}	0.97	0.64	0.52	0.31
MAE	0.33	0.36	0.26	0.17
s	0.29	0.19	0.16	0.09
RMSE	0.43	0.40	0.30	0.19

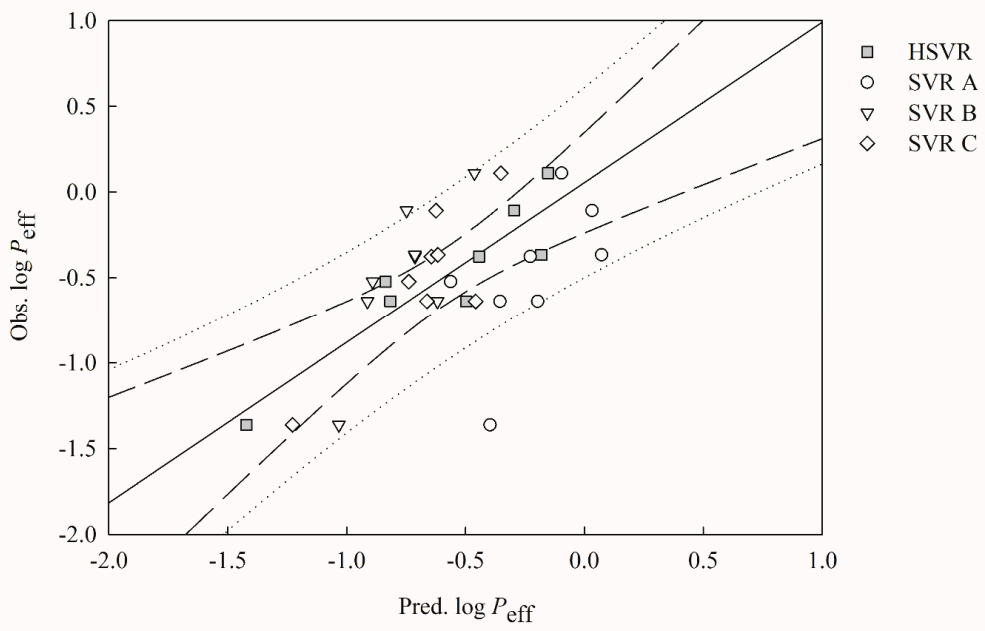

Figure 4. Observed log P_{eff} vs. the log P_{eff} predicted by SVR A (open circle), SVR B (open triangle), HSVR SVR C (open diamond), and HSVR (gray square) for the molecules in the outlier set. The solid line, dashed lines, and dotted lines correspond to the HSVR regression of the data, 95% confidence intervals for the HSVR regression, and 95% confidence interval for the prediction, respectively.

2.3. HSVR

The HSVR model was yielded by the regression of the SVR ensemble based on the predictions of all samples and statistical evaluations in the training set (Table S1 and Table 2) and its optimal runtime conditions are listed in Table S2. HSVR generally executed better than SVR A, SVR B, and SVR C for those training samples, as illustrated by Figure 2, in which it can be observed that the distances between

the predictions by HSVR and regression line generally fall in the range between the largest ones and smallest ones produced by its SVR counterparts in the ensemble. Nevertheless, HSVR predicted better than all of the models in the SVRE in some cases as demonstrated by the prediction of compound **2** (aloin), in which SVR A, SVR B, SVR C, and HSVR yielded absolute residuals of 0.25, 0.20, 0.06, and 0.03, respectively. As such, HSVR statistically functioned better than SVR A, SVR B, and SVR C, as manifested by all parameters listed in Table 2 except MAE values, which were 0.15 and 0.16 for SVR C and HSVR, respectively. In addition, HSVR gave rise to the largest r^2 value (0.93) as compared with its counterparts in the SVRE. In addition, there is a little chance that HSVR was produced by chance correlation, as manifested by its nearly zero value of $<r_s^2>$ (0.02) [34].

When applied to the molecules in the test set, slight performance decreases can be observed for HSVR. For instance, RMSE increased from 0.17 in the training set to 0.26 in the test set (Tables 2 and 3). Nevertheless, the parameter Δ_{Max} declined from 0.50 in the training set to 0.47 in the test set. Figure 3 shows that HSVR performed better than SVR A, SVR B, and SVR C in the test set. The performance dominance of HSVR can be further confirmed by those statistical evaluations listed in Table 3. For example, SVR A, SVR B, SVR C, and HSVR generated s values of 0.29, 0.24, 0.20, and 0.14, respectively. Similar observation that HSVR gave rise to smaller absolute residuals than its counterparts in the SVRE can also be noted in the test set. The absolute prediction deviation of compound **59**, for instance, was 0.04 yielded by HSVR, whereas SVR A, SVR B, and SVR C gave rise to the absolute residuals of 0.35, 0.38, and 0.24, respectively. HSVR normally generated consistent and small errors in both training and test sets, as depicted by those parameters listed in Tables 2 and 3, when compared with its SVR counterparts in the ensemble. Moreover, HSVR yielded the largest q^2 (0.81) in the test set and the smallest difference between r^2 and q^2_{CV} (0.09), suggesting that HSVR was well-trained or no overfitting effect was observed because it would otherwise generate a significant difference between r^2 and q^2 or between r^2 and q^2_{CV}.

When applied to the outliers, HSVR even showed more pounced predominance, as indicated by those statistical parameters listed in Table 5, from which it can be recognized that HSVR generated the largest q^2 values and smallest deviation-related parameters. The superiority of HSVR in the outlier set can be plausibly due to the broad applicability domain encompassed by HSVR as compared with its SVR counterparts in the ensemble and, more importantly, the more robust nature of HSVR makes it more practically useful in real applications [30].

Table 5. Validation verification of HSVR based on prediction performance of the molecules in the training set, test set, and outlier set.

	Training Set	Test Set	Outlier Set
r_o^2	0.94	0.75	0.83
k	1.04	0.99	0.87
r'^2_o	0.93	0.80	0.78
r_m^2	0.88	0.66	0.76
r'^2_m	0.90	0.74	0.67
$\langle r_m^2 \rangle$	0.89	0.70	0.71
Δr_m^2	0.02	0.08	0.08
Equation (16)	X †	X	X
Equation (17)	X	N/A	N/A
Equation (18)	X	X	X
Equation (19)	X	X	X
Equation (20)	X	X	X
Equation (21)	X	X	X
Equation (22)	N/A [a]	X	X

† Fulfilled; [a] Not available.

2.4. Predictive Evaluations

Figure 5 illustrates the scatter plots of the residual vs. the log P_{eff} values predicted by HSVR for the molecules in the training set, test set, and outlier set. It can be conceived that the residuals are approximately evenly allocated on both sides of *x*-axis along the range of predicted values in all datasets, suggesting that there is no significant systematic error associated with HSVR. The unbiased predictions can be further rendered by its almost negligible average residuals that were 0.00, −0.10, and 0.09 in the training set, test set, and outlier set, respectively (Table S1).

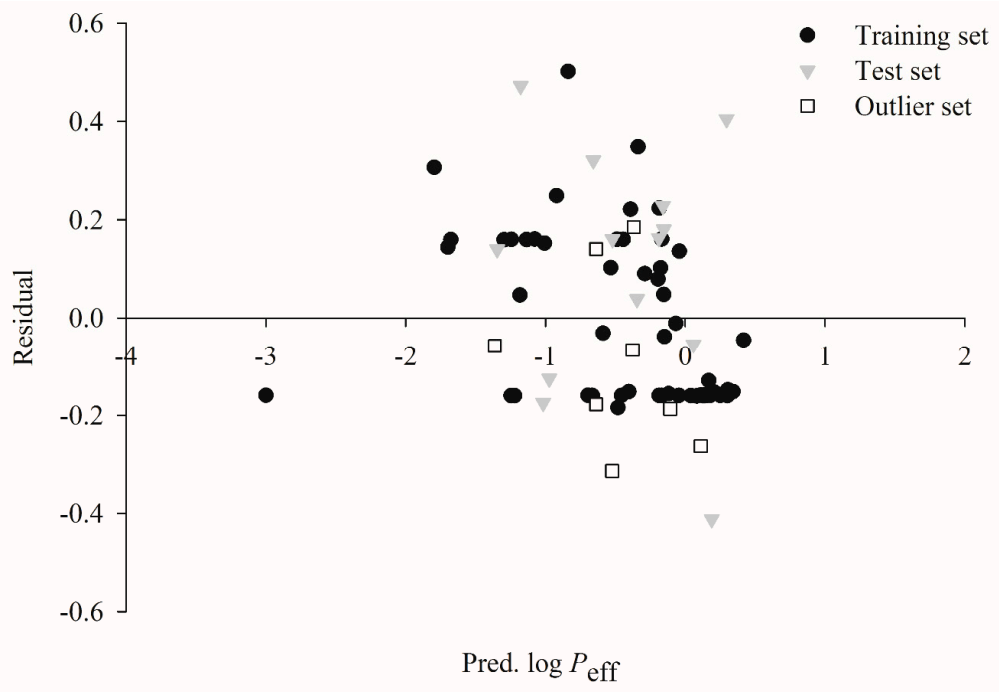

Figure 5. Residual vs. the log P_{eff} predicted by HSVR in the training set (solid circle), test set (gray triangle), and outlier set (open square).

