

Article

Affinity of Compounds for Phosphatidylcholine-Based Immobilized Artificial Membrane—A Measure of Their Bioconcentration in Aquatic Organisms

Anna W. Sobańska 

Department of Analytical Chemistry, Faculty of Pharmacy, Medical University of Lodz, ul. Muszyńskiego 1, 90-151 Lodz, Poland; anna.sobanska@umed.lodz.pl

Abstract: The *BCF* (bioconcentration factor) of solutes in aquatic organisms is an important parameter because many undesired chemicals enter the ecosystem and affect the wildlife. Chromatographic retention factor $\log k_w^{IAM}$ obtained from immobilized artificial membrane (IAM) HPLC chromatography with buffered, aqueous mobile phases and calculated molecular descriptors obtained for a group of 120 structurally unrelated compounds were used to generate useful models of *BCF*. It was established that $\log k_w^{IAM}$ obtained in the conditions described in this study is not sufficient as a sole predictor of bioconcentration. Simple, potentially useful models based on $\log k_w^{IAM}$ and a selection of readily available, calculated descriptors and accounting for over 88% of total variability were generated using multiple linear regression (MLR), partial least squares (PLS) regression and artificial neural networks (ANN). The models proposed in the study were tested on an external group of 120 compounds and on a group of 40 compounds with known experimental *BCF* values. It was established that a relatively simple MLR model containing four independent variables leads to satisfying *BCF* predictions and is more intuitive than PLS or ANN models.



Citation: Sobańska, A.W. Affinity of Compounds for Phosphatidylcholine-Based Immobilized Artificial Membrane—A Measure of Their Bioconcentration in Aquatic Organisms. *Membranes* **2022**, *12*, 1130. <https://doi.org/10.3390/membranes12111130>

Academic Editor: Che-Ming Jack Hu

Received: 29 September 2022

Accepted: 7 November 2022

Published: 11 November 2022

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



Copyright: © 2022 by the author. 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/>).

Keywords: Immobilized artificial membrane; liquid chromatography; multiple linear regression; partial least squares regression; artificial neural networks; bioconcentration factor

1. Introduction

Immobilized artificial membrane (IAM) chromatography is a valuable technique used to predict the behavior of compounds towards biological membranes. IAM stationary phases based on phosphatidylcholine (PC) covalently linked to aminopropyl silica are able to mimic the natural membrane bilayer [1]. Thanks to this ability, they have become widely recognized tools for modeling drug distribution in vitro, with applications in medicinal chemistry including estimation of lipophilicity (a key feature characterizing the biological distribution of compounds), prediction of the ability of compounds to cross biological membranes (skin absorption, blood–brain barrier permeability, oral/human intestinal absorption) and estimation of other biomimetic properties (e.g., volume of distribution or Caco-2 permeability) [2,3]. More recently, immobilized artificial membrane chromatography has attracted the attention of environmental chemists, who used IAM chromatography to study the bioconcentration of pharmaceuticals [4], ecotoxicity of pesticides (expressed as LC₅₀) [5] and mobility of substances in soil [6]. Applications of IAM chromatography and other phospholipid-based in vitro techniques (liposome partitioning and chromatography on unbound phosphatidylcholine stationary phases) in the studies of drug–biomembrane interactions are presented in reviews [2,3].

Anthropogenic compounds enter the aquatic environment via a number of routes, pose a threat to aquatic organisms, accumulate in their tissues and affect their fertility. The risks associated with the exposure of aquatic organisms to chemical compounds released to the environment by humans have been studied extensively, e.g., for organic sunscreens [7–11], per- and polyfluoroalkyl compounds [12], polycyclic aromatic hydrocarbons [13] or antibiotics [14].

