

Article

Prediction Models for Brain Distribution of Drugs Based on Biomimetic Chromatographic Data

Theodosia Vallianatou ^{1,*} , Fotios Tsopelas ²  and Anna Tsantili-Kakoulidou ^{3,*} 

¹ Medical Mass Spectrometry Imaging, Department of Pharmaceutical Biosciences, Uppsala University, 751 24 Uppsala, Sweden

² Laboratory of Inorganic and Analytical Chemistry, School of Chemical Engineering, National Technical University of Athens, 157 80 Athens, Greece; ftsop@central.ntua.gr

³ Faculty of Pharmacy, National and Kapodistrian University of Athens, 157 71 Athens, Greece

* Correspondence: theodosia.vallianatou@farmbio.uu.se (T.V.); tsantili@pharm.uoa.gr (A.T.-K.)

Abstract: The development of high-throughput approaches for the valid estimation of brain disposition is of great importance in the early drug screening of drug candidates. However, the complexity of brain tissue, which is protected by a unique vasculature formation called the blood–brain barrier (BBB), complicates the development of robust *in silico* models. In addition, most computational approaches focus only on brain permeability data without considering the crucial factors of plasma and tissue binding. In the present study, we combined experimental data obtained by HPLC using three biomimetic columns, i.e., immobilized artificial membranes, human serum albumin, and α_1 -acid glycoprotein, with molecular descriptors to model brain disposition of drugs. $K_{p,uu,brain}$, as the ratio between the unbound drug concentration in the brain interstitial fluid to the corresponding plasma concentration, brain permeability, the unbound fraction in the brain, and the brain unbound volume of distribution, was collected from literature. Given the complexity of the investigated biological processes, the extracted models displayed high statistical quality ($R^2 > 0.6$), while in the case of the brain fraction unbound, the models showed excellent performance ($R^2 > 0.9$). All models were thoroughly validated, and their applicability domain was estimated. Our approach highlighted the importance of phospholipid, as well as tissue and protein, binding in balance with BBB permeability in brain disposition and suggests biomimetic chromatography as a rapid and simple technique to construct models with experimental evidence for the early evaluation of CNS drug candidates.

Keywords: biomimetic chromatography; brain disposition; prediction models; brain BBB fraction unbound; brain unbound volume of distribution



Citation: Vallianatou, T.; Tsopelas, F.; Tsantili-Kakoulidou, A. Prediction Models for Brain Distribution of Drugs Based on Biomimetic Chromatographic Data. *Molecules* **2022**, *27*, 3668. <https://doi.org/10.3390/molecules27123668>

Academic Editors: Francesca Grisoni and Angelica Mazzolari

Received: 29 April 2022

Accepted: 3 June 2022

Published: 7 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The brain disposition of drug candidates has been one of the major issues in the pharmaceutical industry in recent decades. The development of novel compounds targeting the central nervous system (CNS) is becoming more and more essential as a consequence of the increasing incidence of neurological and neurodegenerative disorders (e.g., depression, Parkinson disease, and Alzheimer disease). The high attrition rate and failures in the field of CNS agents are current challenges that pharmaceutical companies have to face [1,2]. From this perspective, assessment of brain distribution in the early stages of the drug development process is crucial for novel CNS drug candidates. The most important obstacle for drug delivery in the brain is the blood–brain barrier (BBB), which is formed by the endothelial cells lining the brain micro-vessels [3–5]. The presence of tight junctions between the endothelial cells leads to limited fenestration, making passive diffusion the dominant transport pathway inside the brain, while the paracellular path is of much lower importance. Moreover, the brain transport of certain drugs may be facilitated or restricted by the presence of membrane transport proteins in the BBB [6].

The most common approach to quantifying the BBB permeability of molecules has been the determination of the ratio between the total brain and the total plasma concentration, mostly expressed as $K_{p,brain}$. However, the reliability of $K_{p,brain}$ in quantitative estimations of brain disposition has been criticized, as its determination is based on the total concentrations, ignoring issues like plasma protein binding (PPB) and tissue binding. According to the “free drug hypothesis”, only the unbound drug crosses the biological membranes, thus constituting the unbound concentration (C_u) relevant for pharmacological activity. From this perspective, the ratio between the unbound drug concentration in the interstitial fluid of the brain to the corresponding plasma concentration, $K_{p,uu,brain}$, is considered as a more representative measure [7–13].

Experimental determination of $K_{p,uu,brain}$ through microdialysis exhibits certain limitations as it is technically demanding, time-consuming, and inefficient for very lipophilic compounds [7]. However, $K_{p,uu,brain}$ can be derived from $K_{p,brain}$ by combining the unbound fraction of the drug in the plasma ($f_{u,p}$) and in the brain ($f_{u,brain}$) or the unbound brain volume of distribution, $V_{u,brain}$ [7]. The brain homogenate method is used to estimate the $f_{u,brain}$, considered to reflect only nonspecific binding, while the brain slice method is applied for measuring the $V_{u,brain}$ (mL/g brain), which quantifies the overall cellular uptake of the drug, including active cellular membrane transport and pH partitioning [7,10,12,14]. The relevant equations are included in Section 3. Evidently, $f_{u,p}$ is obtained by standard methods for plasma protein binding.

$K_{p,brain}$, usually converted into the logarithmic form and labeled as logBB, has been widely used to establish *in silico* models in an effort to reveal the favorable physicochemical and molecular properties for BBB permeability [15]. A number of QSAR models have been reported in the literature, most of them involving calculated lipophilicity, expressed as octanol–water partition or distribution coefficients (logP or logD), polar surface area, and molecular weight [16–19]. However, to the best of our knowledge, analogous models for $f_{u,brain}$, $V_{u,brain}$, or $K_{p,uu,brain}$ itself are rather limited [11,14,20].

A potential concern regarding *in silico* predictive models may arise in respect to the reliability of calculated physicochemical properties of drugs, in particular for new chemotypes. It has been reported that small errors in logP or logD predictions may cause substantial errors in the estimation of biological end-points, particularly if limited datasets are analyzed, while experimental lipophilicity requires a rather tedious and time-consuming procedure [21,22]. On the other hand, user-friendly chromatographic techniques may offer a challenging alternative to combining rapid measurements with theoretical descriptors for the construction of ‘hybrid’ *in silico* models based on experimental evidence.

Biomimetic properties, defined as the retention outcome on HPLC stationary phases containing a biologically relevant agent, have attracted considerable interest and are currently used for the rapid evaluation of ADME properties in early drug discovery phases. In particular, immobilized artificial membrane (IAM) chromatography, which uses phospholipid-containing stationary phases, has been applied in investigating permeability as an alternative to traditional octanol–water lipophilicity. Since, however, electrostatic interactions have a strong contribution in retention mechanism, especially in the case of protonated bases, IAM chromatography is considered as also reflecting drug–membrane interactions and tissue binding [23,24]. IAM models for oral absorption, skin partitioning, and brain penetration, mostly expressed as logBB, have been reported in the literature, usually in combination with additional molecular descriptors [23–42]. Classification of CNS⁺ and CNS[−] drugs has been suggested using the IAM retention factor divided by the fourth root of molecular weight [43]. On the other hand, protein-based stationary phases, incorporating human serum albumin (HSA) or α_1 acid glycoprotein (AGP)-, simulate binding to plasma proteins [25,26,39,44–46].

Immobilized HSA retains the characteristics of the protein in solution, as has been proven by zone and frontal analysis experiments, permitting the safe estimation of binding constants to human serum albumin in the plasma [47]. Less investigated is AGP retention due to the polymorphism of α_1 acid glycoprotein and practical issues related to the

immobilization on the silica skeleton of the column. It is well-established, however, that protonated bases are more strongly retained on AGP columns, in agreement with the same binding preference of AGP in solution [47]. In light of the above considerations, HPLC-based biomimetic properties reflect the major factors which govern most biological processes, e.g., passive diffusion and non-specific binding to phospholipids, tissue, or plasma proteins. Accordingly, their combination may be suitable for the estimation of composite pharmacokinetic data. Valko et al. used the weighted sum of IAM and HSA retention to estimate the volume of distribution and tissue binding in different organs, including brain tissue binding [39].

In the present study, we investigated the performance of biomimetic properties in combination with molecular descriptors in the development of 'hybrid' models with experimental evidence for the thorough analysis of composite $K_{p,uu,brain}$ data, and the distinct brain disposition components, $K_{p,brain}$, $f_{u,brain}$, and $V_{u,brain}$. The aim was to bridge the gap with pure in silico prediction models and to contribute to the perception of these crucial experimental end-points for CNS-acting drug candidates. For this purpose, in-house retention factors, determined for a number of pharmaceutical compounds on IAM, has, and AGP stationary phases, were used together with physicochemical/molecular descriptors. Lipophilicity, expressed as octanol–water partition coefficients (logP) or distribution coefficients at pH 7.4 (logD_{7.4}) and pH 5.0 (logD_{5.0}), was incorporated in the pool of descriptors, and its implementation in the models was compared to that of IAM retention. Models were constructed by applying multiple linear regression (MLR) and partial least squares (PLS) analysis. MLR can provide simple models that are easily used by medicinal chemists, while PLS, contributing complementary and supporting MLR models, may permit a deeper insight into the factors underlying brain disposition of drugs. Attention was given to model validation in respect to robustness and applicability domain to offset the drawback that rather limited datasets were analyzed.

2. Results and Discussion

In the present study, we collected experimental biomimetic chromatographic data, previously obtained in our laboratory, for a set of 55 pharmaceutical compounds belonging to a wide range of pharmacological classes. (For further details, see the Section 3 and Table S1.) The investigated compounds are small molecules within a molecular weight (MW) range of ca 129 to 513 Da (Figure 1a). The dataset includes neutral, basic, and acidic compounds; however, most molecules contain one basic ionizable group (Figure 1b).

Experimentally measured logBB, $K_{p,uu,brain}$, $f_{u,brain}$, and $V_{u,brain}$ values were collected from the literature [10–12,48–54] (Table S2). The collected logBB values indicated a sufficient range of BBB permeability, from highly permeable (logBB > > 1) to significantly less permeable (logBB < < 0) compounds (Figure 1c). Nevertheless, the range of log $K_{p,uu,brain}$ values suggested a lower extent of brain distribution of the studied molecules (Figure 1e). The two measures showed a mediocre intercorrelation (Figure 1e), while $f_{u,brain}$ and $V_{u,brain}$ showed significant inverse intercorrelation on the basis of their logarithmic values (Figure 1f–h).

For further statistical analysis, $f_{u,brain}$ values were converted to the thermodynamic constant K_b [55] according to the relevant equation included in Section 3. Its logarithmic form log K_b was used throughout.

2.1. Data Overview with Unsupervised Data Analysis

To obtain an overview of the dataset, unsupervised principal component analysis (PCA) was performed using the pool of descriptors and the response variables as an X matrix. A 5-component PCA model was generated with $R^2 = 0.674$ and $Q^2 = 0.462$. The score plot of the first two components shows a uniform distribution of the data in all four quartiles, with one drug, candesartan, lying outside the Hotelling T^2 ellipse (Figure 2a). As depicted by the coloring of the scores based on the log k_{wIAM} values, IAM retention is a dominant parameter of the data distribution. The loadings plot reflects the correlations

among the chromatographic data, the molecular and physicochemical descriptors, and the experimental brain distribution data in a multilayered fashion (Figure 2b). The coloring of the variables corresponds to their grouping according to hierarchical clustering based on the five principal components. $\log BB$ and $\log V_{u,brain}$ are located in the same quartile with chromatographic factors and lipophilicity parameters and belong to the same hierarchical clustering group. Interestingly, $\log K_{p,uu,brain}$ is also incorporated into the same group, but its loading value in the first component is approximately zero. $\log K_b$ is located in the opposite quartile in high proximity to hydrogen bonding parameters.