The derived HSVR model was further assessed by combining the most rigorous validation criteria collectively suggested by Golbraikh et al. [36], Ojha et al. [37], Roy et al. [38], and Chirico and Gramatica [39] in the training set, test set, and outlier set (Equations (16)–(22)). The results are listed in Table 5, from which it can be found that HSVR showed similar high levels of performance in those three datasets. More importantly, HSVR completely fulfilled all validation requirements, suggesting that this predictive model is highly accurate and predictive.

2.5. Mock Test

To imitate real world challenges, the derived HSVR model was checked by all marketed drugs assayed by Lennernäs [40], of which seven were also included in this investigation, producing a good way to calibrate the challenging system. However, Lennernäs measured the P_{eff} values by the SPIP in human jejunum in contrast to the compounds enrolled in this study, whose P_{eff} values were obtained using the rat jejunum segment. Thus, those drugs assayed by Lennernäs are not suitable as the test set or second external set since their validation assessments (vide supra) are not applicable to these drugs. The subsequent relationship between both measured systems was initially instituted and checked based on those common seven drugs and the resulted scattered plot is displayed in Figure 6. The results show that both systems were modestly correlated with each other with an *r* value of 0.80, suggesting that it is plausible to validate the derived HSVR model by those novel molecules assayed by Lennernäs, which is consistent with the fact that the rat SPIP P_{eff} values can be useful to predict human intestinal permeability [17].

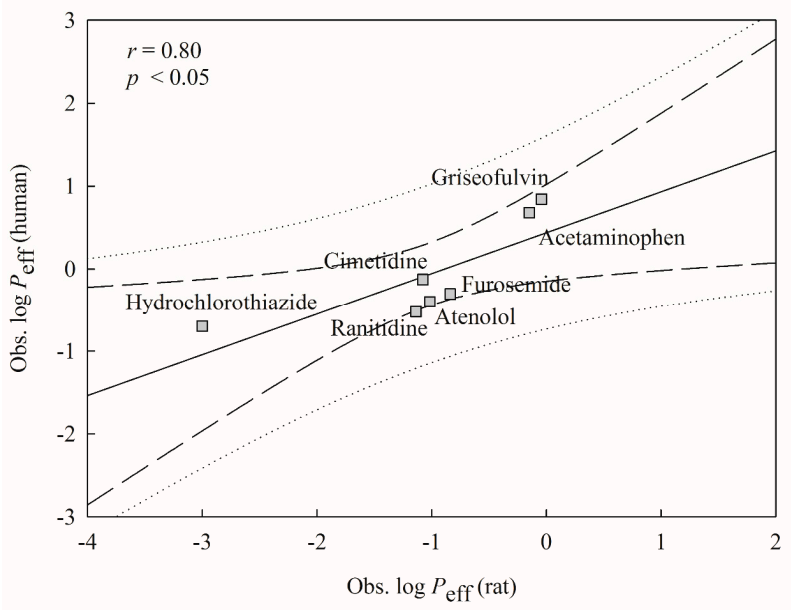

Figure 6. Observed human log P_{eff} versus observed rat log P_{eff} for the molecules in the mock test. The solid line, dashed lines, and dotted lines correspond to the mock test regression of the observed data, 95% confidence interval for the mock test regression, and 95% confidence interval for the observation, respectively.

Figure 7 displays the tested results of 11 novel drugs. It can be observed that the *r* value between experimental human log P_{eff} and predicted rat log P_{eff} was 0.79. The negligible difference between both numbers (0.80 vs. 0.79) suggests that HSVR can almost reproduce the experimental observations. Accordingly, this mock challenge by 11 marketed drugs apparently assured the predictivity of HSVR and it is plausible to adopt this HSVR model as a surrogate for preliminary estimation of human intestine permeability in the process of drug discovery and development.

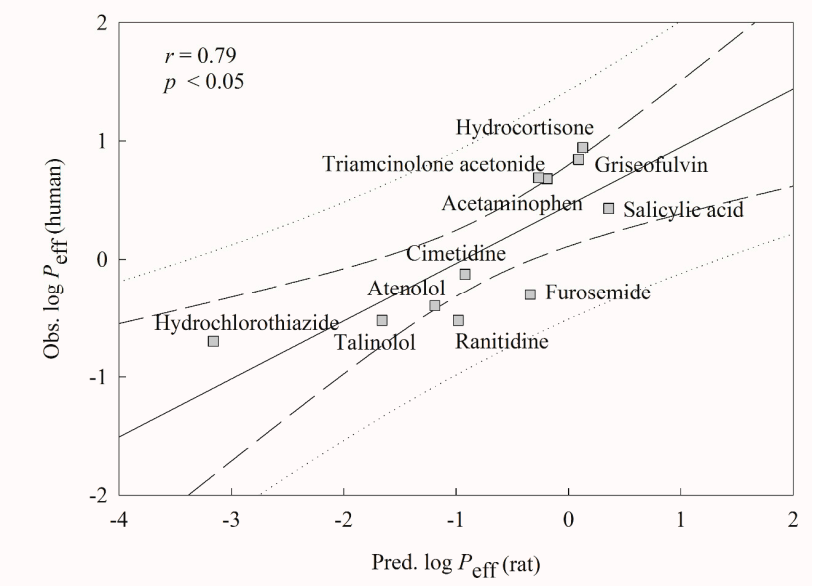

Figure 7. Observed human log P_{eff} versus the rat log P_{eff} predicted by HSVR for the molecules in the mock test. The solid line, dashed lines, and dotted lines correspond to the HSVR regression of the data, 95% confidence interval for the HSVR regression, and 95% confidence interval for the prediction, respectively.

3. Discussion

Numerous in silico models have been reported to predict intestinal permeability [41–55]. However, those published models were based on data assayed by different experimental conditions or different measured parameters, and some of them were qualitative models, making the direct comparison between HSVR and those models extremely difficult. In addition, intestine permeability is an extremely complicated process, which can take place through various active transport and passive diffusion routes (vide supra). As such, it is not uncommon to observe that various descriptor combinations associated with intestinal permeability have been identified. For example, Shultz proposed the significance of HBD, topological polar surface area (TPSA), and log *P* in intestinal permeability [56], whereas Broccatelli et al. recognized the contributions of TPSA, MW, HBD, number of rotamers (n_{rot}), charge, and fraction ionized at pH 7.4 ($cFI_{7.4}$) to intestinal permeability [57].

Drugs must pass through the hydrophobic mucus layer, which is adjacent to the intestinal wall, before they can penetrate through the intestinal epithelial cells [58]. As such, hydrophobicity is of critical importance in intestinal permeability and it can be represented by the *n*-octanol–water partition coefficient (log *P*) and the *n*-octanol–water partition coefficient at pH 6.5 (log *D*). Moreover, it was proposed by Balimane et al. that log *P* and log *D* should be adopted to predict the intestinal permeability since log *P* alone is not sufficient enough to accurately render this complicated process [9]. As such, both log *P* and log *D* were adopted by this study (Table 1). However, the selection of both descriptors can plausibly lead to an overtrained model since the correlation coefficient between log *P* and log *D* was 0.73 for all molecules included in this study. This controversial issue can be eliminated by the fact that log *D* was adopted by SVR A and SVR B, whereas log *P* was selected by SVR C, depicting the fact that no single SVR model included both descriptors simultaneously. In fact, this dilemma of selecting both correlated descriptors to accurately predict intestinal permeability cannot be resolved by any other traditional linear or machine learning-based QSAR schemes but only by any ensemble-based scheme such as HSVR.

It has been observed that PSA is profoundly implicated in membrane permeability in passive diffusion [59], which is completely consistent with the PAMPA study [1] as well as intestinal permeability [56]. In addition, permeability also relies on MW, as proposed in [13]. Nevertheless, neither PSA nor MW was adopted by any of the SVR models in the ensemble (Table 1). Conversely, it is seemingly unusual to observe that the descriptor n_{N+O} was selected by SVR A and yet has hitherto not been adopted by any reported study. These discrepancies can be realized by the fact that n_{N+O} was modestly correlated with PSA and MW with *r* values of 0.88 and 0.71, respectively, for all molecules selected in this study. The empirical observation indicated that models with the selection of n_{N+O} performed better than those with the selection of PSA or MW (data not shown), plausibly due to the descriptor–descriptor interaction [1], suggesting that it is plausible to represent PSA or MW by n_{N+O}. The negative correlation between n_{N+O} and log P_{eff} (−0.29) is also consistent with the fact that permeability can decrease with MW [60].

It has been postulated that hydrogen bond, which can be characterized by HBA and HBD, plays a critical role in intestinal P-gp-mediated transport [61] and HBD makes substantial contributions to intestinal permeability when compared with its HBA counterpart [56]. Accordingly, HBD was adopted in this study. Nevertheless, the relationship between HBD and P_{eff} is seemingly obscure, as illustrated by Figure 8, which shows the average P_{eff} for each histogram bin of HBD for all molecules included in this investigation. This peculiar relationship can be plausibly attributed to the fact that the substrates of PepT1 and P-gp, which are the most abundant SLC and ABC transporters, respectively, in jejunum [15], can interact with their transporter proteins via HBD [25,26]. The complexity can be further increased by taking into the account the fact that P-gp inhibitors, modulators, and substrates can interact with P-gp through HBD [26,62,63]. As such, HBD can simultaneously facilitate and hinder intestinal permeability, leading to a perplexing dependency, which, in turn, can create an unsurmountable hurdle for creating a predictive theoretical model regardless of traditional linear or machine learning-based schemes.