There is a need to identify compounds that are potentially hazardous—bioaccumulative, persistent and toxic in the environment. The fish bioconcentration factor (*BCF*) is the ratio of the chemical concentration in the organism (C_B) and water (C_W), accounting for the absorption via the respiratory route (e.g., gills) and skin. It is commonly used to screen chemicals for their bioaccumulation potential [15], especially in the absence of the bioaccumulation factor (*BAF*), which accounts for dietary, dermal and respiratory exposures. When neither *BAF* nor *BCF* data are available, lipophilicity expressed as the octanol–water partition coefficient K_{ow} is used as a surrogate measure of compounds' ability to bioaccumulate. The criteria of bioaccumulation differ depending on regulatory agency; it is accepted that compounds that bioaccumulate have a $BCF > 5000$ or $BCF > 2000$ [16]. If no *BCF* or *BAF* data are available, it may be assumed that bioaccumulative compounds are those with $\log K_{ow} > 5$ [16,17], >4.5 [18] or >3.3 [19]. Measured and evaluated bioaccumulation data are also used to assign chemicals to three bioaccumulation categories: not significantly bioaccumulative (BCF or $BAF < 1000$), bioaccumulative (BCF or BAF between 1000 and 5000) and highly bioaccumulative (BCF or $BAF > 5000$) [20].

Experimental toxicity data exist for just a fraction of relevant compounds, and in vivo measurements of such data require a lot of time and effort. According to Weisbrod et al., the collection of environmental toxicity data for 1240 potentially bioaccumulative compounds from the Canadian Domestic Substance List would take 82 years [15], and, as estimated in 2013, the average cost of experimental *BCF* determination is EUR 35,000 per compound, with more than 100 fish being sacrificed during tests lasting at least one month [21]. With the difficulties related to experimental *BCF* determination in mind, attention has turned to in vitro or in silico *BCF* models. Log *BCF* can be predicted using descriptors related to the partitioning of molecules between water and lipids, e.g., aqueous solubility [22,23]. However, in the majority of computational *BCF* models, the key descriptor governing the ability of compounds to bioconcentrate is the octanol–water partition coefficient $\log K_{ow}$ (Equations (1)–(6)) [22,24–28].

$$\log BCF = 0.542 \log K_{ow} + 0.124 \quad (n = 8, R^2 = 0.90) \quad (1)$$

$$\log BCF = 0.85 \log K_{ow} - 0.70 \quad (n = 55, R^2 = 0.897) \quad (2)$$

$$\log BCF = \log K_{ow} - 1.32 \quad (n = 63, R^2 = 0.95) \quad (3)$$

$$\log BCF = 0.94 \log K_{ow} - 1.00 \quad (4)$$

$$\log BCF = 0.516 \log K_{ow} + 0.576 \quad (n = 154, R^2 = 0.60) \quad (5)$$

$$\log BCF = 0.80 \log K_{ow} - 0.52 \quad (n = 107, R^2 = 0.81) \quad (6)$$

It was soon noticed [29,30] that the linear $\log BCF$ – $\log K_{ow}$ dependencies fail for more lipophilic compounds ($\log K_{ow} > 6$ to 7), so non-linear relationships were developed (Equations (7)–(9)) [27,28]:

$$\log BCF = -0.164 (\log K_{ow})^2 + 2.069 \log K_{ow} - 2.592 \quad (n = 154, R^2 = 0.83) \quad (7)$$

$$\log BCF = 0.910 \log K_{ow} - 1.975 (6.8 \cdots 10^{-7} K_{ow} + 1) - 0.786 \quad (n = 154, R^2 = 0.90) \quad (8)$$

$$\log BCF = 0.0069 (\log K_{ow})^4 - 0.185 (\log K_{ow})^3 + 1.55 (\log K_{ow})^2 - 4.18 \log K_{ow} + 4.79 \quad (9)$$

The often-observed hydrophobicity “cutoff”, i.e., the significantly reduced ability of lipophilic molecules to bioconcentrate (compared to what might be expected from their lipophilicity) is, however, disputed by some authors who attribute this phenomenon to experimental artifacts [31–33].

Other authors studied the influence of molecular size descriptors on the bioconcentration and bioaccumulation processes. In their opinion, molecular weight or molar volume

should be incorporated in the *BCF* models along with $\log K_{ow}$ to account for the reduced uptake of both large and highly lipophilic molecules (Equation (10) [34]):

$$\log BCF = 3.036 \log K_{ow} - 0.197 (\log K_{ow})^2 - 0.808 V_M \quad (n = 28, R^2 = 0.817) \quad (10)$$

where V_M —molar volume.