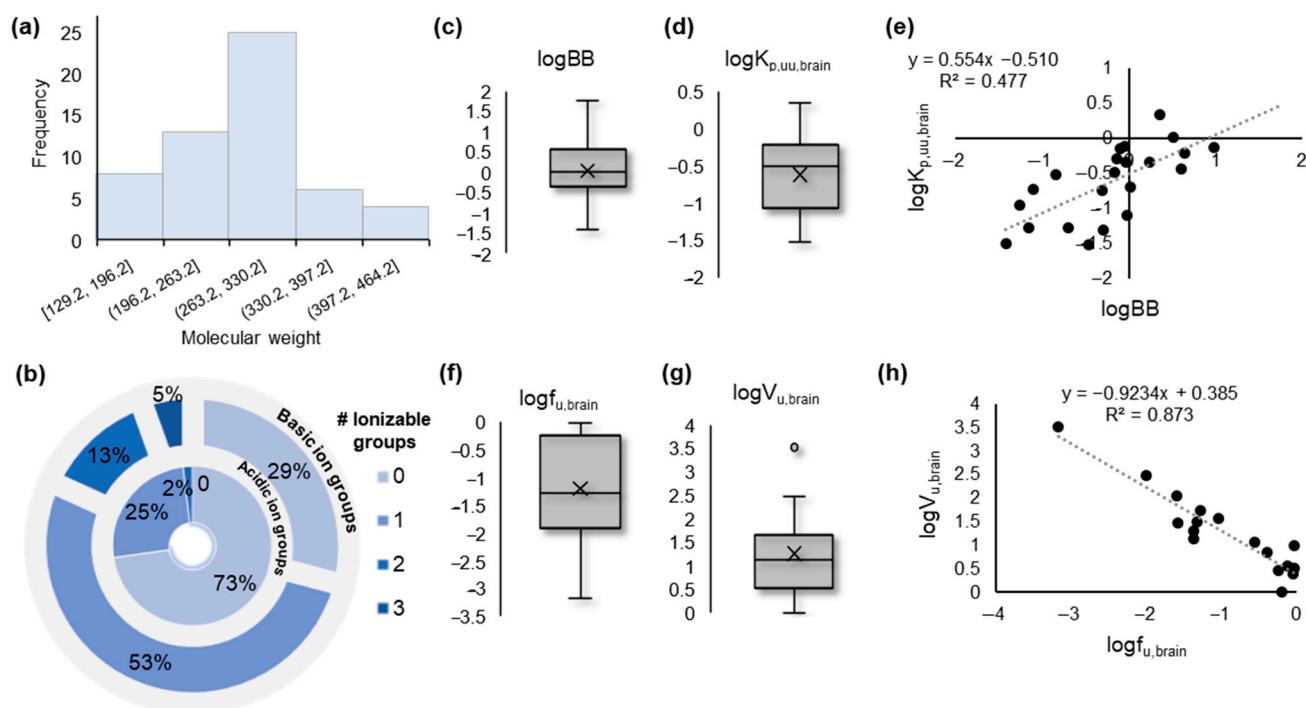


Figure 1. Overview of the data. (a) Histogram showing the distribution of the molecular weight of the investigated compounds; (b) Overlaid pie plots showing the number of ionizable groups (basic, outer layer; acidic, inner layer) present in the investigated compounds; (c) Box plot showing the distribution of the experimental $\log BB$ values collected from the literature; (d) Box plot showing the distribution of the experimental $\log K_{p,uu,brain}$ values collected from the literature; (e) Correlation scatter plot between $\log BB$ and $\log K_{p,uu,brain}$ values; (f) Box plot showing the distribution of the experimental $\log f_{u,brain}$ values collected from the literature; (g) Box plot showing the distribution of the experimental $\log V_{u,brain}$ values collected from the literature; (h) Correlation scatter plot between $\log f_{u,brain}$ and $\log V_{u,brain}$ values.

2.2. Modeling $\log K_{p,uu,brain}$

$\log K_{p,uu,brain}$ is considered to be the most relevant measure of the rate and extent of drug delivery in the brain, and as such, it was examined first. MLR analysis gave poor statistics ($R < 0.5$). Polarity descriptors, e.g., tPSA, display the strongest negative correlation (Figure 3a). Poor correlation was observed between $\log K_{p,uu,brain}$ and $\log k_{wIAM}$, $\log P$ or $\log D_{7.4}$. However, a relatively better correlation was found between $\log D_{5.0}$ and IAM retention at pH 5.0 (Figure 3a).

The above findings were further supported by Partial Least Squares analysis. PLS analysis offers a number of advantages, tolerating intercorrelated variables and missing values. Variables are treated simultaneously to extract the principal components as their linear combinations. Using the entire pool of descriptors and performing variable selection according to the variable importance to projection (VIP) criterion, no satisfactory PLS models could be obtained, with R^2 and Q^2 not exceeding 0.52 and 0.32, respectively. In agreement with MLR, polarity terms (TPSA) show the strongest impact with VIP values > 1 and negative contribution. Biomimetic properties and lipophilicity have positive contributions, albeit with considerably lower impacts. In agreement with MLR results, $\log D_{5.0}$ and IAM retention at pH 5.0 show higher influence. In fact, VIP values decrease in the order of: $\log D_{5.0} > \log k_{wIAM,5.0} > \log D_{7.4} > \log k_{10,HSA} > \log k_{wAGP} > \log P > \log k_{wIAM,7.4}$ (Table S3). The better performance of IAM retention and lipophilicity at pH 5.0 may be explained by the recently reported evidence of the interference of the endo-lysosomal system (lysosomes' pH 5.0–4.5) in vitro blood–brain barrier models [56].

Inspection of the plots of observed vs. predicted $\log K_{p,uu,brain}$ and of DModY plots (distance of observations to the Y model) revealed two strong outliers, quinidine and theophylline, present in any PLS model (figure not shown). Upon exclusion of these drugs from statistical analysis, acceptable PLS models were obtained. Using biomimetic properties and lipophilicity in combination with computational descriptors, a 1-component PLS model with 54 original descriptors was generated with the following statistics (PLS Model 1):

$$A = 1, n = 23, R^2 = 0.694 \quad Q^2 = 0.518, RMSEE = 0.262 \text{ and } RMSEcv = 0.315. \text{ (PLS Model 1)}$$

The separate effect of IAM retention or lipophilicity is shown in PLS Models 2 and 3, which include either IAM retention at pH 7.4 and 5.0 or lipophilicity measures ($\log P$, $\log D_{7.4}$ and $\log D_{5.0}$). PLS Models 2 and 3 denote equal performance of IAM retention and traditional lipophilicity:

$$A = 1, n = 23, R^2 = 0.693 \quad Q^2 = 0.529, RMSEE = 0.263 \text{ and } RMSEcv = 0.311 \text{ (PLS Model 2)}$$

$$A = 1, n = 23, R^2 = 0.694 \quad Q^2 = 0.529, RMSEE = 0.263 \text{ and } RMSEcv = 0.311 \text{ (PLS Model 3)}$$

Finally, the use of only a few computational descriptors with high VIP values, including polarity, flexibility, and shape indices, proved sufficient to produce an improved, 1-component PLS model (PLS Model 4):

$$A = 1, n = 23, R^2 = 0.686 \quad Q^2 = 0.621, RMSEE = 0.266 \text{ and } RMSEcv = 0.279 \text{ (PLS Model 4)}$$

In Figure 3b,c the plot of observed vs. predicted $\log K_{p,uu,brain}$ values generated by PLS Model 4 and the coefficients of the original variables are depicted, respectively. The predicted $\log K_{p,uu,brain}$ values are presented in Table S4. The applicability domain was defined using DModY plot, and the doubled average DModY value, equal to $0.822 \times 2 = 1.644$, was used as a critical value. Two drugs, acetaminophen and midazolam, exceed the critical value (Figure S1).

PLS Model 4 was validated by permutation tests (Figure S2) and by dividing the dataset into a training set and a test set, using 4 different training/test sets (see Section 3). The external validation demonstrated robust statistics except for 1 iteration, in which the drug acyclovir was included in the test set (Table 1). In Figure S3, a representative plot of observed vs. predicted values is presented.

Table 1. Ranges of statistical values for training and external test sets used for validation of model 4 for $\log K_{p,u,u,brain}$.

Validated PLS Model: 4 Different Test Sets with n = 6	R^2_{train}/Q^2_{train}	RMSEE	R^2_{test}	RMSEP
PLS Model 4 (based on computational descriptors)	0.687–0.831/0.597–0.736	0.212–0.276	0.212 *–0.824	0.229–0.407 *

R^2_{train} , coefficient of determination of the training sets; R^2_{test} , coefficient of determination of the test sets; Q^2_{train} , cross-validated coefficient of determination in PLS models; RMSEE, root mean square error of estimation; RMSEP, root mean square error of prediction; * statistics deteriorated due to the presence of acyclovir in the test set.

In fact, the construction of PLS Model 4 indicates that biomimetic properties and lipophilicity are less crucial in $\log K_{p,u,u,brain}$ modeling, as shown also by the failure to generate successful MLR models. Previous studies have likewise identified the lack of strong correlations between $\log K_{p,u,u,brain}$ and physicochemical descriptors, especially lipophilicity, while hydrogen bonding has been found to be the strongest contributor [11,20]. Loryan, et al. reported a PLS model with two descriptors related to polarity (tPSA) and hydrogen bonding capacity [14].

Considering the complex nature of $K_{p,u,u,brain}$ as the outcome of opposing brain disposition end-points and plasma binding [11] (see relevant equations in Section 3), modeling $K_{p,brain}$, expressed as $\log BB$, $f_{u,brain}$, and $V_{u,brain}$, can indirectly assist in the exploration of $\log K_{p,u,u,brain}$. In the next sections, the generation and validation of the MLR and PLS models of these distinct brain disposition end-points are discussed.

2.3. Modeling $\log BB$

2.3.1. Multiple Linear Regression Models

Partial correlations between $\log BB$, IAM retention, $\log D_{7.4}$, and physicochemical/molecular properties revealed the crucial effect of polarity, with TPSA showing the highest correlation coefficient ($r = -0.777$, $p < 0.001$), followed by the sum of nitrogen and oxygen atoms, NO ($r = -0.711$, $p < 0.001$), and dipole moment ($r = -0.693$, $p < 0.001$), while for $\log kwIAM$ and $\log D_{7.4}$ $r = 0.617$ ($p < 0.001$) and $r = 0.687$ ($p < 0.001$), respectively. Their combination with TPSA led to Equations (1) and (2), respectively. The term h^* in Equations (1) and (2) expresses the critical leverage value, which defines the applicability domain (AD) [57]. (See Section 3.)

$$\log BB = 0.16(\pm 0.07) \log kwIAM - 0.02(\pm 0.003) TPSA + 0.69(\pm 0.27)$$

$$n = 41^*, R = 0.806, R^2 = 0.650, R^2_{adj} = 0.632, s = 0.472, F = 35.31, h^* = 0.073 \quad (1)$$

(* Piracetam did not have a $\log kwIAM$ measured value).

Eight drugs had h values higher than the threshold h^* value and were beyond the AD of the model, as also shown in a Williams plot (Figure S4).

$$\log BB = 0.15(\pm 0.05) \log D_{7.4} - 0.02(\pm 0.003) TPSA + 0.72(\pm 0.23)$$

$$n = 42, R = 0.816, R^2 = 0.665, R^2_{adj} = 0.648, s = 0.456, F = 38.74, h^* = 0.071 \quad (2)$$

Six drugs were outside the AD of Equation (2) (Figure S4).