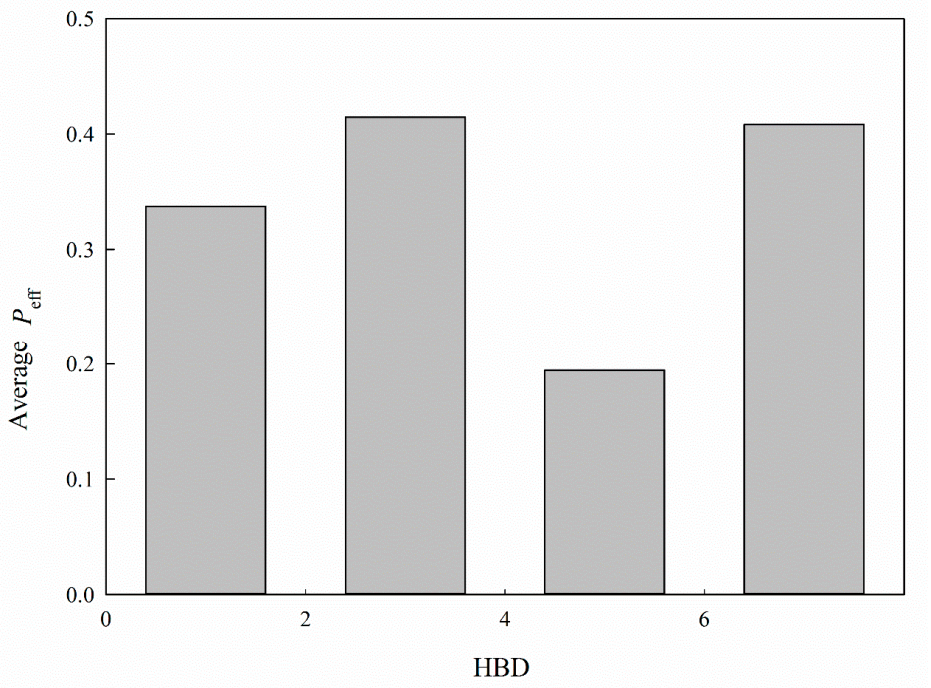

Figure 8. Average P_{eff} vs. the distribution of HBD.

Shadow-*ν* is a size-related descriptor which measures the ratio of largest to smallest dimension. It can be observed in Figure 9, which displays the average P_{eff} for each histogram bin of shadow-*ν*, that P_{eff} generally increased with shadow-*ν* for all molecules selected in this investigation, suggesting that molecules with larger shadow-*ν* have higher permeability than their smaller counterparts.

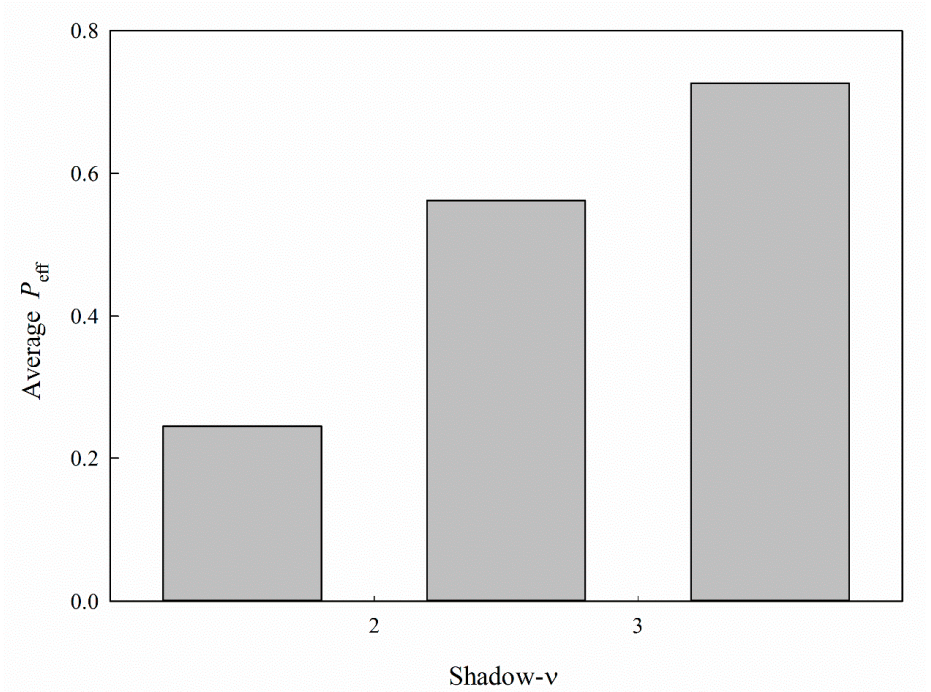

Figure 9. Average P_{eff} vs. the distribution of shadow-*ν*.

It has been observed that molar refractivity (MR), which is possibly associated with molecular size, polarity, and/or polarizability [64], is closely related to ligand-P-gp interactions [65,66]. Nevertheless,

little correlation manifested between MR and log P_{eff} for all molecules enrolled in this study, with an insignificant r value of −0.12 (Table 1). This incongruity can be resolved by the nonlinearity between MR and P_{eff}, as demonstrated in Figure 10, which illustrates the average P_{eff} for each histogram bin of MR. It can be observed that P_{eff} marginally increased with MR and substantially decreased afterwards, suggesting the nonlinear relationship between MR and P_{eff}. Thus, linear models cannot properly render such a complicated relationship.

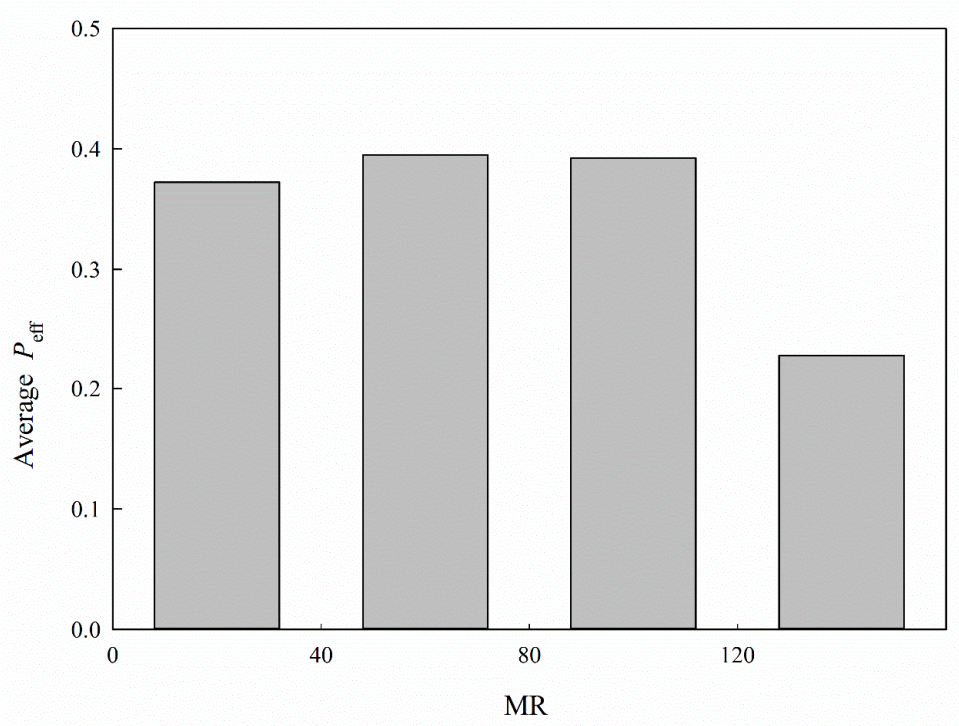

Figure 10. Average P_{eff} vs. the distribution of MR.

The significance of the descriptor μ in intestinal permeability has been recognized [67] since μ can describe the solute-solute and solute–solvent dipole interactions [68], as demonstrated in PAMPA permeability [1], leading to nonlinear relationship between μ and permeability. In addition, it has been observed that ligands can interact with the efflux transporter P-gp and the influx transporter PepT1 through dipole interactions [69–71], giving rise to the complex role played by μ in intestinal permeability.

It is of interest that most of descriptors selected in this study are associated with passive diffusion, which is consistent with the fact that passive diffusion is the major route for intestinal permeability for many administrated drugs [12]. Additionally, MR, shadow-ν, and n_{N+O}, which was selected in place of MW in this study (vide supra), are closely linked to molecular size, and the molecular size is a determining barrier factor in intestinal permeability as postulated [72–74].

CSA, which is also another characteristic associated with molecular size, has been implicated in membrane permeability [75]. Figure 11 exhibits the average P_{eff} for each histogram bin of CSA. It can be observed that P_{eff} did not show substantial variations with CSA initially, yet the P_{eff} value greatly dropped once CSA was larger than 75, which is very similar to the trend observed for MR (Figure 10), suggesting that it is more difficult to penetrate the intestinal wall once the CSA values are larger than a threshold. Nevertheless, the empirical observation has indicated that HSVR with the selections of MR executed better than those with the selection of CSA (data not shown), presumably because of the descriptor–descriptor interaction [1].