Dimitrov reported that the threshold value of 1.5 nm for the maximal cross-section diameter discriminates between compounds with $\log BCF >$ and <3.3 [35]. Further research by Dimitrov was concerned with the influence of chemicals' metabolism in fish liver on their ability to bioconcentrate (*BCF* was calculated using K_{ow} as the most important descriptor, with molecular size and ionization taken into account and a simulator for fish liver used to reproduce the fish metabolism) [36].

In search for models capable of addressing the hydrophobicity cutoff problem observed for highly lipophilic molecules, QSAR *BCF* studies were reported by several authors [18,21,37–40]. The most widely recognized models accounting for this phenomenon are:

- The model developed by Meylan [41], including different relationships depending on the compounds' properties (Table 1);
- The model developed on the basis of Meylan's work recommended by US EPA and available as EPI Suite™ BCFBAF v. 3.02 freeware [42] (Table 1);

Table 1. $\log BCF$ vs. $\log K_{ow}$ according to Meylan (a) and US EPA (b) models.

$\log K_{ow}$	Non-Ionic		$\log K_{ow}$	Ionic	
	Meylan	US EPA		Meylan ^a	US EPA ^b
below 1	0.50	0.50	below 5		0.50
1 to 7	$0.77 \log K_{ow} - 0.70 + \Sigma F_i$	$0.6598 \log K_{ow} - 0.333 + \Sigma F_i$	5 to 6		0.75
7 to 10.5	$-1.37 \log K_{ow} + 14.4 + \Sigma F_i$	$-0.49 \log K_{ow} + 7.554 + \Sigma F_i$	6 to 7 ^a or 8 ^b		1.75
			7 ^a or 8 ^b to 9		1.00
above 10.5	0.50		above 9		0.50

where ΣF_i —sum of correction factors.

- CAESAR method (Equations (11)–(13)) based on eight descriptors: *MlogP* (Moriguchi log of the octanol–water partition coefficient), *BEHp2* (highest eigenvalue $n. 2$ of Burden matrix/weighted by atomic polarizabilities), *AEige* (absolute eigenvalue sum from electronegativity weighted distance matrix), *GATS5v* (Geary autocorrelation—lag 5/weighted by atomic van der Waals volumes), *CI-089* (Cl attached to Cl(sp²)), X0sol (solvation connectivity index chi-0), *MATS5v* (Moran autocorrelation—lag 5/weighted by atomic van der Waals volumes), *SsCl* (sum of all (–Cl) E-state values in molecule) [37,43]. According to the CAESAR method, *BCF* is calculated according to two models, A and B, which differ in the selection of descriptors (with *MlogP* and *BEHp2* being common to A and B), and the *BCF* value is finally predicted as follows:

$$\log K_{ow} \leq 1.355 \quad \log BCF = 0.936 \log BCF_{mean} - 0.123 \quad (11)$$

$$1.355 \leq \log K_{ow} \leq 2.410 \quad \log BCF = 0.996 \min(\log BCF_A, \log BCF_B) \quad (12)$$

$$\log K_{ow} > 2.410 \quad \log BCF = 1.052 \log BCF_{mean} \quad (13)$$

- The model suggested by the Technical Guidance Document (TGD) on risk assessment [18] (Equations (14)–(17)):

$$\log K_{ow} < 1 \quad \log BCF = 0.15 \quad (14)$$

$$1 \leq \log K_{ow} \leq 6 \quad \log BCF = 0.85 \log K_{ow} - 0.70 \quad (15)$$

$$6 < \log K_{ow} < 10 \quad \log BCF = -0.20 (\log K_{ow})^2 + 2.74 \log K_{ow} - 4.72 \quad (16)$$

$$\log K_{ow} \geq 10 \quad \log BCF = 2.68 \quad (17)$$

Alternative approaches to *BCF* predictions involve molecular connectivity indices [44–46] or solvation parameters [47]. *BCF* may also be estimated on the basis of quantum chemical descriptors [48] (Equation (18)):

$$BCF = 0.00250 M_w - 0.0724 E_T - 0.214 E_{HOMO} - 0.892 E_{LUMO} - 2.58 \quad (18)$$

where E_T —total energy (hartree), E_{HOMO} —energy of the highest occupied molecular orbital (eV) and E_{LUMO} —energy of the lowest unoccupied molecular orbital (eV).