In both equations, TPSA had a higher significance than $\log kwIAM$ or $\log D_{7.4}$, with Student's test-t values of $|t_{TPSA}| = 5.4$ and 4.8 , respectively, while $|t_{\log kwIAM}| = 2.22$, and $|t_{\log D_{7.4}}| = 2.69$.

External validation of Equations (1) and (2) was performed by dividing the dataset into a training set and a test set. This procedure was repeated up to six times for each equation. Practically speaking, the same equations were generated with comparable statistical data (Table 2). In all cases, observed vs. predicted values showed a 1:1 correlation. In Figure S5, a representative plot of observed vs. predicted values from Equations (1) and (2) is illustrated.

Table 2. Ranges of statistical values for training and external test sets for logBB validated models.

Validated MLR Models: 6 Different Test Sets with n = 7 or n = 6	R^2_{train}	s_{train}	R^2_{test}	s_{test}
Equation (1) (based on IAM retention)	0.626–0.705	0.388–0.499	0.583–0.932	0.167–0.669
Equation (2) (based on lipophilicity)	0.639–0.692	0.416–0.472	0.532–0.848	0.244–0.510
Validated PLS Models: 5 Different Test Sets with n = 8 or n = 9	$R^2_{\text{train}}/Q^2_{\text{train}}$	RMSEE	R^2_{test}	RMSEP
PLS Model 5 (based on IAM retention)	0.850–0.863/ 0.724–0.793	0.272–0.316	0.457*–0.936	0.302–0.525*
PLS Model 6 (based on lipophilicity)	0.850–0.865/ 0.725–0.791	0.272–0.328	0.600**–0.919	0.317–0.516**

R^2_{train} , coefficient of determination of the training sets; s_{train} , standard deviation of the training sets; R^2_{test} , coefficient of determination of the test sets; s_{test} , standard deviation of the test sets; Q^2_{train} , cross-validated coefficient of determination in PLS models; RMSEE, root mean square error of estimation; RMSEP, root mean square error of prediction. * Deteriorated statistics due to the presence of quinidine. ** Deteriorated statistics due to the presence of morphine.

2.3.2. Partial Least Squares Models

Initially, PLS analysis was performed considering both biomimetic properties and lipophilicity in the pool of descriptors, so as to rank their impact in modeling logBB. Polarity descriptors (TPSA and dipole moment) were found to have the highest VIP values (VIP = 1.2). AGP retention showed equally high impact, followed by $\log D_{7.4}$ (VIP = 1.05), and IAM retention (VIP = 0.95). HSA retention and $\log P$ had a lower influence in the model with $\text{VIP} < 0.9$.

Keeping AGP retention as the most influential parameter and including either $\log k_{w\text{IAM}}$ (PLS Model 5, Figure 4a,b) or $\log D_{7.4}$ (PLS Model 6, Figure 4c,d), 2 component PLS models with 19 descriptors were obtained after variable selection, with practically equal statistical quality:

$$A = 2, n = 42, R^2 = 0.846, Q^2 = 0.792, \text{RMSEE} = 0.313, \text{RMSEcv} = 0.351 \text{ (PLS Model 5)}$$

$$A = 2, n = 42, R^2 = 0.846, Q^2 = 0.791, \text{RMSEE} = 0.313, \text{RMSEcv} = 0.351 \text{ (PLS Model 6)}$$

The applicability domain was defined by the critical value D_{ModY} being $0.767 \times 2 = 1.534$ for Model 5 and $0.789 \times 2 = 1.58$ for Model 6 (Figure S6). A total of 4 drugs, i.e., atenolol, candesartan, haloperidol, and maprotiline, were outside the AD of PLS Model 5. The same drugs, plus indomethacin, were outside the AD of PLS Model 6 (Figure S6).

Considering the large percentage of missing $\log k_{w\text{AGP}}$ values, and in order to confirm its essential contribution, PLS analysis was repeated including only drugs with available $\log k_{w\text{AGP}}$ data. The generated models proved the crucial contribution of AGP retention (Figure S7).

The contribution of $\log k_{w\text{AGP}}$ in logBB was further confirmed in MLR models for the drugs with available data:

$$\log \text{BB} = 0.467(\pm 0.182) \log k_{w\text{AGP}} - 0.013(\pm 0.05) \text{TPSA} + 0.018(\pm 0.483)$$

$$n = 19, R = 0.871, R^2 = 0.758, R^2_{\text{adj}} = 0.728, s = 0.431, F = 25.1 \quad (3)$$

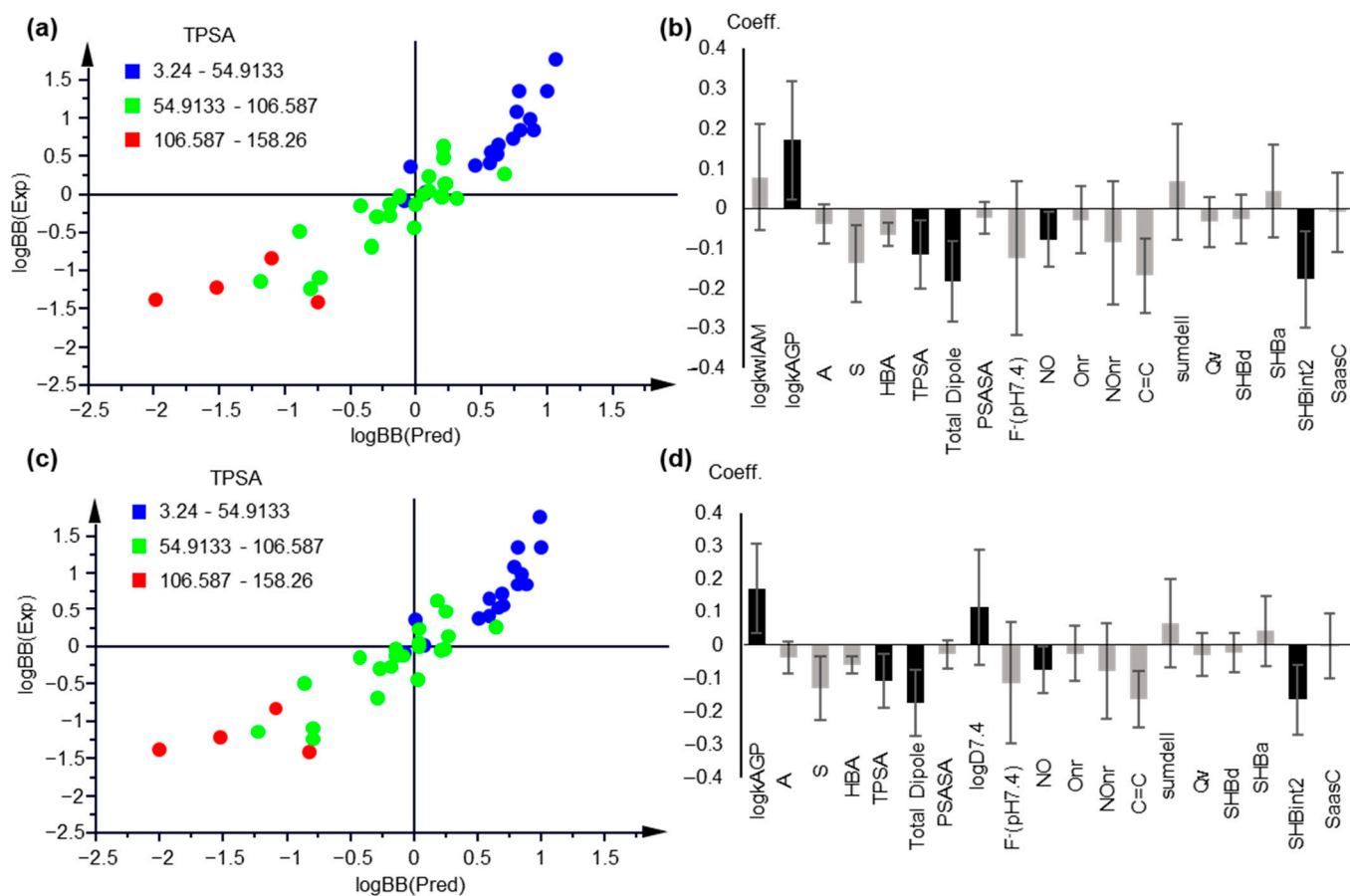


Figure 4. PLS modeling of logBB based on chromatographic and lipophilicity data. (a) Observed vs. predicted logBB values plot, based on PLS Model 5; (b) Coefficient plot of the original variables of PLS Model 5; (c) Observed vs. predicted logBB values plot, based on PLS Model 6; (d) Coefficient plot of the original variables of PLS Model 6. In (a,c), the compounds are colored based on their TPSA values. In (b,d), variables with VIP > 1 are highlighted with black.

The high impact of AGP retention may be related to the presence of a strong negative charge in AGP stationary phases due to their high content in sialic acid. Brain cell membranes are also the most anionic and have their lipids mostly exposed, thus explaining the reason that lipophilic cationic compounds are more prone to cross the BBB.

PLS Models 5 and 6 were validated with permutation tests (Figure S8) and by external validation upon dividing the dataset into 5 different training and test sets. Practically speaking, the same models were generated by an external validation procedure, with comparable statistical data (Table 2), with the exception of iteration, including quinidine (PLS Model 5) and morphine (PLS Model 6). In all cases, observed vs. predicted values showed a 1:1 correlation. In Figure S9, a representative plot of observed vs. predicted values is illustrated. The predicted logBB values from the MLR and PLS models are provided in Table S5.

2.4. Modeling $\log K_b$

2.4.1. Multiple Linear Regression

Since $\log K_b$ is related to tissue binding, a satisfactory negative correlation with $\log k_{w,IAM}$ values was obtained. Strong IAM retention corresponds to high tissue phospholipid binding and thereupon to a lower fraction being unbound:

$$\log K_b = -0.984(\pm 0.059) \log k_{w,IAM} + 0.966(\pm 0.131)$$

$$n = 39, R = 0.940, R^2 = 0.884, R^2_{adj} = 0.881, s = 0.543, F = 281.786 \quad (4)$$

A negative correlation with inferior but still acceptable statistics was obtained using HSA retention factors:

$$\log K_b = -1.291(\pm 0.105) \log k_{10, \text{HSA}} - 0.551(\pm 0.092)$$

$$n = 39, R = 0.897, R^2 = 0.804, R^2_{\text{adj}} = 0.799, s = 0.543, F = 152.245 \quad (5)$$

Further stepwise MLR analysis led to a three-parameter equation, combining the IAM and HSA retention factors with the count of nitrogen and oxygen atoms present in the molecule (NO):

$$\log K_b = -0.617(\pm 0.093) \log k_{w, \text{IAM}} - 0.450(\pm 0.126) \log k_{10, \text{HSA}} + 0.111(\pm 0.034) \text{NO} - 0.026(\pm 0.230)$$

$$n = 39, R = 0.966, R^2 = 0.932, R^2_{\text{adj}} = 0.927, s = 0.328, F = 160.776 \quad (6)$$

The positive sign of the coefficient of NO indicated that a higher hydrogen bond acceptor potential is not favorable for unspecific hydrophobic tissue and plasma protein binding, increasing the amount of the unbound fraction. It should be mentioned that $\log k_{w, \text{IAM}}$ and $\log k_{10, \text{ACN, HSA}}$ showed a considerable degree of intercorrelation ($r = 0.824$). To overcome the collinearity problem, and considering the low differentiation of regression coefficients of $\log k_{w, \text{IAM}}$ and $\log k_{10, \text{ACN, HSA}}$ in Equation (5) was within statistical limits, their sum $\text{Sum}_{\text{IAM-HSA}}$ was used instead:

$$\log K_b = -0.55(\pm 0.03) \text{Sum}_{\text{IAM-HSA}} + 0.11(\pm 0.03) \text{NO} - 0.14(\pm 0.18)$$

$$n = 39, R = 0.965, R^2 = 0.931, R^2_{\text{adj}} = 0.927, s = 0.327, F = 243.095, h^* = 0.231 \quad (7)$$

All drugs were within the AD (Figure S10).