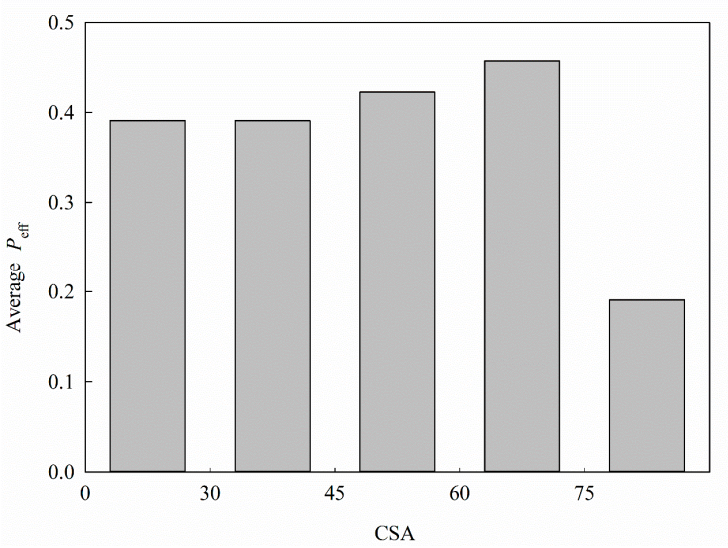

Figure 11. Average P_{eff} vs. the distribution of CSA.

It has been indicated that ion class is one of critical factors in physiological-based pharmacokinetic (PBPK) models and ADME/Tox properties that should be taken into account [20,76]. Actually, it has been demonstrated that neutral compounds show higher passive diffusion [1]. Accordingly, all molecules enrolled in this investigation were subjected to ion class analysis. Figure 12 displays the box plot of the log P_{eff} minimum, maximum, mean, median, the 25th percentile, and the 75th percentile for each ion class. The log P_{eff} values of neutral compounds are statistically greater than the other ion classes, depicting that neutral compounds show higher intestinal permeability. It is possible to improve the compound's intestinal permeability of the other ion classes by chemical modification to produce neutral compounds when they show low intestinal permeability.

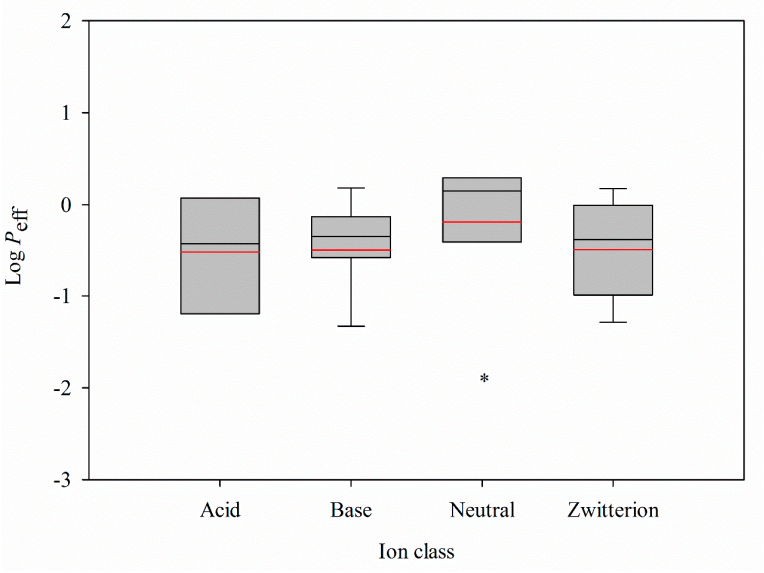

Figure 12. Box plot of log P_{eff} values for different ion classes, where the boxes indicate the distribution of log P_{eff} from the 25th to the 75th percentile, the black and red lines represent the median and mean values, the whiskers depict the minimum and maximum values, and the asterisk denotes significant difference between neutral and the others ($p < 0.05$).

Initially, massive attempts were made in this investigation to construct various 2-QSAR models by adopting numerous partial least squares (PLSs), but no acceptable models were yielded

(data not shown) [29]. This challenge was due to little correlation between the selected descriptors and log P_{eff} for those molecules selected in this study and the largest absolute maximum r was merely 0.29 between n_{N+O} and log P_{eff} (Table 1), depicting the highly nonlinear relationship between them. More importantly, the substantial difference in 2-QSAR development between the passive diffusion, viz. the PAMPA system, and intestinal permeability can be greatly attributed to the significant and complex roles played by those active (influx and efflux) transporters. As such, it is extremely difficult to build a sound linear model to predict intestinal permeability. Conversely, the accurate and predictive HSVR model can considerably delineate such nonlinear dependence of log P_{eff} on descriptors.

4. Materials and Methods

4.1. Data Compilation

A comprehensive literature search was carried out to retrieve in vivo permeability data from a variety of sources to construct quality data for this investigation. Of various assay systems, 74 compounds, which were measured by SPIP in rat jejunum with pH 6.5 phosphate buffered saline (PBS), were adopted from various sources [77–98]. The mean value was taken to assure better consistency if there were more than one P_{eff} values for a given compound within close range. Chemical structures without defined stereochemistry (e.g., racemates) were discarded from the collection. All molecules included in this study, IUPAC names, SMILES strings, CAS registry numbers, logarithm of observed P_{eff} values, and references to the literature are listed in Table S1.

4.2. Descriptor Enumeration

All molecules selected in this study were subjected to full geometry optimization using the density functional theory (DFT) B3LYP method with the basis set 6-31G(*d*,*p*) by the *Gaussian* package (Gaussian, Wallingford, CT) since it has been demonstrated elsewhere that predictive models with the selection of DFT descriptors can perform better [29]. To mimic the assay conditions, the water solvent system was considered by the polarizable continuum model (PCM) [99,100]. The minimum of optimized geometry on the potential energy surface was verified by force calculations in case no imaginary frequency was obtained. Furthermore, atomic charges were also determined by the molecular electrostatic potential-based method of Merz and Kollman [101]. The highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), free energy (ΔG), molecular dipole (μ), and its maximum absolute components ($|\mu|_{Max}$) of each molecule were also retrieved from the optimization calculations.

More than 200 one-, two-, and three-dimensional molecular descriptors, which can be categorized as electronic descriptors, spatial descriptors, structural descriptors, thermodynamic descriptors, topological descriptors, and *E*-state indices, were evaluated by the *Discovery Studio* package (BIOVIA, San Diego, CA, USA) and *E-Dragon* (available at the website: http://www.vcclab.org/lab/edragon/). Additionally, the cross-sectional area (CSA) was also computed using the method modified by Muehlbacher et al. since it was implicated in membrane permeability [102]. Molecules were further placed into four classes, namely neutral, zwitterion, acid, and base, by their pK_a values. Specifically, molecules with only one pK_a value are defined as neutrals, whereas those with more than one pK_a value are designated as zwitterions, acids, and bases when their largest pK_a values are larger than 7 and the smallest pK_a values are smaller than 7; the largest and smallest pK_a values are smaller than 7; and the largest and smallest pK_a values are larger than 7, respectively.

Descriptor filtration was initiated by discarding those descriptors missing for at least one sample or barely displaying discrimination against most of samples. It was suggested by Topliss and Edwards that the probability of spurious correlations can be reduced by removing those descriptors with intercorrelation values of $r^2 > 0.80$ [103]. Accordingly, the Spearman's matrix was constructed and a more conservative threshold of $r^2 \geqq 0.64$ was adopted. Descriptor normalization, which was designated

to reduce the possibility of descriptors with broader ranges outweighing those with narrower ranges, was implemented by

$$\chi_{ij} = \left(x_{ij} - \left\langle x_j \right\rangle\right) / \left[\sum_{i=1}^{n} \left(x_{ij} - \left\langle x_j \right\rangle\right)^2 / (n-1)\right]^{1/2} \quad (3)$$

where x_{ij} and χ_{ij} present the jth original and normalized descriptors of the ith sample, respectively; $\left\langle x_j \right\rangle$ is the mean value of the original jth descriptor; and n is the number of samples.

Descriptors are one of the critical factors influencing the performance of predictive models [104]. The initial descriptor selection was executed by genetic function approximation (GFA) using the QSAR module of *Discovery Studio* due to its effectiveness and efficiency [105]. Further selection was done by the recursive feature elimination (RFE) scheme, in which the model was repeatedly built by all but one of descriptors. Their contributions to the predictive performance were then evaluated, and the descriptor with the smallest contribution was discarded [106].

4.3. Sample Partition

It is of necessity to identify the outliers prior to the model development, which can be done by examining the molecular distribution in the chemical space [107]. As such, the chemical space was initially constructed based on principal components (PCs) using the Diverse Molecules/Principal Component Analysis module of *Discovery Studio* and the outliers were determined. The remaining compounds were randomly divided into the training set and test set with a ca. 4:1 ratio as proposed to build and validate the models, respectively [108], using the Diverse Molecules/Library Analysis module of *Discovery Studio*. In addition, it was suggested by Golbraikh et al. that only high levels of chemical and biological similarity between the training samples and test samples can develop a sound model [36]. As such, the data distribution was carefully examined to guarantee the high levels of biological and chemical similarity in both datasets.

4.4. Hierarchical Support Vector Regression

Hierarchical support vector regression (HSVR), which was originally proposed by Leong et al. [27], was evolved from support vector machine (SVM), which was originally invented by Vapnik et al. [109]. Initially, SVM was devised for classification and then modified for regression, termed support vector regression (SVR) [110]. The most distinguished difference between SVR and HSVR is that the latter can simultaneously take into account the characteristics of a global model, viz. broader applicability domain (AD), and a local model, viz. higher predictivity, as compared with the former [28]. More importantly, it has been demonstrated by a number of studies that predictive models based on the HSVR scheme perform extremely well [1,26–28].