The *BCF* can be estimated using reversed-phase chromatographic retention data. In particular, retention parameters derived from HPLC chromatography on C_{18} , C_8 , C_2 and phenyl-bonded silica sorbents were used as predictors of *BCF* of aromatic hydrocarbons [49]. C_{18} and cyanopropyl- and phenyl-bonded silica were used in chromatographic bioconcentration studies of aromatic hydrocarbons, alkylbenzenes, chlorinated benzenes, phthalates, nitroaromatics, phenols and aniline [50]. RP-18 TLC chromatographic descriptors were used to investigate the bioconcentration factors of organic sunscreens and cosmetic preservatives [51]. IAM chromatographic descriptors were used to study the *BCF* of structurally unrelated chemicals [4].

The objective of this study was to develop useful and easy-to-use predictive models of the bioconcentration factor of structurally diverse solutes based on their affinity for phosphatidylcholine-based artificial membranes. Novel models proposed in this study were generated using multiple linear regression (MLR), partial least squares (PLS) and artificial neural network (ANN) techniques. It is the first report on PLS and ANN approaches to bioconcentration studies involving chromatographic and calculated physico-chemical data.

2. Materials and Methods

2.1. Compounds, IAM Chromatographic Data, Reference *BCF* Values

The first stage of this study was intended to involve 175 compounds, whose IAM chromatographic retention factors obtained for purely aqueous mobile phases ($\log k_w^{IAM}$) were compiled by Sprunger et al. [52]. Because of the lack of experimental *BCF* data for the whole group of 175 compounds, $\log BCF$ (denoted later as $\log BCF_{EPI}$) was calculated using the commonly accepted computational approach (EPI SuiteTM, BCFBAF module v. 3.02) [42] based on Meylan's model [41]. A large number of compounds considered at this stage of the study, however, were molecules with arbitrarily assigned $\log BCF = 0.50$. The majority of such compounds were excluded from the training set because it was suspected that their theoretical $\log BCF$ value may not truly reflect their ability to bioconcentrate [4], and the models generated in this study were finally based upon a solute set containing 120 compounds from different chemical families (**1** to **120**). The excluded compounds were later combined with solutes, whose $\log k_w^{IAM}$ values were reported by other authors [53,54], to form an external test set also containing 120 compounds (**121** to **240**) with and without known experimental values of $\log BCF$ ($\log BCF_{vivo}$). Reliable reference $\log BCF$ values ($\log BCF_{EPI}$) were available for compounds **1** to **187**, and, for the compounds **188** to **240**, $\log BCF$ was calculated de novo. The external set of compounds included more lipophilic molecules, whose $\log k_w^{IAM}$ could not be measured directly by chromatography with 100% aqueous mobile phase and could only be calculated by extrapolation of $\log k^{IAM}$ vs. the φ plots obtained for a series of chromatographic experiments with mobile phases containing different concentrations φ of a water-miscible organic solvent, usually according to the linear Soczewiński–Wachmeister Equation (19) [55]:

$$\log k = \log k_w + S\varphi \quad (19)$$

The values of $\log BCF_{vivo}$ were taken from the literature sources and the EPISuite™ database [4,29,36,42]. The reference $\log BCF_{EPI}$ and the experimental $\log BCF_{vivo}$ values for compounds **1** to **240** (where available) are given in Table S1 (Supplementary Materials); the IAM chromatographic retention factors are given in Table S2 (Supplementary Materials).

2.2. Calculated Descriptors

Molecular weight (M_w), heavy atom count ($\#HvAt$), aromatic heavy atom count ($\#ArHvAt$), fraction of sp^3 carbons (F_{Csp3}), rotatable bond count (FRB), hydrogen donor count (HD), hydrogen acceptor count (HA), molecular refractivity (MR) and topological polar surface area ($TPSA$) were calculated using SwissADME software available freely on-line [56]. Total energy (E_t), energy of the highest occupied molecular orbital (E_{HOMO}), energy of the lowest unoccupied molecular orbital (E_{LUMO}), dipole point charge ($DiPCh$), dipole hybrid ($DipH$) and dipole sum ($DipS$) were of Mopac 2016 type and were calculated using the OCHEM platform [57]. Octanol–water partition coefficient ($\log K_{ow}$) was calculated according to the KOWWIN algorithm [58] using EPI Suite™ software [42] (Table S2).