Satisfactory correlation was also obtained with AGP retention factors for 20 drugs with available data:

$$\log K_b = -1.240(\pm 0.133) \log k_{w, \text{AGP}} + 1.034(\pm 0.264)$$

$$n = 20, R = 0.910, R^2 = 0.827, R^2_{\text{adj}} = 0.818, s = 0.458, F = 160.863 \quad (8)$$

Hydrophobic binding, a major force in drug-tissue and drug-protein interactions, was better simulated by $\log P$ of the neutral form than $\log D_{7.4}$. Thus, a satisfactory $\log K_b / \log P$ relationship was obtained—although with a lower correlation coefficient, compared to Equation (4):

$$\log K_b = -0.580(\pm 0.042) \log P + 0.614(\pm 0.139)$$

$$n = 39, R = 0.915, R^2 = 0.836, R^2_{\text{adj}} = 0.832, s = 0.497, F = 189.249, h^* = 0.153 \quad (9)$$

Further stepwise regression did not lead to an improved equation.

Correlation with the distribution coefficient $\log D_{7.4}$ led to an inferior relationship:

$$\log K_b = -0.555(\pm 0.042) \log D_{7.4} - 0.147(\pm 0.142)$$

$$n = 39, R = 0.823, R^2 = 0.677, R^2_{\text{adj}} = 0.668, s = 0.698, F = 77.626 \quad (10)$$

Equation (10), however, can be improved if the fraction of the ionized molecular species F^+ and F^- is introduced in combination with NO. The introduction of fractions F^+ and F^- reflects the binding conditions in respect to basic and acidic drugs, the first interacting stronger with phospholipids and the latter with serum albumin:

$$\log K_b = -0.54(\pm 0.05) \log D_{7.4} - 0.77(\pm 0.18) F^+ - 0.98(\pm 0.30) F^-_{7.4} + 0.19(\pm 0.05) \text{NO} - 0.30(\pm 0.28)$$

$$n = 39, R = 0.931, R^2 = 0.867, R^2_{\text{adj}} = 0.852, s = 0.467, F = 55.57, h^* = 0.307 \quad (11)$$

Evidently, all of the regression coefficients in Equation (11) had negative signs, except NO, which contributed positively. Three drugs were outside the AD, as shown in Figure S10.

External validation of Equations (7) and (11) was performed by dividing the dataset into the training set and the test set, using 5 different test sets (Table 3). In all training sets, observed vs. predicted values showed a 1:1 correlation. In Figure S11, a representative plot of observed vs. predicted values from Equations (7) and (11) is illustrated.

Table 3. Ranges of statistical values for the training and external test sets for $\log K_b$ validated models.

Validated MLR Models: 5 Different Test Sets with n = 8 or n = 9	R^2_{train}	S_{train}	R^2_{test}	S_{test}
Equation (7) (IAM retention)	0.924–0.937	0.306–0.341	0.823–0.966	0.270–0.377
Equation (11) (lipophilicity)	0.848–0.890	0.403–0.462	0.797–0.955	0.310–0.664
Validated PLS Models: 5 different test sets with n = 8 or = 9	$R^2_{\text{train}}/Q^2_{\text{train}}$	RMSEE	R^2_{test}	RMSEP
PLS Model 8 (based on IAM retention)	0.919–0.946/ 0.851–0.927	0.286–0.353	0.822–0.988	0.277–0.519
PLS Model 9 (based on lipophilicity)	0.751–0.931/ 0.669–0.857	0.358–0.610	0.711–0.939	0.402–0.676

R^2_{train} , coefficient of determination of the training sets; S_{train} , standard deviation of the MLR models on the training sets; R^2_{test} , coefficient of determination of the test sets; S_{test} , standard deviation of the test sets; Q^2_{train} , cross-validated coefficient of determination for PLS models; RMSEE, root mean square error of estimation; RMSEP, root mean square error of prediction.

2.4.2. Partial Least Squares Models

Application of PLS to model $\log K_b$ led to a two-component model (PLS Model 7) if both biomimetic properties and lipophilicity were included in the pool of descriptors:

$$n = 39, A = 2, R^2 = 0.929, Q^2 = 0.904, \text{RMSEE} = 0.332, \text{RMSEECv} = 0.371 \text{ (PLS Model 7)}$$

Biomimetic properties and lipophilicity were the most crucial parameters ($\text{VIP} > 1$), with decreasing importance following the order: $\log k_{\text{ACFP}} > \log k_{\text{wIAM}} > \log P > \log k_{10\text{HSA}} > \log D_{7.4}$ (Figure S12). In accordance with the MLR equations, the three biomimetic properties and lipophilicity had negative contributions, while the hydrogen bonding acceptor descriptor NO had a positive effect. More to the point, $\log P$ showed a higher impact in respect to $\log D_{7.4}$ as a better lipophilicity expression for hydrophobic binding. Three bulk descriptors, CMR, arC6, and nPSA, included in the model had a negative sign (Figure S12).

Including only biomimetic properties in the pool of descriptors, a 3-component model (PLS Model 8) was obtained with improved statistics:

$$A = 3, n = 39, R^2 = 0.938, Q^2 = 0.913, \text{RMSEE} = 0.315, \text{RMSEECv} = 0.353 \text{ (PLS Model 8)}$$

As illustrated in Figure 5, the same signs were kept in PLS Model 8 for biomimetic properties and NO, while a balance was observed between the non-polar descriptors, with nPSA having a positive sign.

According to the DModY criterion (critical value = $0.812 \times 2 = 1.624$), three drugs, i.e., fluoxetine, ranitidine, and theophylline were outside of the AD of the model (Figure S13).

Replacement of IAM retention by lipophilicity, expressed both as $\log P$ and $\log D_{7.4}$ led to a 3-component model (PLS Model 9):

$$A = 3, n = 39, R^2 = 0.925, Q^2 = 0.844, \text{RMSEE} = 0.347, \text{RMSEECv} = 0.478 \text{ (PLS Model 9)}$$

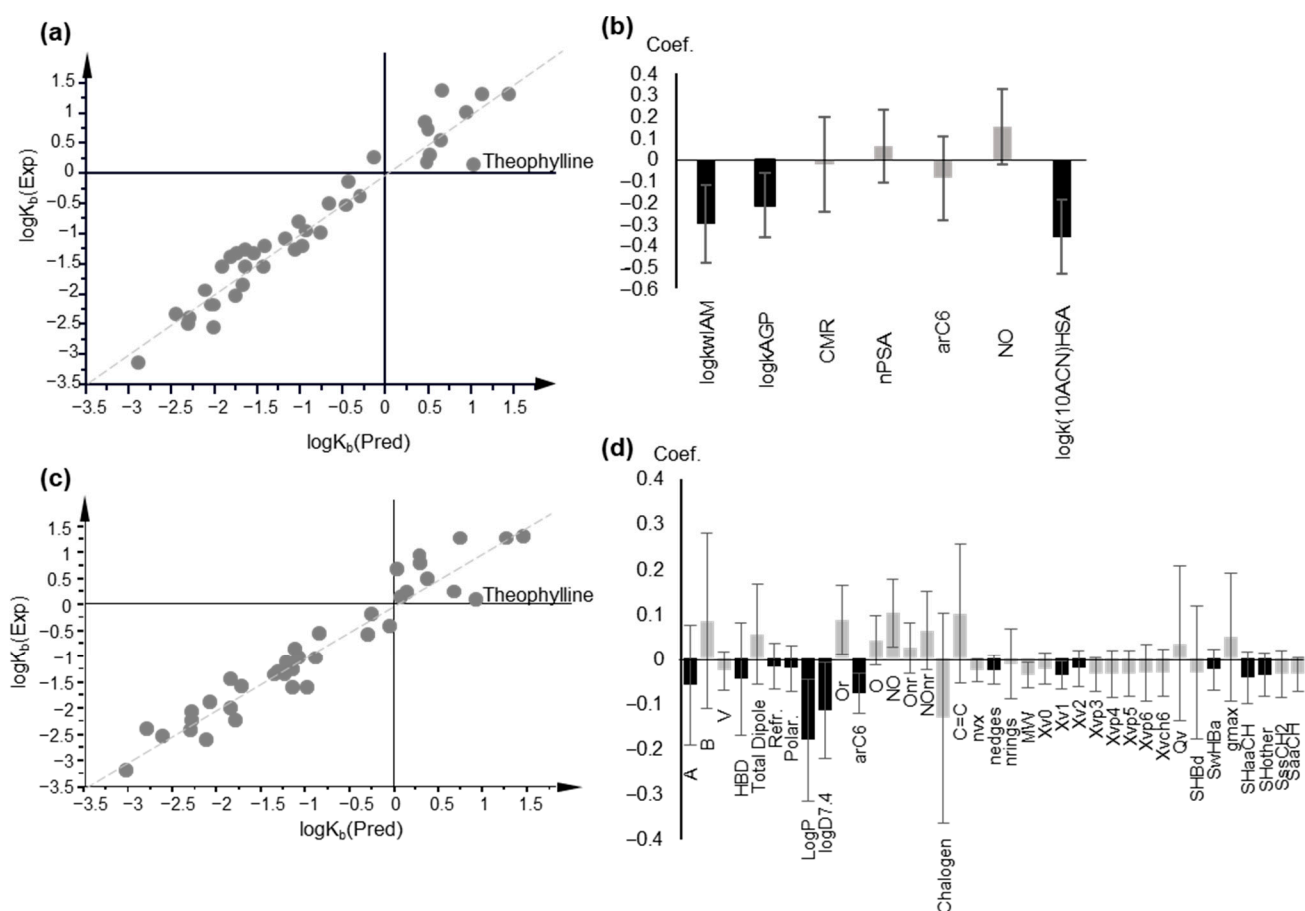


Figure 5. PLS modeling of $\log K_b$, based on chromatographic data and lipophilicity parameters. (a) Observed vs. predicted $\log K_b$ values plot based on PLS Model 8; (b) Coefficient plot of the original variables of PLS Model 8; (c) Plot of observed vs. predicted $\log K_b$ values, based on PLS Model 9; (d) Coefficient plot of the original variables of PLS Model 9; In (b,d), variables with $VIP > 1$ are highlighted in black.

Five drugs—acetaminophen, atenolol, neostigmine, propranolol, and theophylline—are the beyond AD, with D_{ModY} being higher than the critical value (2×0.768) (Figure S13).

The $\log K_b$ PLS Models 8 and 9 were validated with permutation tests (Figure S14), and upon dividing the dataset into 5 different training and test sets. Robust statistical data were obtained in all cases (Table 3). Representative plots of observed vs. predicted $\log K_b$ values from external validation are presented in Figure S15. The predicted $\log K_b$ values from the MLR and PLS models are presented in Table S6.

Back calculation of $f_{u,\text{brain}}$ values using the predicted $\log K_b$ were successfully correlated with the corresponding experimental values, approximating a 1:1 correlation (Table S7).