The principle and implementation of HSVR has been detailed elsewhere and the architecture of HSVR schemeis depicted in Figure 1 of Leong et al. [27]. Briefly, SVR models were developed by the *svm-train* module in *LIBSVM* (software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm) using those samples in the training set with various descriptor combinations and SVR run parameters. The derived models were validated by *svm-predict* using those samples in the test set. The regression functions, namely ε-SVR and γ-SVR, were tried, and radial basis function (RBF) was selected as the kernel function due to its simplicity and greater performance [111]. The runtime parameters including regression modes ε-SVR and ν-SVR, the corresponding ε and ν, cost C, and the kernel width γ were automatically run by the systemic grid search algorithm in a parallel fashion.

The principle of Occam's razor was applied to the construction of SVR ensemble (SVRE), which suggests that the fewer the ensemble members, the better [112]. More specifically, two SVR models were selected to develop an SVRE, which was further subjected to regression by another SVR to give rise to the final HSVR model. The construction of two-member SVREs was repeated until the HSVR model executed well. Otherwise, three- or even four-member ensembles would be built by

adding one or more SVR models, respectively, in the case all two-member ensembles failed to show good performance.

4.5. Predictive Assessment

The difference between the observed value (y_i) and the predicted value ($\hat{y}_i$), viz. the residual, was calculated

$$\Delta_i = y_i - \hat{y}_i \tag{4}$$

The root mean square error (RMSE) and the mean absolute error (MAE) for n samples in the dataset were computed

$$\text{RMSE} = \left[\sum_{i=1}^{n} \Delta_i^2 / n\right]^{1/2} \tag{5}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\Delta_i| \tag{6}$$

The developed models were further assessed by the correlation coefficients r^2 and q^2 in the training set and external set, respectively, by the following equation

$$r^2,\ q^2 = 1 - \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 / \sum_{i=1}^{n} (y_i - \langle \hat{y} \rangle)^2 \tag{7}$$

where $\langle \hat{y} \rangle$ represents the average predicted value.

Additionally, Ojha et al. [37] also proposed various modified versions of r^2 to gauge the model performance

$$r_m^2 = r^2\left(1 - \sqrt{\left|r^2 - r_o^2\right|}\right) \tag{8}$$

$$r'^2_m = r^2\left(1 - \sqrt{\left|r^2 - r'^2_o\right|}\right) \tag{9}$$

$$\left\langle r_m^2 \right\rangle = \left(r_m^2 + r'^2_m\right)/2 \tag{10}$$

$$\Delta r_m^2 = \left|r_m^2 - r'^2_m\right| \tag{11}$$

where the correlation coefficient r_o^2 and the slope k were derived from the regression line (predicted vs. observed values) without the intercept, whereas r'^2_o was calculated from the regression line (observed vs. predicted values) without the intercept.

The developed model was further subjected to internal validation using the leave-one-out cross-validation scheme, producing the correlation coefficient q^2_{CV}. Furthermore, *Y*-scrambling or *Y*-randomization was applied by randomly permuting the log P_{eff} values, viz. Y values, to refit the previously derived models while the descriptors remained unchanged. This procedure was repeated 25 times, as suggested in [34], to give rise the average correlation coefficient $\left\langle r_s^2 \right\rangle$ to ensure no chance correlation was associated with those derived models.

Additionally, the predictivity in the external dataset was evaluated by the correlation coefficients q^2_{F1}, q^2_{F2}, and q^2_{F3}, and concordance correlation coefficient (*CCC*) using *QSARINS* [32,113]

$$q^2_{\text{F1}} = 1 - \sum_{i=1}^{n_{\text{EXT}}} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n_{\text{EXT}}} (y_i - \langle y_{\text{TR}} \rangle)^2 \tag{12}$$

$$q^2_{\text{F2}} = 1 - \sum_{i=1}^{n_{\text{EXT}}} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n_{\text{EXT}}} (y_i - \langle y_{\text{EXT}} \rangle)^2 \tag{13}$$

$$q^2_{F3} = 1 - \left[\sum_{i=1}^{n_{EXT}} (y_i - \hat{y}_i)^2 / n_{EXT}\right] / \left[\sum_{i=1}^{n_{TR}} (y_i - \langle y_{TR}\rangle)^2 / n_{TR}\right] \quad (14)$$

$$CCC = \frac{2\sum_{i=1}^{n_{EXT}} (y_i - \langle y_{EXT}\rangle)(\hat{y}_i - \langle \hat{y}_{EXT}\rangle)}{\sum_{i=1}^{n_{EXT}} (y_i - \langle y_{EXT}\rangle)^2 + (\hat{y}_i - \langle \hat{y}_{EXT}\rangle)^2 + n_{EXT}(\langle y_{EXT}\rangle - \langle \hat{y}_{EXT}\rangle)^2} \quad (15)$$

where n_{TR} and n_{EXT} are the numbers of samples in the training set and external set, respectively; $\langle \hat{y}_{TR}\rangle$ stands for the average predicted value in the training set; and $\langle y_{EXT}\rangle$ and $\langle \hat{y}_{EXT}\rangle$ represent the average observed and predicted values in the external set, respectively.

Finally, the predictivity of developed models were further assessed by combining the most stringent criteria collectively suggested by Golbraikh et al. [36], Ojha et al. [37], Roy et al. [38], and Chirico and Gramatica [39]

$$r^2, q^2_{CV},\ q^2, q^2_{Fn} \geq 0.70 \quad (16)$$

$$\left|r^2 - q^2_{CV}\right| < 0.10 \quad (17)$$

$$\left(r^2 - r^2_o\right)/r^2 < 0.10 \text{ and } 0.85 \leq k \leq 1.15 \quad (18)$$

$$\left|r^2_o - r'^2_o\right| < 0.30 \quad (19)$$

$$r^2_m \geq 0.65 \quad (20)$$

$$\left\langle r^2_m \right\rangle \geq 0.65 \text{ and } \Delta r^2_m < 0.20 \quad (21)$$

$$CCC \geq 0.85 \quad (22)$$

where r in Equations (18)–(21) denotes r and q in the training set and external set, respectively. q_{Fn} in Equation (16) symbolizes q^2_{F1}, q^2_{F2}, and q^2_{F3}.

5. Conclusions

Intestinal permeability plays a pivotal role in systemic drug absorption that, in turn, can be of critical importance to drug efficacy. As such, intestinal permeability is one of the important drug metabolism and pharmacokinetics parameters that should be assessed in the process of drug discovery and development. An in silico model to predict the intestinal permeability can be of great value to drug discovery and development. Nevertheless, intestinal permeability is an extremely complex process that can take place through various routes, namely passive diffusion and carrier-mediated active transport. thus distinct descriptor combinations as well as various relationships are needed to depict these variations in different mechanisms. The novel machine learning-based HSVR scheme, which concurrently possesses the advantageous characteristics of a local model and a global model, namely larger coverage of applicability domain and higher degree of predictivity, respectively, was adopted in this investigation to construct a QSAR model to predict the intestinal permeability. The constructed HSVR showed great prediction accuracy for the molecules in the training set and test set, respectively, with superior predictivity and statistical significance. When subjected to a mock test by a group of molecules to mimic real challenges, the built HSVR model also performed equivalently well. In addition, the selected descriptors can render those interactions associated with passive diffusion and active transport. Accordingly, it can be asserted that this HSVR model can be utilized as an accurate and dependable predictive tool, even in the high throughput environment, to facilitate drug discovery and development by predicting the intestinal permeability of hit and lead compounds even when they are virtual. Additionally, the development of HSVR also paves the way to create more in silico models to predict oral absorption, drug stability in stomach, and bioavailability in the future.

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/21/10/3582/s1, Table S1. Selected compounds for this study; their names, SMILES strings, CAS numbers, and observed log P_e values; their predicted values by SVR A, SVR B, HSVR, and PLS; data partitions; and references; Table S2. Optimal runtime parameters for the SVR models; Figure S1. Histograms of: (A) log P_{eff}; (B) molecular weight (MW); (C) log *D*; (D) log *P*; (E) hydrogen-bond acceptor (HBA); (F) hydrogen-bond donor (HBD); and (G) polar surface area (PSA) in the training set and test set.

Author Contributions: M.-H.L., C.-F.W., and M.K.L. conceived and designed the study; M.-H.L., G.H.T., and M.K.L. performed the experiments and analyzed the data; and M.-H.L., G.H.T., C.F.W., and M.K.L. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Technology, Taiwan.