2.3. Partial Least Squares Approach

Multiple linear regression (MLR) is a common approach used in QSAR studies. It is based on the assumption that the effect of a set of molecule's properties on its activity is additive, and the properties are (almost) independent. The conditions that must be satisfied to generate reliable MLR models are severe—standard regression techniques based on the least squares estimation give unstable and unreliable results when independent variables are colinear, and the number of cases must exceed the number of variables (ideally, it should be at least five times greater). In order to overcome the colinearity problem, partial least square (PLS) regression was developed. PLS replaces the original variables with “components”—linear combinations of the variables based on the correlation between the dependent variable and the independent variable(s) [59,60].

Regression models based on PLS estimation must be optimized in terms of the number of components—if too many are used, a model is over-fitted (it perfectly fits the training dataset, but it gives poor prediction results for new cases); if too few components are used, the model is under-fitted (it is not sufficiently large to capture the important data variability). Models based on the same number of components can be compared using RSS (residual sum of squares) or R^2 , but these parameters are unsuitable for models with different numbers of components. PLS models are often evaluated using RMSEP calculated for a separate test set and/or using cross-validation—RMSEP usually decreases as more variables are added to a small model, then it stabilizes around the optimum number of components, and it increases when the model becomes over-fitted [61].

2.4. Statistical Tools

Multiple linear regression (MLR) models were generated using Statistica v. 13 by StatSoft Polska, Kraków, Poland, stepwise forward regression mode. Partial least squares (PLS) models were generated using Statistica v. 13, NIPALS algorithm with auto-scaling. Multilayer Perceptron (MLP) artificial neural networks (ANNs), with the number of inputs the same as the number of variables, the varying number of hidden units and one output unit, were generated using Statistica v. 13 (regression mode, Automated Network Search—ANS module, 1000 networks to train, 50 networks to retain). The neuron activation functions were selected from the following group: identity, logistic, hyperbolic tangent and exponential. The BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm was used to train the network together with the sum of squares (SOS) error function.

The models considered in this study were evaluated using the following procedures and statistical parameters:

- K-fold cross-validation, with n compounds from the initial training set split into k even subsets, $(k - 1)$ of which were used to train a new model and the remaining one to test it; the procedure was repeated k times, each time using a different subset

of compounds as a test set. After each cross-validation step, the RMSE (root mean squared error) was calculated for the particular N-compound test subset according to the following Equation (20):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i^{\text{pred}} - y_i^{\text{ref}})^2}{N}} \quad (20)$$

The overall root mean squared error of k-fold cross-validation (RMSECV) is calculated as follows:

$$\text{RMSECV} = \sqrt{\frac{\sum_{j=1}^k \text{RMSE}_j^2}{k}} \quad (21)$$

In this study, $n = 120$, $k = 5$ and $N = 24$; y_i^{pred} and y_i^{ref} are $\log BCF_{\text{pred}}$ and $\log BCF_{\text{EPI}}$, respectively.

- Relationship between the predicted $\log BCF_{\text{pred}}$ values (computed for the external test set of 67 compounds 121 to 187 that were not used to build models) with the reference values $\log BCF_{\text{EPI}}$ —using root mean squared error of prediction ($\text{RMSEP}_{\text{ext}}$), calculated according to Equation (20);
- Comparison of the predicted $\log BCF_{\text{pred}}$ values (calculated for 40 compounds, whose experimental $\log BCF_{\text{vivo}}$ data are available), and these data—using squared coefficient of determination (R^2_{vivo}) and root-mean-squared error of prediction ($\text{RMSEP}_{\text{vivo}}$), calculated according to Equation (20).