2.5. Unbound Volume of Distribution, V_u , Brain

2.5.1. Correlation between fraction unbound and unbound volume of distribution in the brain

The negative correlation of $\log V_{u,\text{brain}}$ with $\log f_{u,\text{brain}}$, shown also in Figure 1h, is reflected in Equation (12):

$$\log V_{u,\text{brain}} = -0.922(\pm 0.091) \log f_{u,\text{brain}} + 0.386(\pm 0.117)$$

$$n = 17, R = 0.934, R^2 = 0.873, R^2_{\text{adj}} = 0.864, s = 0.319, F = 102.678 \quad (12)$$

Equation (12) is considerably improved upon the inclusion of F^+ , which has a positive effect, indicating the importance of positive charge in the overall cellular uptake (Equation (13):

$$\log V_{u,\text{brain}} = -0.950(\pm 0.049) \log f_{u,\text{brain}} + 0.632(\pm 0.102) F^+_{7.4} - 0.113(\pm 0.102)$$

$$n = 17, R = 0.983, R^2 = 0.966, R^2_{\text{adj}} = 0.961, s = 0.167, F = 199.802 \quad (13)$$

The high quality of Equation (13) permits the safe prediction of $\log V_{u,\text{brain}}$ from $\log f_{u,\text{brain}}$. We further attempted to construct models for the unbound volume of distribution using biomimetic properties, lipophilicity, and computational descriptors although, in this case, the dataset was limited, including only 17 drugs with available experimental data.

2.5.2. Multiple Linear Regression

Direct correlation of $\log V_{u,\text{brain}}$ with $\log k_{w,\text{IAM}}$ led to Equation (14), with moderate statistics:

$$\log V_{u,\text{brain}} = 0.624(\pm 0.093) \log k_{w,\text{IAM}} + 0.226(\pm 0.190)$$

$$n = 17, R = 0.866, R^2 = 0.750, R^2_{\text{adj}} = 0.734, s = 0.446, F = 45.043 \quad (14)$$

Using stepwise regression, a considerably improved regression equation was obtained upon inclusion of the count of the hydrogen bond acceptors (HBA) as an additional parameter (Equation (15)):

$$\log V_{u,\text{brain}} = 0.623(\pm 0.071) \log k_{w,\text{IAM}} - 0.214(\pm 0.062) \text{HBA} + 0.958(\pm 0.256)$$

$$n = 17, R = 0.930, R^2 = 0.865, R^2_{\text{adj}} = 0.846, s = 0.339, F = 44.983, h^* = 0.53 \quad (15)$$

All compounds were within the AD of Equation (15) (Figure S16).

HSA retention was not significant, while AGP retention led to a very good correlation for the limited dataset of seven compounds:

$$\log V_{u,\text{brain}} = 0.822(\pm 0.155) \log k_{w,\text{AGP}} + 0.127(\pm 0.280)$$

$$n = 7, R = 0.921, R^2 = 0.849, R^2_{\text{adj}} = 0.819, s = 0.284, F = 28.07 \quad (16)$$

Replacing $\log k_{w,\text{IAM}}$ with lipophilicity, a moderate regression equation was obtained (Equation (17)), with $\log P$ in combination with the fraction protonated, including F^+ as an additional parameter:

$$\log V_{u,\text{brain}} = 0.336(\pm 0.054) \log P + 0.859(\pm 0.285) F^+_{7.4} - 0.115(\pm 0.290)$$

$$n = 17, R = 0.864, R^2 = 0.746, R^2_{\text{adj}} = 0.710, s = 0.465, F = 20.61, h^* = 0.53 \quad (17)$$

All compounds were within the AD of Equation (17) (Figure S16).

The combination of $\log P$ with the fraction protonated at pH7.4, both with a positive contribution, has previously been reported for a PLS model of the apparent volume of distribution [58].

Owing to the reduced dataset, test sets for the external validation of Equations (15) and (17) contained 4 to 6 compounds. Robust statistical data were obtained for validated equations, except when metformin was included in the test set, for both Equations (15) and (17) (Table 4). In Figure S17, a representative plot of observed vs. predicted values from Equations (15) and (17) is illustrated:

Table 4. Ranges of statistical values for the training and external test sets for models validated by $\log V_{u,brain}$.

Validated MLR Model: 3 Different Test Sets with n = 4–6	R^2_{train}	S_{train}	R^2_{test}	S_{test}
Equation (15) (IAM retention)	0.686–0.923	0.262–0.376	0.433 *–0.920	0.257–0.434
Equation (17) (lipophilicity)	0.526 *–0.852	0.364–0.485	0.437 *–0.950	0.132–0.429
Validated PLS Model: 2 different test sets with n = 8–9	R^2_{train}/Q^2_{train}	RMSEE	R^2_{test}	RMSEP
PLS Model 10 (based on IAM retention)	0.927, 0.962/ 0.690, 0.922	0.160, 0.268	0.718, 0.956	0.316, 0.656
PLS Model 11 (based on lipophilicity)	0.747, 0.962/ 0.747, 0.880	0.160, 0.251	0.718, 0.782	0.650, 0.656

R^2_{train} , coefficient of determination of the training sets; S_{train} , standard deviation of the MLR models on the training sets; R^2_{test} , coefficient of determination of the test sets; S_{test} , standard deviation of the test sets; Q^2_{train} , cross-validated coefficient of determination in PLS models; RMSEE, root mean square error of estimation; RMSEP, root mean square error of prediction; * Deteriorated statistics due to the presence of metformin.

2.5.3. Partial Least Squares (PLS) Models

A one-component PLS model, based on IAM retention, was obtained after variable selection with very good statistics (Figure 6a,b):

$$A = 1, n = 17, R^2 = 0.932, Q^2 = 0.901, RMSEE = 0.232, RMSEECv = 0.264 \text{ (PLS Model 10)}$$

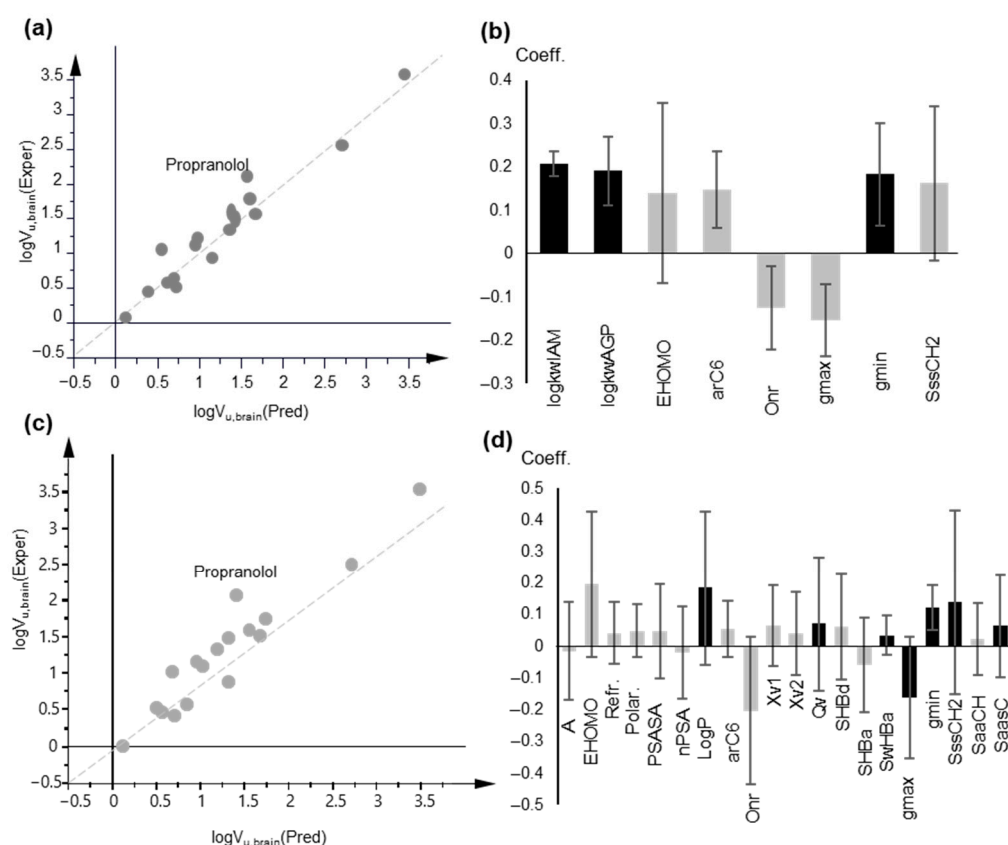


Figure 6. PLS modeling of $\log V_{u,brain}$ based on chromatographic data and lipophilicity parameters. (a) Observed vs. predicted $\log V_{u,brain}$ values plot based on PLS Model 10; (b) Coefficient plot of the original variables of PLS Model 10; (c) Observed vs. predicted $\log V_{u,brain}$ values plot based on PLS Model 11; (d) Coefficient plot of the original variables of PLS Model 11; In (b,d), variables with $VIP > 1$ are highlighted in black.

In agreement with the MLR models, IAM and AGP retention were the most influential variables with a positive contribution, while HSA retention factors were not included in the final model. Most polar descriptors had a negative effect, and the opposite was true for non-polar descriptors.

According to the DModY criterion (critical value: $0.802 \times 2 = 1.604$), 2 drugs (metformin and propranolol) were outside the AD of the model (Figure S18).

The use of logP in place of biomimetic properties led to a three-component model with satisfactory statistics although they had inferior cross-validation results (Figure 6c,d):

$$A = 3, n = 17, R^2 = 0.912, Q^2 = 0.723, RMSEE = 0.285, RMSEECv = 0.403 \text{ (PLS Model 11)}$$

According to the DModY criterion (critical value 1.524), 2 drugs (pindolol and propranolol) were outside the AD of the model (Figure S18).

The lower Q^2 and higher RMSEECv indicate inferior predictability, reflected also in the higher intercepts of the corresponding permutation tests (Figure S19). The PLS models were also validated by external validation, dividing the dataset into two training and test sets. In Table 4, the statistical data of the new models are given. In Figure S20, representative plots of observed vs. predicted $\log V_{u,brain}$ are provided. The predicted $\log V_{u,brain}$ values from the MLR and PLS models are presented in Table S8.

Considering the potential of Equation (13) for safe predictions of $\log V_{u,brain}$, we used Equation (13) in order to extend the dataset. Drugs with predicted $\log V_{u,brain}$ were used as a blind test set. As illustrated in Figure S21, the blind test set was well accommodated in the model with the exception of two drugs, clomipramine and fluoxetine, which were strong outliers. Excluding clomipramine and fluoxetine, the combination of the blind test set with the training set led to PLS Model 12, which was practically the same as PLS Model 10:

$$A = 1, n = 37, R^2 = 0.858, Q^2 = 0.845, RMSEE = 0.355, RMSEECv = 0.361 \text{ (PLS Model 12)}$$

For 4 drugs, DModY exceeded the critical value of 1.45, being outside the AD (Figure S22).

3. Materials and Methods

3.1. Dataset and Chromatographic Data

A dataset of 55 pharmaceutical compounds, belonging to a wide range of pharmacological classes, was used in the present investigation. The dataset exhibited adequate structural diversity, consisting of acidic, neutral, basic, and zwitterionic molecules. The compounds had been previously studied in our laboratory with respect to their retention profiles in 3 biomimetic chromatographic columns, namely an IAM.PC.DD2 (Regis Technology, Morton Grove, IL, USA), a ChromTech CHIRAL-HSA column (50 mm \times 4 mm i.d.), and a ChromTech CHIRAL-AGP column (50 \times 4 mm i.d.), under experimental conditions as described in the corresponding references [23,24,27,29,45].