Acknowledgments: Parts of calculations were performed at the National Center for High-Performance Computing, Taiwan. The authors are grateful to Paola Gramatica for providing the free license of *QSARINS*.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

QSAR	Quantitative structure–activity relationship
HSVR	Hierarchical support vector regression
ADME/Tox	Absorption, distribution, metabolism and excretion, and toxicity
BDDCS	Biopharmaceutics drug disposition classification system
SLC	Solute carrier
ABC	ATP-binding cassette
Caco-2	Colorectal adenocarcinoma
MDCK	Madin–Darby canine kidney
PAMPA	Parallel artificial membrane permeability assay
SPIP	Single-pass intestinal perfusion

References

1. Chi, C.-T.; Lee, M.-H.; Weng, C.-F.; Leong, M.K. In Silico Prediction of PAMPA Effective Permeability Using a Two-QSAR Approach. *Int. J. Mol. Sci.* **2019**, *20*, 3170. [CrossRef]
2. Van de Waterbeemd, H.; Smith, D.A.; Beaumont, K.; Walker, D.K. Property-based design: Optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* **2001**, *44*, 1313–1333. [CrossRef]
3. Soleimani, V.; Sahebkar, A.; Hosseinzadeh, H. Turmeric (*Curcuma longa*) and its major constituent (curcumin) as nontoxic and safe substances: Review. *Phytother. Res.* **2018**, *32*, 985–995. [CrossRef]
4. Khaerunnisa, S.; Kurniawan, H.; Awaluddin, R.; Suhartati, S.; Soetjipto, S. Potential inhibitor of COVID-19 main protease (M^{pro}) from several medicinal plant compounds by molecular docking study. *Preprints* **2020**. [CrossRef]
5. Cheng, D.; Li, W.; Wang, L.; Lin, T.; Poiani, G.; Wassef, A.; Hudlikar, R.; Ondar, P.; Brunetti, L.; Kong, A.-N. Pharmacokinetics, pharmacodynamics, and pkpd modeling of curcumin in regulating antioxidant and epigenetic gene expression in healthy human volunteers. *Mol. Pharm.* **2019**, *16*, 1881–1889. [CrossRef]
6. Zwerling, A. Costs of tuberculosis screening among inpatients with HIV. *Lancet Glob. Health* **2019**, *7*, e163–e164. [CrossRef]
7. Otu, A.; Hashmi, M.; Mukhtar, A.M.; Kwizera, A.; Tiberi, S.; Macrae, B.; Zumla, A.; Dünser, M.W.; Mer, M. The critically ill patient with tuberculosis in intensive care: Clinical presentations, management and infection control. *J. Crit. Care* **2018**, *45*, 184–196. [CrossRef]
8. Schneckener, S.; Grimbs, S.; Hey, J.; Menz, S.; Osmers, M.; Schaper, S.; Hillisch, A.; Göller, A.H. Prediction of oral Bioavailability in Rats: Transferring insights from in vitro correlations to (deep) machine learning models using in silico model outputs and chemical structure parameters. *J. Chem. Inf. Model.* **2019**, *59*, 4893–4905. [CrossRef]
9. Balimane, P.V.; Chong, S.; Morrison, R.A. Current methodologies used for evaluation of intestinal permeability and absorption. *J. Pharm. Toxicol. Methods* **2000**, *44*, 301–312. [CrossRef]
10. Wu, C.-Y.; Benet, L.Z. Predicting drug disposition via application of BCS: Transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm. Res.* **2005**, *22*, 11–23. [CrossRef]

11. Dahlgren, D.; Lennernäs, H. Intestinal permeability and drug absorption: Predictive experimental, computational and in vivo approaches. *Pharmaceutics* **2019**, *11*, 411. [CrossRef]
12. Sugano, K.; Kansy, M.; Artursson, P.; Avdeef, A.; Bendels, S.; Di, L.; Ecker, G.F.; Faller, B.; Fischer, H.; Gerebtzoff, G.; et al. Coexistence of passive and carrier-mediated processes in drug transport. *Nat. Rev. Drug Discov.* **2010**, *9*, 597–614. [CrossRef]
13. Bergström, C.A.S.; Haeberlein, M.; Norinder, U. Computational Absorption Prediction. In *Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability*; van de Waterbeemd, H., Testa, B., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2009; Volume 40, pp. 409–432.
14. Meier, Y.; Eloranta, J.J.; Darimont, J.; Ismair, M.G.; Hiller, C.; Fried, M.; Kullak-Ublick, G.A.; Vavricka, S.R. Regional DISTRIBUTION of solute carrier mRNA expression along the human intestinal tract. *Drug Metab. Dispos.* **2007**, *35*, 590–594. [CrossRef]
15. Seithel, A.; Karlsson, J.; Hilgendorf, C.; Björquist, A.; Ungell, A.-L. Variability in mRNA expression of ABC- and SLC-transporters in human intestinal cells: Comparison between human segments and Caco-2 cells. *Eur. J. Pharm. Sci.* **2006**, *28*, 291–299. [CrossRef]
16. Englund, G.; Rorsman, F.; Rönnblom, A.; Karlbom, U.; Lazorova, L.; Gråsjö, J.; Kindmark, A.; Artursson, P. Regional levels of drug transporters along the human intestinal tract: Co-expression of ABC and SLC transporters and comparison with Caco-2 cells. *Eur. J. Pharm. Sci.* **2006**, *29*, 269–277. [CrossRef]
17. Zakeri-Milani, P.; Valizadeh, H.; Tajerzadeh, H.; Azarmi, Y.; Islambolchilar, Z.; Barzegar, S.; Barzegar-Jalali, M. Predicting human intestinal permeability using single-pass intestinal perfusion in rat. *J. Pharm. Pharm. Sci.* **2007**, *10*, 368–379.
18. Zhu, W.; Jeong, E.J. Intestinal Perfusion Methods for Oral Drug Absorptions. In *Oral Bioavailability: Basic Principles, Advanced Concepts, and Applications*; Hu, M., Li, X., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2011; pp. 461–473.
19. Komiya, I.; Park, J.; Kamani, A.; Ho, N.; Higuchi, W. Quantitative mechanistic studies in simultaneous fluid flow and intestinal absorption using steroids as model solutes. *Int. J. Pharm.* **1980**, *4*, 249–262. [CrossRef]
20. Peters, S.A.; Jones, C.R.; Ungell, A.-L.; Hatley, O.J.D. Predicting drug extraction in the human gut wall: Assessing contributions from drug metabolizing enzymes and transporter proteins using preclinical models. *Clin. Pharmacokinet.* **2016**, *55*, 673–696. [CrossRef]
21. Chen, X.-Q.; Ziemba, T.; Huang, C.; Chang, M.; Xu, C.; Qiao, J.X.; Wang, T.C.; Finlay, H.J.; Salvati, M.E.; Adam, L.P.; et al. Oral delivery of highly lipophilic, poorly water-soluble drugs: Self-emulsifying drug delivery systems to improve oral absorption and enable high-dose toxicology studies of a cholesteryl ester transfer protein inhibitor in preclinical species. *J. Pharm. Sci.* **2018**, *107*, 1352–1360. [CrossRef]
22. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; et al. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977–5010. [CrossRef]
23. Paranjpe, P.V.; Grass, G.M.; Sinko, P.J. In silico tools for drug absorption prediction. *Am. J. Drug Deliv.* **2003**, *1*, 133–148. [CrossRef]
24. Varma, M.V.; Ambler, C.M.; Ullah, M.; Rotter, C.J.; Hao, S.; Litchfield, J.; Fenner, K.S.; El-Kattan, A.F. Targeting intestinal transporters for optimizing oral drug absorption. *Curr. Drug Metab.* **2010**, *11*, 730–742. [CrossRef]
25. David, W.F.; Jeyaganesh, R.; Patrick, D.B.; David, M. Bioavailability through PepT1: The role of computer modelling in intelligent drug design. *Curr. Comput. Aided Drug Des.* **2010**, *6*, 68–78.
26. Chen, C.; Lee, M.-H.; Weng, C.-F.; Leong, M.K. Theoretical prediction of the complex P-glycoprotein substrate efflux based on the novel hierarchical support vector regression scheme. *Molecules* **2018**, *23*, 1820. [CrossRef]
27. Leong, M.K.; Chen, Y.-M.; Chen, T.-H. Prediction of human cytochrome p450 2b6-substrate interactions using hierarchical support vector regression approach. *J. Comput. Chem.* **2009**, *30*, 1899–1909. [CrossRef]
28. Leong, M.K.; Lin, S.-W.; Chen, H.-B.; Tsai, F.-Y. Predicting mutagenicity of aromatic amines by various machine learning approaches. *Toxicol. Sci.* **2010**, *116*, 498–513. [CrossRef]
29. Ding, Y.-L.; Lyu, Y.-C.; Leong, M.K. In silico prediction of the mutagenicity of nitroaromatic compounds using a novel two-QSAR approach. *Toxicol. Vitro* **2017**, *40*, 102–114. [CrossRef]
30. Gnanadesikan, R.; Kettenring, J.R. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **1972**, *28*, 81–124. [CrossRef]
31. Scott, D.W. Averaged shifted histogram. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 160–164. [CrossRef]