3. Results and Discussion

3.1. Multiple Linear Regression (MLR) Models

The values of $\log BCF_{\text{EPI}}$, calculated for compounds 1 to 120 using EPISuite™ software [42], were plotted against the IAM retention factors obtained for aqueous mobile phases $\log k_w^{\text{IAM}}$ and compiled by Sprunger [52]. The linear relationship between $\log BCF_{\text{EPI}}$ and $\log k_w^{\text{IAM}}$ (Equation (22), MLR1, Figure 1) accounted for 80% of total $\log BCF_{\text{EPI}}$ variability.

$$\log BCF_{\text{EPI}} = 0.18 (\pm 0.06) + 0.70 (\pm 0.03) \log k_w^{\text{IAM}} \quad (22)$$

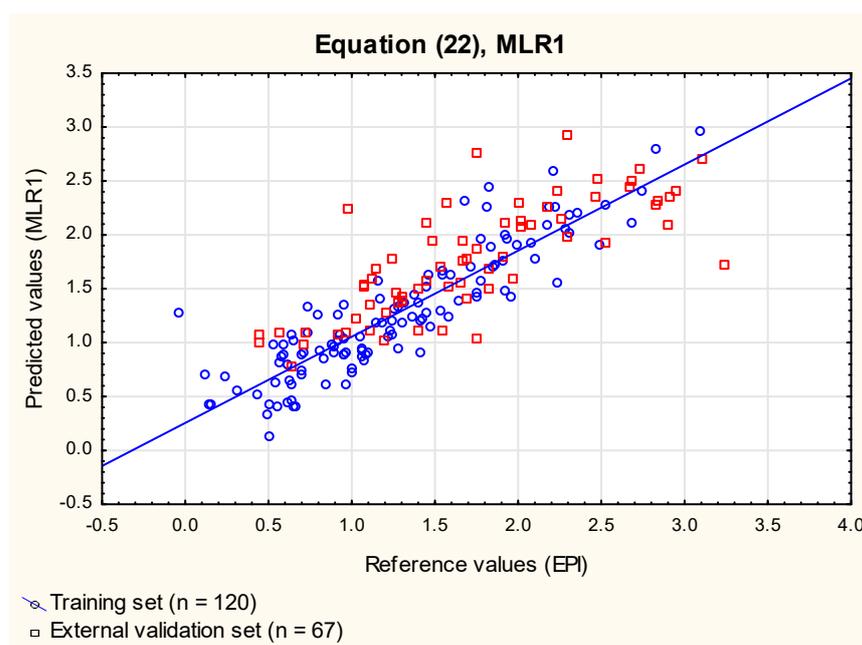


Figure 1. MLR1 model, Equation (22)—predicted vs. reference $\log BCF$ values.

($n = 120$, $R^2 = 0.80$, $R^2_{\text{adj}} = 0.80$, $\text{RMSECV} = 0.30$, $\text{RMSEP}_{\text{ext}} = 0.45$, $\text{RMSEP}_{\text{vivo}} = 0.35$, $R^2_{\text{vivo}} = 0.74$, $F = 470.5$, $p < 0.01$).

The results of $\log BCF_{\text{EPI}}$ modeling using a single chromatographic descriptor ($\log k_w^{\text{IAM}}$) obtained in this study (Equation (22), **MLR1**) are similar to those reported by Tsopelas [4] ($R^2 = 0.74$, $n = 77$). $\log k_w^{\text{IAM}}$ accounted for ca. 74% of variability of $\log BCF_{\text{vivo}}$ data (which proves the importance of the chromatographic parameter), but it was hoped that the model can be improved by incorporating some additional independent variables expected to influence the ability of compounds to be absorbed by aquatic animals from the surrounding water via the respiratory route and skin. It is likely that, similarly to pharmacokinetic processes of compound absorption and distribution in humans, the key features responsible for the ability of molecules to bioconcentrate in aquatic organisms are their lipophilicity (which, indeed, is the main parameter in the majority of BCF in silico models), ability to form hydrogen bonds and molecule flexibility and size. Apart from $\log k_w^{\text{IAM}}$, which is strongly related to solutes' lipophilicity, several molecular descriptors calculated using SwissADME software were investigated. The improved Equation (23) (**MLR2**, Figure 2) was generated using forward stepwise regression:

$$\log BCF_{\text{EPI}} = 0.27 (\pm 0.06) + 0.71 (\pm 0.03) \log k_w^{\text{IAM}} - 0.0043 (\pm 0.0009) \text{TPSA} + 0.24 (\pm 0.06) F_{\text{Csp3}} - 0.089 (\pm 0.036) \text{HD} \quad (23)$$