For IAM chromatography, retention factors $\log k_{wIAM}$ corresponding to a 100% aqueous mobile phase were used. PBS was used as a buffer at two pH values, i.e., the physiological pH 7.4 (data labeled as $\log k_{wIAM}$) and pH 5.0 (data labeled as $\log k_{wIAM5.0}$), with the latter being associated with intestinal absorption and lysosomal trapping [59]. In the case of HSA chromatography, isocratic $\log k$ values, measured in the presence of 10% acetonitrile (ACN), were used ($\log k_{10,HSA}$), as they showed a highly significant 1:1 correlation with plasma protein binding data in our previous study [45]. $\log k_{10,HSA}$ data were available for 37 compounds in the dataset. For the remaining compounds, $\log k_{10ACN,HSA}$ values were calculated based on the highly significant equation reported in the same study [45]. $\log k_{wAGP}$ retention factors corresponding to a 100% aqueous mobile phase were available for 28 drugs. Thus, they could be used in the MLR models only for restricted datasets. However, their performance could be evaluated in PLS models since this type of analysis tolerates missing values. All chromatographic indices are included in Table S1.

3.2. Brain Disposition Data

Experimental $K_{p,uu,brain}$ values measured in rat brain tissue were collected for 22 compounds [10–12,60,61] and converted to the logarithmic form. Experimentally determined $K_{p,brain}$ values for 42 compounds were obtained from the literature and were converted to $\log BB$. They preferably referred to rat studies in order to achieve homogeneity in the data [19,31,32]. Experimental $f_{u,brain}$ values were available from the literature for 39 compounds in the dataset [10,50–53,62]. When more than one value was available, the average was calculated. Most values were determined on rats. Since there are only small inter-species differences, values from species other than rats were included (seven cases) for compounds where rat values were not provided. The $f_{u,brain}$ values were converted to the thermodynamic constant K_b according to the following equation [55]:

$$K_b = \frac{f_{u,brain}}{1.001 - f_{u,brain}}$$

The denominator was set to $1.001 - f_{u,brain}$ in the case of compounds exhibiting an $f_{u,brain}$ value equal to unity. The decimal logarithm of K_b , $\log K_b$, was used for the development of the models.

$V_{u,brain}$ values, determined using the brain slice method, were collected from the literature for 17 compounds in the dataset and converted to the corresponding logarithm, $\log V_{u,brain}$ [10,12].

$K_{p,brain}$, $K_{p,uu,brain}$, $f_{u,brain}$, and $V_{u,brain}$ values are presented in Table S2. They span a sufficient range and are evenly distributed, including drugs with different brain distribution profiles.

$K_{p,uu,brain}$ can be derived from $K_{p,brain}$ by combining the unbound fractions of the drug in the plasma ($f_{u,p}$) and in the brain ($f_{u,brain}$) or the unbound brain volume of distribution, $V_{u,brain}$ [7]:

$$K_{p,uu,brain} \sim \frac{K_{p,brain}}{V_{u,brain} \times f_{u,plasma}}$$

$V_{u,brain}$ shows an inverse relation to $V_{u,brain}$ shows an inverse relation to $f_{u,brain}$ approximating under circumstances [63]. $f_{u,brain} \sim [1/V_{u,brain}]$ range 0-1.

3.3. Physicochemical and Molecular Descriptors

A pool of 121 descriptors, including 1D, 2D and 3D descriptors, was calculated using appropriate software. The software ADME Boxes 3.0 was used to calculate the topological polar surface area (TPSA), hydrogen bond acceptor and donor sites (HBA and HBD), the rotatable bonds (RB), the number of ionizable groups, and the fraction of molecular species at pH 7.4 (with the fraction of negatively charged species being $F_{7.4}^-$ and that of positively charged species being $F_{7.4}^+$). The Abraham's solvation parameters (hydrogen bond acidity (A), hydrogen bond basicity (B), excess molar refraction (E), McGowan's volume (V), and dipolarity/polarizability (S)) were calculated with the module ABSOLV implemented using the same software. Topological indices and electrotopological state indices [64] were computed with Molconn-Z (v4.12 eduSoft, LC, La Jolla, CA 92037 USA) software Hyperchem v.5.0/Chemplus v.1.6 software (HYPERCUBE Inc., Waterloo, ON, Canada) was used for the calculation of 3D descriptors. Molecular size descriptors, energy parameters, and dipole moment were calculated using the lowest energetic conformation. Descriptors based on essential structural characteristics, such as the total number of rings (nRings), the number of phenyl rings, the number of heteroatoms inside the rings (Nr, Or, Sr), the number of heteroatoms outside of the rings (Nnr, Onr, and Snr), or the sum of them (N, O, and S), the number of different halogens (Cl, F, I, and Br) and the sum of them (halogen), and the number of double bonds between carbon atoms were derived manually.

Experimental lipophilicity data, $\log P$, $\log D_{7.4}$, and $\log D_{5.0}$, were taken from references [23,27,29] and the therein-cited literature sources. They are included in Table S1.

3.4. Statistical Analysis

Multiple linear regression analysis (MLR) was performed using SPSS v.22.0 software. Variable selection was performed by applying the stepwise algorithm, first in groups of descriptors, classified according to their physicochemical content, in order to exclude the less-relevant variables for the final selection. Variables with zero or very small variance were excluded, as were collinear variables, considering a correlation coefficient lower than 0.8. The models were evaluated by considering the values of R , R^2 , s (standard deviation), and F -test. For the significance of each individual variable, a t -test value of $|t| \geq 2$ was considered. The applicability domain (AD) of the regression equations was defined by calculating the critical leverage h^* according to the formula $h^* = 3(p + 1)/n$, where p is the number of parameters, and n is the number of compounds [65].

Models were validated by dividing the dataset into different training and test sets of 5 to 9 drugs, depending on the overall sample size. The test sets were randomly selected, taking however into account to cover all four quartiles of the PCA scores plot. Subsequently, each model derived by the training set was used to predict the response variable of the corresponding test set. The R , R^2 , and s for both the training and test sets were considered and compared to the original model. In addition, models were evaluated by the proximity of the relation of observed vs. predicted values to a 1:1 correlation, reflected by a slope close to 1 and an intercept close to 0.

Multivariate data analysis was performed using Simca-P 14.1 (Sartorius Stedim Biotech, Umeå, Sweden). Prior to analysis, data were scaled to unit variance. Principal component analysis (PCA) was performed, considering all columns of the table as X variables. PCA is a projection method resulting in dimensional reduction. The principal components, derived through projection, represent the new (latent) variables that summarize the information included in the initial set of descriptors. A PCA scores plot provides a useful data overview, which was also considered for the test set selection. Partial least squares analysis (PLS), a regression extension of PCA, was applied in order to construct prediction models. The variable selection was based on the values of the variable's influence to projection (VIP), the weight (w) in the loadings plot, and the size of coefficients. From this perspective, variables with variables with $VIP < 0.8$, low weight in the loading plot or low coefficient. Moreover, between descriptors encoding the same information, those which performed better were chosen. It should be mentioned, however, that PLS, as a projection method, can tolerate intercorrelated variables, as compared to multiple linear regression. The predictive ability and robustness of the models was evaluated using cross-validation as an internal validation according to the seven-fold option of Simca-P. The sum of the squared differences between the measured response and the predicted value of the omitted data, defined as predicting residuals sum of squares (PRESS), is used to calculate the cross-validated correlation coefficient, Q^2 : $Q^2 = 1 - \text{PRESS}/\text{SSY}$, with SSY representing the variation in Y after mean centering and scaling. Permutation tests (100 permutations) were applied by randomly re-ordering the response variables, and the newly derived R^2 and Q^2 were plotted against the degree of correlation between the permuted and the original data. The models were validated by external test sets, as already described for multiple linear regression. The root mean square error of prediction (RMSEP) was considered as an index of the predictability of the models. The applicability domain was defined using the double of the average value of the distance of observation to model Y (DMod Y) as the critical value.

4. Conclusions

In the present study, we combined bio-chromatographic retention factors, determined on IAM, *has*, and AGP stationary phases, with computed descriptors to develop MLR/PLS models for the rapid estimation of drug brain disposition. Our aim was to suggest a novel 'hybrid' modelling approach which combines high-throughput experimental accuracy with theoretical descriptors. At the same time, the well-investigated information content of the individual biomimetic properties [24,29,44,45] permitted a deeper insight into the underlying

ing mechanisms of biological processes. In this sense, diverse aspects of drug disposition in the CNS were explored, including the modeling of experimental $\log K_{p,uu,brain}$, $\log BB$, $f_{u,brain}$, and $V_{u,brain}$ data, upon the application of two statistical techniques, considered as contributing in a complementary way. MLR led to simple models that are easy to use by the medicinal chemist, whereas PLS served to strengthen the MLR models and to further scrutinize the biological issues inherent in brain disposition measurements.

$\log K_{p,uu,brain}$, as the outcome of contradictory factors related to permeability and tissue and plasma binding could not be efficiently modeled by biomimetic properties and lipophilicity, while computational descriptors alone were sufficient for model construction. Yet, PLS analysis revealed the greater effect of IAM, HSA, and AGP retention factors, as well as $\log D$, if measured at pH 5.0, supporting the potential interference of the endo-lysosomal system in vitro brain penetration models. In regard to $\log BB$, which measures permeability to CNS, traditional lipophilicity and IAM retention showed equal performance with positive contributions. However, in the cases of both $\log K_b$ and $\log V_{u,brain}$, which also depend on binding, IAM retention performed better than lipophilicity. More to the point, a lesser number of total descriptors was required in the corresponding PLS models, reflecting the higher information content of IAM retention. HSA retention was found to be important in $\log K_b$ modeling, while AGP retention influenced all brain disposition data, confirming the role of basicity as an essential CNS-drug-like characteristic. Polarity and hydrogen bond descriptors proved crucial in all models, with an opposite effect in regard to biomimetic properties and lipophilicity.

In view of the above findings, biomimetic properties are well-justified in modeling distinct brain disposition data as they reflect the major factors which govern biological processes, e.g., passive diffusion and binding. More to the point, IAM retention performs equally well or better than traditional lipophilicity. Hence, biomimetic chromatography can be suggested as a rapid and simple tool for the early evaluation of CNS drug candidates, permitting the construction of evidence-based 'hybrid' models and bridging the gap with in silico modeling, built solely on theoretical descriptors.