32. Gramatica, P.; Cassani, S.; Chirico, N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J. Comput. Chem.* **2014**, *35*, 1036–1044. [CrossRef]
33. Gajewicz, A. How to judge whether QSAR/read-across predictions can be trusted: A novel approach for establishing a model's applicability domain. *Environ. Sci. Nano* **2018**, *5*, 408–421. [CrossRef]
34. Rücker, C.; Rücker, G.; Meringer, M. y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.* **2007**, *47*, 2345–2357. [CrossRef] [PubMed]
35. Caudill, M. Using neural networks: Hybrid expert networks. *AI Expert* **1990**, *5*, 49–54.
36. Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.-D.; Lee, K.-H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided Mol. Des.* **2003**, *17*, 241–253. [CrossRef] [PubMed]
37. Ojha, P.K.; Mitra, I.; Das, R.N.; Roy, K. Further exploring r_m^2 metrics for validation of QSPR models. *Chemom. Intell. Lab. Syst.* **2011**, *107*, 194–205. [CrossRef]
38. Roy, K.; Mitra, I.; Kar, S.; Ojha, P.K.; Das, R.N.; Kabir, H. Comparative studies on some metrics for external validation of QSPR models. *J. Chem. Inf. Model.* **2012**, *52*, 396–408. [CrossRef] [PubMed]
39. Chirico, N.; Gramatica, P. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J. Chem. Inf. Model.* **2012**, *52*, 2044–2058. [CrossRef]
40. Lennernäs, H. Regional intestinal drug permeation: Biopharmaceutics and drug development. *Eur. J. Pharm. Sci.* **2014**, *57*, 333–341. [CrossRef]
41. Zhao, Y.H.; Le, J.; Abraham, M.H.; Hersey, A.; Eddershaw, P.J.; Luscombe, C.N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; et al. Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* **2001**, *90*, 749–784. [CrossRef]
42. Verma, R.; Hansch, C.; Selassie, C. Comparative QSAR studies on PAMPA/modified PAMPA for high throughput profiling of drug absorption potential with respect to Caco-2 cells and human intestinal absorption. *J. Comput. Aided Mol. Des.* **2007**, *21*, 3–22. [CrossRef]
43. Polley, M.J.; Burden, F.R.; Winkler, D.A. Predictive human intestinal absorption QSAR models using bayesian regularized neural networks. *Aust. J. Chem.* **2005**, *58*, 859–863. [CrossRef]
44. Zhao, Y.H.; Abraham, M.H.; Le, J.; Hersey, A.; Luscombe, C.N.; Beck, G.; Sherborne, B.; Cooper, I. Rate-limited steps of human oral absorption and QSAR studies. *Pharm. Res.* **2002**, *19*, 1446–1457. [CrossRef] [PubMed]
45. Gunturi, S.B.; Ramamurthi, N. In silico adme modeling 3: Computational models to predict human intestinal absorption using sphere exclusion and kNN QSAR methods. *QSAR Comb. Sci.* **2007**, *26*, 653–668. [CrossRef]
46. Moda, T.L.; Andricopulo, A.D. Consensus hologram QSAR modeling for the prediction of human intestinal absorption. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 2889–2893. [CrossRef]
47. Basant, N.; Gupta, S.; Singh, K.P. Predicting human intestinal absorption of diverse chemicals using ensemble learning based QSAR modeling approaches. *Comput. Biol. Chem.* **2016**, *61*, 178–196. [CrossRef] [PubMed]
48. Silva, F.T.; Trossini, G.H.G. The survey of the use of qsar methods to determine intestinal absorption and oral bioavailability during drug design. *Med. Chem.* **2014**, *10*, 441–448. [CrossRef] [PubMed]
49. Deconinck, E.; Coomans, D.; Vander Heyden, Y. Exploration of linear modelling techniques and their combination with multivariate adaptive regression splines to predict gastro-intestinal absorption of drugs. *J. Pharm. Biomed. Anal.* **2007**, *43*, 119–130. [CrossRef]
50. Reynolds, D.P.; Lanevskij, K.; Japertas, P.; Didziapetris, R.; Petrauskas, A. Ionization-specific analysis of human intestinal absorption. *J. Pharm. Sci.* **2009**, *98*, 4039–4054. [CrossRef]
51. Suenderhauf, C.; Hammann, F.; Maunz, A.; Helma, C.; Huwyler, J. Combinatorial QSAR modeling of human intestinal absorption. *Mol. Pharm.* **2011**, *8*, 213–224. [CrossRef]
52. Ghafourian, T.; Freitas, A.A.; Newby, D. The impact of training set data distributions for modelling of passive intestinal absorption. *Int. J. Pharm.* **2012**, *436*, 711–720. [CrossRef]
53. Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Dorne, J.L. SAR for gastro-intestinal absorption and blood-brain barrier permeation of pesticides. *Chem. Biol. Interact.* **2018**, *290*, 1–5. [CrossRef] [PubMed]
54. Wang, N.-N.; Huang, C.; Dong, J.; Yao, Z.-J.; Zhu, M.-F.; Deng, Z.-K.; Lv, B.; Lu, A.-P.; Chen, A.F.; Cao, D.-S. Predicting human intestinal absorption with modified random forest approach: A comprehensive evaluation of molecular representation, unbalanced data, and applicability domain issues. *RSC Adv.* **2017**, *7*, 19007–19018. [CrossRef]

55. Kumar, R.; Sharma, A.; Siddiqui, M.H.; Tiwari, R.K. Prediction of human intestinal absorption of compounds using artificial intelligence techniques. *Curr. Drug Discov. Technol.* **2017**, *14*, 244–254. [CrossRef] [PubMed]
56. Shultz, M.D. Two decades under the influence of the rule of five and the changing properties of approved oral drugs. *J. Med. Chem.* **2019**, *62*, 1701–1714. [CrossRef]
57. Broccatelli, F.; Salphati, L.; Plise, E.; Cheong, J.; Gobbi, A.; Lee, M.L.; Aliagas, I. Predicting passive permeability of drug-like molecules from chemical structure: Where are we? *Mol. Pharm.* **2016**, *13*, 4199–4208. [CrossRef]
58. Kishimoto, H.; Miyazaki, K.; Shirasaka, Y.; Inoue, K. Effect of mucus layer on the transcellular absorption of lipophilic drugs in rat small intestine. *FASEB J.* **2018**, *32*, 761.1. [CrossRef]
59. Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714–3717. [CrossRef]
60. Meanwell, N.A. Improving drug candidates by design: A focus on physicochemical properties as a means of improving compound disposition and safety. *Chem. Res. Toxicol.* **2011**, *24*, 1420–1456. [CrossRef]
61. Desai, P.V.; Raub, T.J.; Blanco, M.-J. How hydrogen bonds impact P-glycoprotein transport and permeability. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 6540–6548. [CrossRef]
62. Ferreira, R.J.; dos Santos, D.J.V.A.; Ferreira, M.-J.U.; Guedes, R.C. Toward a better pharmacophore description of p-glycoprotein modulators, based on macrocyclic diterpenes from euphorbia species. *J. Chem. Inf. Model.* **2011**, *51*, 1315–1324. [CrossRef]
63. Leong, M.K.; Chen, H.-B.; Shih, Y.-H. Prediction of promiscuous P-glycoprotein inhibition using a novel machine learning scheme. *PLoS ONE* **2012**, *7*, e33829. [CrossRef] [PubMed]
64. Didziapetris, R.; Japertas, P.; Avdeef, A.; Petrauskas, A. Classification analysis of P-glycoprotein substrate specificity. *J. Drug Target.* **2003**, *11*, 391–406. [CrossRef] [PubMed]
65. Tmej, C.; Chiba, P.; Huber, M.; Richter, E.; Hitzler, M.; Schaper, K.-J.; Ecker, G. A Combined hansch/free-wilson approach as predictive tool in qsar studies on propafenone-type modulators of multidrug resistance. *Arch. Pharm.* **1998**, *331*, 233–240. [CrossRef]
66. Hiessböck, R.; Wolf, C.; Richter, E.; Hitzler, M.; Chiba, P.; Kratzel, M.; Ecker, G. Synthesis and in vitro multidrug resistance modulating activity of a series of dihydrobenzopyrans and tetrahydroquinolines. *J. Med. Chem.* **1999**, *42*, 1921–1926. [CrossRef]
67. Agatonovic-Kustrin, S.; Beresford, R.; Yusof, A.P.M. Theoretically-derived molecular descriptors important in human intestinal absorption. *J. Pharm. Biomed. Anal.* **2001**, *25*, 227–237. [CrossRef]
68. Hou, T.; Wang, J.; Zhang, W.; Wang, W.; Xu, X. Recent advances in computational prediction of drug absorption and permeability in drug discovery. *Curr. Med. Chem.* **2006**, *13*, 2653–2667. [CrossRef]
69. Bain, L.J.; McLachlan, J.B.; LeBlanc, G.A. Structure-activity relationships for xenobiotic transport substrates and inhibitory ligands of P-glycoprotein. *Environ. Health Perspect.* **1997**, *105*, 812–818. [CrossRef]
70. Mollazadeh, S.; Moosavi, F.; Hadizadeh, F.; Seifi, M.; Behravan, J.; Iman, M. Synthesis and DFT study on hantzsch reaction to produce asymmetrical compounds of 1,4-dihydropyridine derivatives for P-glycoprotein inhibition as anticancer agent. *Recent Pat. Anticancer Drug Discov.* **2018**, *13*, 255–264. [CrossRef]
71. Newstead, S.; Drew, D.; Cameron, A.D.; Postis, V.L.G.; Xia, X.; Fowler, P.W.; Ingram, J.C.; Carpenter, E.P.; Sansom, M.S.P.; McPherson, M.J.; et al. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide–proton symporters, PepT1 and PepT2. *EMBO J.* **2011**, *30*, 417–426. [CrossRef]
72. Golin, J.; Ambudkar, S.V.; Gottesman, M.M.; Habib, A.D.; Sczepanski, J.; Ziccardi, W.; May, L. Studies with novel Pdr5p substrates demonstrate a strong size dependence for xenobiotic efflux. *J. Biol. Chem.* **2003**, *278*, 5963–5969. [CrossRef]
73. Varma, M.V.S.; Sateesh, K.; Panchagnula, R. Functional role of P-glycoprotein in limiting intestinal absorption of drugs: Contribution of passive permeability to P-glycoprotein mediated efflux transport. *Mol. Pharm.* **2005**, *2*, 12–21. [CrossRef] [PubMed]
74. Pauletti, G.M.; Okumu, F.W.; Borchardt, R.T. Effect of size and charge on the passive diffusion of peptides across Caco-2 cell monolayers via the paracellular pathway. *Pharm. Res.* **1997**, *14*, 164–168. [CrossRef]
75. Gerebtzoff, G.; Seelig, A. In silico prediction of blood-brain barrier permeation using the calculated molecular cross-sectional area as main parameter. *J. Chem. Inf. Model.* **2006**, *46*, 2638–2650. [CrossRef] [PubMed]
76. Leeson, P.D.; St-Gallay, S.A.; Wenlock, M.C. Impact of ion class and time on oral drug molecular properties. *Med. Chem. Comm.* **2011**, *2*, 91–105. [CrossRef]