The 'hybrid' complementary models constructed in this study can be further applied to, and validated with, a wider range of pharmaceutical compounds. Combined with relevant plasma protein binding models also based on biomimetic properties [45], they can serve as a sound basis for exploring the composite brain disposition end-points.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27123668/s1>: Figure S1: DmodY bar plot for PLS Model 4 of $\log K_{p,uu,brain}$, including computational descriptors; Figure S2: Permutation test (based on 100 permutations) for PLS Model 4 of $\log K_{p,uu,brain}$; Figure S3: Representative plot of observed vs. predicted $\log K_{p,uu,brain}$ values from the external validation of PLS Model 4; Figure S4: Williams plots of standardized residuals vs. the leverage; Figure S5: Representative plot of observed vs. predicted $\log BB$ values from external validation of (a) Equation (1) and (b) Equation (2); Figure S6: (a) DmodY bar plot for PLS Model 5 of $\log BB$ prediction, including IAM retention factors; (b) DmodY bar plot for PLS Model 6 of $\log BB$ prediction, including $\log D_{7.4}$; Figure S7: PLS models of $\log BB$ prediction including only the compounds with experimental $\log k_{wAGP}$ values; Figure S8: Permutation tests for the PLS models of $\log BB$ prediction; Figure S9: Representative plot of observed vs. predicted $\log BB$ values from the external validation of the PLS model; Figure S10: Williams plots of standardized residuals vs. the leverage; Figure S11: Representative plot of observed vs. predicted $\log K_b$ values from external validation of (a) Equation (7) and (b) Equation (11); Figure S12: (a) Variable Importance ranking and (b) Coefficient plot of the original variables in PLS Model 7; Figure S13: (a) DmodY bar plot for PLS Model 8 of $\log K_b$ prediction, including IAM retention factors and (b) DmodY bar plot for PLS Model 9 of $\log K_b$ prediction, including $\log P$ and $\log D_{7.4}$; Figure S14: Permutation tests for the PLS models of $\log K_b$ prediction; Figure S15: Representative plot of observed vs. predicted $\log K_b$ values from the external validation of the PLS model; Figure S16: Williams plots of standardized residuals vs. the leverage; Figure S17: Representative plot of observed vs. predicted $\log V_{u,brain}$ values from the external validation of (a) Equation (15) and (b) Equation (17); Figure S18: (a) DmodY bar plot for PLS Model 10 of $\log V_{u,brain}$ prediction, including IAM retention factors; (b) DmodY

bar plot for PLS Model 11 of logV_{u,brain} prediction, including logP; Figure S19: Permutation tests for the PLS models of logV_{u,brain} prediction; Figure S20: Representative plot of observed vs. predicted logV_{u,brain} values from the external validation of the PLS model; Figure S21: Plot of observed vs. predicted logV_{u,brain} values from PLS Model 12; Figure S22: DmodY bar plot for PLS Model 12 of logV_{u,brain}. Table S1: Compounds included in the dataset: pharmaceutical classification, chromatographic data, and lipophilicity parameters; Table S2: Brain disposition data, i.e., K_{p,brain}, K_{p,uu,brain}, f_{u,brain}, and V_{u,brain} for the investigated compounds; Table S3: Variable Importance to Projection (VIP) values in PLS Model 4 for logK_{p,uu,brain} on the basis of biomimetic properties, lipophilicity, and computational descriptors; Table S4: Observed and predicted logK_{p,uu,brain} values by PLS Model 4, based on computational descriptors; Table S5: Experimental and predicted logBB values by MLR and PLS models; Table S6: Experimental and predicted logK_b values by MLR and PLS models; Table S7: Observed vs. predicted f_{u,brain} correlation: f_{u,brain} = a × f_{u,brain,pred} + b; Table S8: Experimental and predicted logV_{u,brain} values by MLR and PLS models.

Author Contributions: Conceptualization, T.V. and A.T.-K.; methodology, T.V., F.T. and A.T.-K.; validation, T.V. and A.T.-K.; formal analysis, T.V. and A.T.-K.; investigation, T.V. and A.T.-K.; data curation, F.T. and A.T.-K.; writing—original draft preparation, T.V.; writing—review and editing, T.V., F.T. and A.T.-K.; visualization, T.V. and A.T.-K.; supervision, A.T.-K. All authors have read and agreed to the published version of the manuscript.

Funding: T.V. is grateful for the funding provided to Per by the Swedish Research Council (grant 2018-03320) and the Science for Life Laboratory.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All experimental and predicted brain disposition data are provided in the Supplementary Materials. All experimental chromatographic and lipophilicity data are provided in the Supplementary Materials.

Acknowledgments: T.V. would like to acknowledge the kind and valuable support provided by Per Andrén, head of Medical Mass Spectrometry Imaging, Dept. of Pharmaceutical Biosciences, Uppsala University.

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

Sample Availability: Not applicable.

References

1. Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **2004**, *3*, 711–715. [[CrossRef](#)]
2. Choi, D.W.; Armitage, R.; Brady, L.S.; Coetzee, T.; Fisher, W.; Hyman, S.; Pande, A.; Paul, S.; Potter, W.; Roin, B.; et al. Medicines for the mind: Policy-based “pull” incentives for creating breakthrough CNS drugs. *Neuron* **2014**, *84*, 554–563. [[CrossRef](#)] [[PubMed](#)]
3. Abbott, N.J.; Friedman, A. Overview and introduction: The blood-brain barrier in health and disease. *Epilepsia* **2012**, *53* (Suppl. 6), 1–6. [[CrossRef](#)]
4. Abbott, N.J.; Patabendige, A.A.; Dolman, D.E.; Yusof, S.R.; Begley, D.J. Structure and function of the blood-brain barrier. *Neurobiol. Dis.* **2010**, *37*, 13–25. [[CrossRef](#)]
5. Abbott, N.J. Astrocyte-endothelial interactions and blood-brain barrier permeability. *J. Anat.* **2002**, *200*, 629–638. [[CrossRef](#)]
6. Eyal, S.; Hsiao, P.; Unadkat, J.D. Drug interactions at the blood-brain barrier: Fact or fantasy? *Pharmacol. Ther.* **2009**, *123*, 80–104. [[CrossRef](#)]
7. Loryan, I.; Sinha, V.; Mackie, C.; Van Peer, A.; Drinkenburg, W.; Vermeulen, A.; Morrison, D.; Monshouwer, M.; Heald, D.; Hammarlund-Udenaes, M. Mechanistic understanding of brain drug disposition to optimize the selection of potential neurotherapeutics in drug discovery. *Pharm. Res.* **2014**, *31*, 2203–2219. [[CrossRef](#)]
8. Hammarlund-Udenaes, M.; Bredberg, U.; Friden, M. Methodologies to assess brain drug delivery in lead optimization. *Curr. Top. Med. Chem.* **2009**, *9*, 148–162. [[CrossRef](#)] [[PubMed](#)]
9. Hammarlund-Udenaes, M.; Friden, M.; Syvanen, S.; Gupta, A. On the rate and extent of drug delivery to the brain. *Pharm. Res.* **2008**, *25*, 1737–1750. [[CrossRef](#)] [[PubMed](#)]
10. Friden, M.; Bergstrom, F.; Wan, H.; Rehngrén, M.; Ahlin, G.; Hammarlund-Udenaes, M.; Bredberg, U. Measurement of unbound drug exposure in brain: Modeling of pH partitioning explains diverging results between the brain slice and brain homogenate methods. *Drug Metab. Dispos.* **2011**, *39*, 353–362. [[CrossRef](#)]

11. Friden, M.; Winiwarter, S.; Jerndal, G.; Bengtsson, O.; Wan, H.; Bredberg, U.; Hammarlund-Udenaes, M.; Antonsson, M. Structure-brain exposure relationships in rat and human using a novel data set of unbound drug concentrations in brain interstitial and cerebrospinal fluids. *J. Med. Chem.* **2009**, *52*, 6233–6243. [[CrossRef](#)]
12. Friden, M.; Ducrozet, F.; Middleton, B.; Antonsson, M.; Bredberg, U.; Hammarlund-Udenaes, M. Development of a high-throughput brain slice method for studying drug distribution in the central nervous system. *Drug Metab. Dispos.* **2009**, *37*, 1226–1233. [[CrossRef](#)] [[PubMed](#)]
13. Luptáková, D.; Vallianatou, T.; Nilsson, A.; Shariatgorji, R.; Hammarlund-Udenaes, M.; Loryan, I.; Andrén, P.E. Neuropharmacokinetic visualization of regional and subregional unbound antipsychotic drug transport across the blood-brain barrier. *Mol. Psychiatry* **2021**, *26*, 7732–7745. [[CrossRef](#)]
14. Loryan, I.; Sinha, V.; Mackie, C.; Van Peer, A.; Drinkenburg, W.H.; Vermeulen, A.; Heald, D.; Hammarlund-Udenaes, M.; Wassvik, C.M. Molecular properties determining unbound intracellular and extracellular brain exposure of CNS drug candidates. *Mol. Pharm.* **2015**, *12*, 520–532. [[CrossRef](#)] [[PubMed](#)]
15. Abbott, N.J. Prediction of blood-brain barrier permeation in drug discovery from in vivo, in vitro and in silico models. *Drug Discov. Today Technol.* **2004**, *1*, 407–416. [[CrossRef](#)]
16. Wang, Z.; Yang, H.; Wu, Z.; Wang, T.; Li, W.; Tang, Y.; Liu, G. In Silico Prediction of Blood-Brain Barrier Permeability of Compounds by Machine Learning and Resampling Methods. *ChemMedChem* **2018**, *13*, 2189–2201. [[CrossRef](#)]
17. Clark, D.E. In silico prediction of blood-brain barrier permeation. *Drug Discov. Today* **2003**, *8*, 927–933. [[CrossRef](#)]
18. Zhang, L.; Zhu, H.; Oprea, T.I.; Golbraikh, A.; Tropsha, A. QSAR modeling of the blood-brain barrier permeability for diverse organic compounds. *Pharm. Res.* **2008**, *25*, 1902–1914. [[CrossRef](#)] [[PubMed](#)]
19. Bergstrom, C.A.; Charman, S.A.; Nicolazzo, J.A. Computational prediction of CNS drug exposure based on a novel in vivo dataset. *Pharm. Res.* **2012**, *29*, 3131–3142. [[CrossRef](#)]
20. Varadharajan, S.; Winiwarter, S.; Carlsson, L.; Engkvist, O.; Anantha, A.; Kogej, T.; Friden, M.; Stalring, J.; Chen, H. Exploring in silico prediction of the unbound brain-to-plasma drug concentration ratio: Model validation, renewal, and interpretation. *J. Pharm. Sci.* **2015**, *104*, 1197–1206. [[CrossRef](#)]
21. Chrysanthakopoulos, M.; Kolletsou, A.; Nicolaou, I.; Demopoulos, V.J.; Tsantili-Kakoulidou, A. Lipophilicity Studies on Pyrrolyl-Acetic Acid Derivatives. Experimental Versus Predicted logP Values in Relationship with Aldose Reductase Inhibitory Activity. *QSAR Comb. Sci.* **2009**, *28*, 551–560. [[CrossRef](#)]
22. Lanevskij, K.; Dapkunas, J.; Juska, L.; Japertas, P.; Didziapetris, R. QSAR analysis of blood-brain distribution: The influence of plasma and brain tissue binding. *J. Pharm. Sci.* **2011**, *100*, 2147–2160. [[CrossRef](#)] [[PubMed](#)]
23. Vrakas, D.; Giaginis, C.; Tsantili-Kakoulidou, A. Different retention behavior of structurally diverse basic and neutral drugs in immobilized artificial membrane and reversed-phase high performance liquid chromatography: Comparison with octanol-water partitioning. *J. Chromatogr. A* **2006**, *1116*, 158–164. [[CrossRef](#)]
24. Vrakas, D.; Giaginis, C.; Tsantili-Kakoulidou, A. Electrostatic interactions and ionization effect in immobilized artificial membrane retention. A comparative study with octanol-water partitioning. *J. Chromatogr. A* **2008**, *1187*, 67–78. [[CrossRef](#)] [[PubMed](#)]
25. Valkó, K.L. Lipophilicity and biomimetic properties measured by HPLC to support drug discovery. *J. Pharm. Biomed. Anal.* **2016**, *130*, 35–54. [[CrossRef](#)]
26. Valko, K.; Nunhuck, S.; Bevan, C.; Abraham, M.H.; Reynolds, D.P. Fast gradient HPLC method to determine compounds binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity. *J. Pharm. Sci.* **2003**, *92*, 2236–2248. [[CrossRef](#)]
27. Tsopelas, F.; Vallianatou, T.; Tsantili-Kakoulidou, A. The potential of immobilized artificial membrane chromatography to predict human oral absorption. *Eur. J. Pharm. Sci.* **2016**, *81*, 82–93. [[CrossRef](#)] [[PubMed](#)]
28. Tsopelas, F.; Vallianatou, T.; Tsantili-Kakoulidou, A. Advances in immobilized artificial membrane (IAM) chromatography for novel drug discovery. *Expert Opin. Drug Discov.* **2016**, *11*, 473–488. [[CrossRef](#)] [[PubMed](#)]
29. Tsopelas, F.; Malaki, N.; Vallianatou, T.; Chrysanthakopoulos, M.; Vrakas, D.; Ochsenkuhn-Petropoulou, M.; Tsantili-Kakoulidou, A. Insight into the retention mechanism on immobilized artificial membrane chromatography using two stationary phases. *J. Chromatogr. A* **2015**, *1396*, 25–33. [[CrossRef](#)]
30. Grumetto, L.; Russo, G.; Barbato, F. Indexes of polar interactions between ionizable drugs and membrane phospholipids measured by IAM-HPLC: Their relationships with data of Blood-Brain Barrier passage. *Eur. J. Pharm. Sci.* **2014**, *65*, 139–146. [[CrossRef](#)]
31. Grumetto, L.; Carpentiero, C.; Di Vaio, P.; Frecentese, F.; Barbato, F. Lipophilic and polar interaction forces between acidic drugs and membrane phospholipids encoded in IAM-HPLC indexes: Their role in membrane partition and relationships with BBB permeation data. *J. Pharm. Biomed. Anal.* **2013**, *75*, 165–172. [[CrossRef](#)]
32. Grumetto, L.; Carpentiero, C.; Barbato, F. Lipophilic and electrostatic forces encoded in IAM-HPLC indexes of basic drugs: Their role in membrane partition and their relationships with BBB passage data. *Eur. J. Pharm. Sci.* **2012**, *45*, 685–692. [[CrossRef](#)] [[PubMed](#)]
33. Giaginis, C.; Tsantili-Kakoulidou, A. Alternative measures of lipophilicity: From octanol-water partitioning to IAM retention. *J. Pharm. Sci.* **2008**, *97*, 2984–3004. [[CrossRef](#)]
34. Verzele, D.; Lynen, F.; De Vrieze, M.; Wright, A.G.; Hanna-Brown, M.; Sandra, P. Development of the first sphingomyelin biomimetic stationary phase for immobilized artificial membrane (IAM) chromatography. *Chem. Commun.* **2012**, *48*, 1162–1164. [[CrossRef](#)] [[PubMed](#)]

35. Salminen, T.; Pulli, A.; Taskinen, J. Relationship between immobilised artificial membrane chromatographic retention and the brain penetration of structurally diverse drugs. *J. Pharm. Biomed. Anal.* **1997**, *15*, 469–477. [[CrossRef](#)]
36. Reichel, A.; Begley, D.J. Potential of immobilized artificial membranes for predicting drug penetration across the blood-brain barrier. *Pharm. Res.* **1998**, *15*, 1270–1274. [[CrossRef](#)]
37. Norinder, U.; Osterberg, T. Theoretical calculation and prediction of drug transport processes using simple parameters and partial least squares projections to latent structures (PLS) statistics. The use of electrotopological state indices. *J. Pharm. Sci.* **2001**, *90*, 1076–1085. [[CrossRef](#)] [[PubMed](#)]
38. Lázaro, E.; Ràfols, C.; Abraham, M.H.; Rosés, M. Chromatographic estimation of drug disposition properties by means of immobilized artificial membranes (IAM) and C18 columns. *J. Med. Chem.* **2006**, *49*, 4861–4870. [[CrossRef](#)]
39. Hollósy, F.; Valkó, K.; Hersey, A.; Nunhuck, S.; Kéri, G.; Bevan, C. Estimation of volume of distribution in humans from high throughput HPLC-based measurements of human serum albumin binding and immobilized artificial membrane partitioning. *J. Med. Chem.* **2006**, *49*, 6958–6971. [[CrossRef](#)]
40. De Vrieze, M.; Lynen, F.; Chen, K.; Szucs, R.; Sandra, P. Predicting drug penetration across the blood-brain barrier: Comparison of micellar liquid chromatography and immobilized artificial membrane liquid chromatography. *Anal. Bioanal. Chem.* **2013**, *405*, 6029–6041. [[CrossRef](#)]
41. Dash, A.K.; Elmquist, W.F. Separation methods that are capable of revealing blood-brain barrier permeability. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2003**, *797*, 241–254. [[CrossRef](#)]
42. Valko, K.; Rava, S.; Bunally, S.; Anderson, S. Revisiting the application of Immobilized Artificial Membrane (IAM) chromatography to estimate in vivo distribution properties of drug discovery compounds based on the model of marketed drugs. *ADMET DMPK* **2020**, *8*, 78–97. [[CrossRef](#)] [[PubMed](#)]
43. Yoon, C.H.; Kim, S.J.; Shin, B.S.; Lee, K.C.; Yoo, S.D. Rapid screening of blood-brain barrier penetration of drugs using the immobilized artificial membrane phosphatidylcholine column chromatography. *J. Biomol. Screen* **2006**, *11*, 13–20. [[CrossRef](#)]
44. Chrysanthakopoulos, M.; Vallianatou, T.; Giaginis, C.; Tsantili-Kakoulidou, A. Investigation of the retention behavior of structurally diverse drugs on alpha(1) acid glycoprotein column: Insight on the molecular factors involved and correlation with protein binding data. *Eur. J. Pharm. Sci.* **2014**, *60*, 24–31. [[CrossRef](#)]
45. Chrysanthakopoulos, M.; Giaginis, C.; Tsantili-Kakoulidou, A. Retention of structurally diverse drugs in human serum albumin chromatography and its potential to simulate plasma protein binding. *J. Chromatogr. A* **2010**, *1217*, 5761–5768. [[CrossRef](#)] [[PubMed](#)]
46. Bteich, M. An overview of albumin and alpha-1-acid glycoprotein main characteristics: Highlighting the roles of amino acids in binding kinetics and molecular interactions. *Heliyon* **2019**, *5*, e02879. [[CrossRef](#)] [[PubMed](#)]
47. Lambrinidis, G.; Vallianatou, T.; Tsantili-Kakoulidou, A. In vitro, in silico and integrated strategies for the estimation of plasma protein binding. A review. *Adv. Drug Deliv. Rev.* **2015**, *86*, 27–45. [[CrossRef](#)] [[PubMed](#)]
48. Wan, H.; Ahman, M.; Holmen, A.G. Relationship between brain tissue partitioning and microemulsion retention factors of CNS drugs. *J. Med. Chem.* **2009**, *52*, 1693–1700. [[CrossRef](#)]
49. Wan, H.; Rehgren, M.; Giordanetto, F.; Bergstrom, F.; Tunek, A. High-throughput screening of drug-brain tissue binding and in silico prediction for assessment of central nervous system drug delivery. *J. Med. Chem.* **2007**, *50*, 4606–4615. [[CrossRef](#)]
50. Summerfield, S.G.; Stevens, A.J.; Cutler, L.; del Carmen Osuna, M.; Hammond, B.; Tang, S.P.; Hersey, A.; Spalding, D.J.; Jeffrey, P. Improving the in vitro prediction of in vivo central nervous system penetration: Integrating permeability, P-glycoprotein efflux, and free fractions in blood and brain. *J. Pharmacol. Exp. Ther.* **2006**, *316*, 1282–1290. [[CrossRef](#)]
51. Mateus, A.; Matsson, P.; Artursson, P. Rapid measurement of intracellular unbound drug concentrations. *Mol. Pharm.* **2013**, *10*, 2467–2478. [[CrossRef](#)]
52. Longhi, R.; Corbioli, S.; Fontana, S.; Vinco, F.; Braggio, S.; Helmdach, L.; Schiller, J.; Boriss, H. Brain tissue binding of drugs: Evaluation and validation of solid supported porcine brain membrane vesicles (TRANSIL) as a novel high-throughput method. *Drug Metab. Dispos.* **2011**, *39*, 312–321. [[CrossRef](#)] [[PubMed](#)]
53. Kalvass, J.C.; Maurer, T.S. Influence of nonspecific brain and plasma binding on CNS exposure: Implications for rational drug discovery. *Biopharm. Drug Dispos.* **2002**, *23*, 327–338. [[CrossRef](#)] [[PubMed](#)]
54. Ball, K.; Bouzom, F.; Scherrmann, J.M.; Walther, B.; Declèves, X. Physiologically based pharmacokinetic modelling of drug penetration across the blood-brain barrier—towards a mechanistic IVIVE-based approach. *AAPS J.* **2013**, *15*, 913–932. [[CrossRef](#)] [[PubMed](#)]
55. Lanevskij, K.; Japertas, P.; Didziapetris, R.; Petrauskas, A. Ionization-specific prediction of blood-brain permeability. *J. Pharm. Sci.* **2009**, *98*, 122–134. [[CrossRef](#)] [[PubMed](#)]
56. Toth, A.E.; Nielsen, S.S.E.; Tomaka, W.; Abbott, N.J.; Nielsen, M.S. The endo-lysosomal system of bEnd.3 and hCMEC/D3 brain endothelial cells. *Fluids Barriers CNS* **2019**, *16*, 14. [[CrossRef](#)]
57. Eriksson, L.; Jaworska, J.; Worth, A.P.; Cronin, M.T.; McDowell, R.M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361–1375. [[CrossRef](#)]
58. Karalis, V.; Tsantili-Kakoulidou, A.; Macheras, P. Multivariate statistics of disposition pharmacokinetic parameters for structurally unrelated drugs used in therapeutics. *Pharm. Res.* **2002**, *19*, 1827–1834. [[CrossRef](#)] [[PubMed](#)]
59. Daniel, W.A.; Wojcikowski, J. Lysosomal trapping as an important mechanism involved in the cellular distribution of perazine and in pharmacokinetic interaction with antidepressants. *Eur. Neuropsychopharmacol.* **1999**, *9*, 483–491. [[CrossRef](#)]

60. Friden, M.; Gupta, A.; Antonsson, M.; Bredberg, U.; Hammarlund-Udenaes, M. In vitro methods for estimating unbound drug concentrations in the brain interstitial and intracellular fluids. *Drug Metab. Dispos.* **2007**, *35*, 1711–1719. [[CrossRef](#)]
61. Friden, M.; Ljungqvist, H.; Middleton, B.; Bredberg, U.; Hammarlund-Udenaes, M. Improved measurement of drug exposure in the brain using drug-specific correction for residual blood. *J. Cereb. Blood Flow Metab.* **2010**, *30*, 150–161. [[CrossRef](#)] [[PubMed](#)]
62. Chen, H.; Winiwarter, S.; Friden, M.; Antonsson, M.; Engkvist, O. In silico prediction of unbound brain-to-plasma concentration ratio using machine learning algorithms. *J. Mol. Graph. Model.* **2011**, *29*, 985–995. [[CrossRef](#)] [[PubMed](#)]
63. Spreafico, M.; Jacobson, M.P. In silico prediction of brain exposure: Drug free fraction, unbound brain to plasma concentration ratio and equilibrium half-life. *Curr. Top. Med. Chem.* **2013**, *13*, 813–820. [[CrossRef](#)]
64. Kier, L.B.; Hall, L.H. An electrotopological-state index for atoms in molecules. *Pharm. Res.* **1990**, *7*, 801–807. [[CrossRef](#)]
65. Gramatica, P. On the development and validation of QSAR models. *Methods Mol. Biol.* **2013**, *930*, 499–526. [[CrossRef](#)] [[PubMed](#)]