77. Chu, X.-Y.; Sánchez-Castaño, G.P.; Higaki, K.; Oh, D.-M.; Hsu, C.-P.; Amidon, G.L. Correlation between epithelial cell permeability of cephalexin and expression of intestinal oligopeptide transporter. *J. Pharmacol. Exp. Ther.* **2001**, *299*, 575–582.
78. Dahan, A.; Amidon, G. Grapefruit juice and its constituents augment colchicine intestinal absorption: Potential hazardous interaction and the role of P-glycoprotein. *Pharm. Res.* **2009**, *26*, 883–892. [CrossRef]
79. Deng, D.-D.; Liu, C.-Y.; Liao, Z.-X.; Liang, J.-Y.; Lu, J.-Z. Prediction of passive transport of 10 natural products by liposome-based fluorescence technique. *Chin. J. Anal. Chem.* **2007**, *35*, 1696–1700.
80. Fagerholm, U.; Johansson, M.; Lennernäs, H. Comparison between permeability coefficients in rat and human jejunum. *Pharm. Res.* **1996**, *13*, 1336–1342. [CrossRef]
81. Fang, L.; Wang, Q.; Bi, K.; Zhao, X. Simultaneous determination of procaspase activating compound 1 and permeability markers in intestinal perfusion samples and application to a rat intestinal absorption study. *Chromatographia* **2016**, *79*, 1659–1663. [CrossRef]
82. Incecayir, T.; Tsume, Y.; Amidon, G.L. Comparison of the permeability of metoprolol and labetalol in rat, mouse, and Caco-2 cells: Use as a reference standard for BCS classification. *Mol. Pharm.* **2013**, *10*, 958–966. [CrossRef]
83. Jain, R.; Agarwal, S.; Mandava, N.K.; Sheng, Y.; Mitra, A.K. Interaction of dipeptide prodrugs of saquinavir with multidrug resistance protein-2 (MRP-2): Evasion of MRP-2 mediated efflux. *Int. J. Pharm.* **2008**, *362*, 44–51. [CrossRef] [PubMed]
84. Kang, M.J.; Kim, H.S.; Jeon, H.S.; Park, J.H.; Lee, B.S.; Ahn, B.K.; Moon, K.Y.; Choi, Y.W. In situ intestinal permeability and in vivo absorption characteristics of olmesartan medoxomil in self-microemulsifying drug delivery system. *Drug Dev. Ind. Pharm.* **2012**, *38*, 587–596. [CrossRef] [PubMed]
85. Kim, J.-S.; Mitchell, S.; Kijek, P.; Tsume, Y.; Hilfinger, J.; Amidon, G.L. The suitability of an in situ perfusion model for permeability determinations: Utility for BCS class I biowaiver requests. *Mol. Pharm.* **2006**, *3*, 686–694. [CrossRef] [PubMed]
86. Krondahl, E.; Orzechowski, A.; Ekström, G.; Lennernäs, H. Rat jejunal permeability and metabolism of μ-selective tetrapeptides in gastrointestinal fluids from humans and rats. *Pharm. Res.* **1997**, *14*, 1780–1785. [CrossRef] [PubMed]
87. Lindahl, A.; Sandström, R.; Ungell, A.L.; Lennernäs, H. Concentration-and region-dependent intestinal permeability of fluvastatin in the Rat. *J. Pharm. Pharmacol.* **1998**, *50*, 737–744. [CrossRef]
88. Lindahl, A.; Persson, B.; Ungell, A.-L.; Lennernäs, H. Surface activity and concentration dependent intestinal permeability in the rat. *Pharm. Res.* **1999**, *16*, 97–102. [CrossRef]
89. Liu, Y.; Hu, M. Absorption and metabolism of flavonoids in the caco-2 cell culture model and a perused rat intestinal model. *Drug Metab. Dispos.* **2002**, *30*, 370–377. [CrossRef]
90. Liu, C.; Liu, D.; Bai, F.; Zhang, J.; Zhang, N. In vitro and in vivo studies of lipid-based nanocarriers for oral N3-o-toluyl-fluorouracil delivery. *Drug Deliv.* **2010**, *17*, 352–363. [CrossRef]
91. Lozoya-Agullo, I.; Zur, M.; Wolk, O.; Beig, A.; González-Álvarez, I.; González-Álvarez, M.; Merino-Sanjuán, M.; Bermejo, M.; Dahan, A. In-situ intestinal rat perfusions for human Fabs prediction and BCS permeability class determination: Investigation of the single-pass vs. the doluisio experimental approaches. *Int. J. Pharm.* **2015**, *480*, 1–7. [CrossRef]
92. Masaoka, Y.; Tanaka, Y.; Kataoka, M.; Sakuma, S.; Yamashita, S. Site of drug absorption after oral administration: Assessment of membrane permeability and luminal concentration of drugs in each segment of gastrointestinal tract. *Eur. J. Pharm. Sci.* **2006**, *29*, 240–250. [CrossRef]
93. Nagare, N.; Damre, A.; Singh, K.; Mallurwar, S.; Iyer, S.; Naik, A.; Chintamaneni, M. Determination of site of absorption of propranolol in rat gut using in situ single-pass intestinal perfusion. *Indian J. Pharm. Sci.* **2010**, *72*, 625–629.
94. Ozawa, M.; Tsume, Y.; Zur, M.; Dahan, A.; Amidon, G.L. Intestinal permeability study of minoxidil: Assessment of minoxidil as a high permeability reference drug for biopharmaceutics classification. *Mol. Pharm.* **2014**, *12*, 204–211. [CrossRef]
95. Patel, J.; Barve, K.H. Intestinal permeability of lamivudine using single pass intestinal perfusion. *Indian J. Pharm. Sci.* **2012**, *74*, 478–481.
96. Shashikanth, P.; Mohan, P.C.; Karunakar, K.; Sagi, S.R. Paclitaxel disposition studies using P-Gp inhibtor & inducer by single pass intestinal perfusion in rats. *Asian J. Pharm. Clin. Res.* **2013**, *1*, 199–203.

97. Steffansen, B.; Lepist, E.-I.; Taub, M.E.; Larsen, B.D.; Frokjaer, S.; Lennernäs, H. Stability, metabolism and transport of d-Asp (OBzl)–Ala—a model prodrug with affinity for the oligopeptide transporter. *Eur. J. Pharm. Sci.* **1999**, *8*, 67–73. [CrossRef]
98. Stewart, B.H.; Chan, O.H.; Lu, R.H.; Reyner, E.L.; Schmid, H.L.; Hamilton, H.W.; Steinbaugh, B.A.; Taylor, M.D. Comparison of intestinal permeabilities determined in multiple in vitro and in situ models: Relationship to absorption in humans. *Pharm. Res.* **1995**, *12*, 693–699. [CrossRef]
99. Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129. [CrossRef]
100. Cammi, R.; Tomasi, J. Remarks on the use of the apparent surface charges (ASC) methods in solvation problems: Iterative versus matrix-inversion procedures and the renormalization of the apparent charges. *J. Comput. Chem.* **1995**, *16*, 1449–1458. [CrossRef]
101. Besler, B.H.; Merz, K.M.J.; Kollman, P.A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431–439. [CrossRef]
102. Muehlbacher, M.; Spitzer, G.; Liedl, K.; Kornhuber, J. Qualitative prediction of blood–brain barrier permeability on a large and refined dataset. *J. Comput. Aided Mol. Des.* **2011**, *25*, 1095–1106. [CrossRef]
103. Topliss, J.G.; Edwards, R.P. Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem.* **1979**, *22*, 1238–1244. [CrossRef] [PubMed]
104. Tseng, Y.J.; Hopfinger, A.J.; Esposito, E.X. The great descriptor melting pot: Mixing descriptors for the common good of QSAR models. *J. Comput. Aided Mol. Des.* **2012**, *26*, 39–43. [CrossRef] [PubMed]
105. Rogers, D.; Hopfinger, A.J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866. [CrossRef]
106. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
107. Hubert, M.; Engelen, S. Robust PCA and classification in biosciences. *Bioinformatics* **2004**, *20*, 1728–1736. [CrossRef]
108. Tropsha, A.; Gramatica, P.; Gombar, V.K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77. [CrossRef]
109. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
110. Vapnik, V.; Golowich, S.; Smola, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. In *Advances in Neural Information Processing Systems 9*; Mozer, M.C., Jordan, M.I., Petsche, T., Eds.; MIT Press: Cambridge, MA, USA, 1997; pp. 281–287.
111. Kecman, V. *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*; MIT Press: Cambridge, MA, USA, 2001; 576p.
112. Dearden, J.C.; Cronin, M.T.D.; Kaiser, K.L.E. How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20*, 241–266. [CrossRef]
113. Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34*, 2121–2132. [CrossRef]