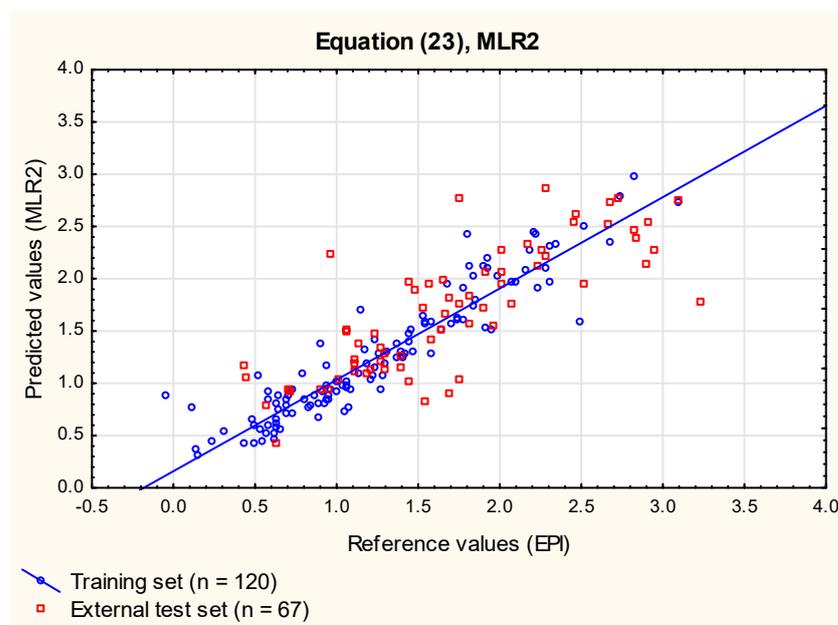


Figure 2. MLR2 model, Equation (23)—predicted vs. reference $\log BCF$ values.

($n = 120$, $R^2 = 0.87$, $R^2_{\text{adj}} = 0.87$, $\text{RMSECV} = 0.25$, $\text{RMSEP}_{\text{ext}} = 0.42$, $\text{RMSEP}_{\text{vivo}} = 0.27$, $R^2_{\text{vivo}} = 0.83$, $F = 198.4$, $p < 0.01$).

The additional independent variables incorporated into Equation (23) (**MLR2**) were statistically significant and accounted for ca. 7% of total variability. They were introduced in the following order: TPSA , F_{Csp3} and HD , which confirmed the relationship between TPSA and the phenomenon of bioconcentration reported earlier by Tsopelas [4] (who also demonstrated the contribution of a biodegradation estimate, BioWin5 , calculated using EPISuite™ software). Polar surface area is an important parameter that defines the polar part of a molecule. It is strongly related to the passive transport of molecules through membranes, and it is known to influence the ADME processes in humans (e.g., the blood and brain barrier permeability, transdermal or intestinal absorption [62–64]). Other BCF predictors incorporated in Equation (23) (**MLR2**, Figure 2) are the fraction of sp^3 carbons F_{Csp3} (which, in simple terms, can be considered a measure of molecule's flexibility and is positively correlated with $\log BCF$) and the count of H-bond donors HD .

The coefficients for both *HD* and *PSA* in Equation (MLR2) are negative—high polar surface area and the molecule's strong tendency to form hydrogen bonds reduce its uptake by aquatic organisms.

Further attempts to improve the MLR models by incorporating other parameters expected to influence the compounds' ability to bioconcentrate were not very successful—Equation (24) (MLR3, Figure 3), obtained using six variables selected by forward stepwise regression, had slightly better parameters of cross-validation than the model MLR2 (Equation (23)), but this gain did not justify the risk of over-fitting related to incorporation of two more parameters (*FRB* and *DipH*) that, although both statistically significant, accounted together for only slightly over 1% of total variability. The ability of Equation (24) to predict log *BCF* for new cases (the external test set) and the relationship between log *BCF* values predicted using this model and the experimental values were comparable to those reported for Equation (23) (MLR2).

$$\log BCF_{EPI} = 0.14 (\pm 0.06) + 0.74 (\pm 0.03) \log k_w^{IAM} - 0.0037 (\pm 0.0011) TPSA + 0.35 (\pm 0.07) F_{Csp3} - 0.16 (\pm 0.04) HD - 0.026 (\pm 0.011) FRB + 0.29 (\pm 0.08) DipH \quad (24)$$

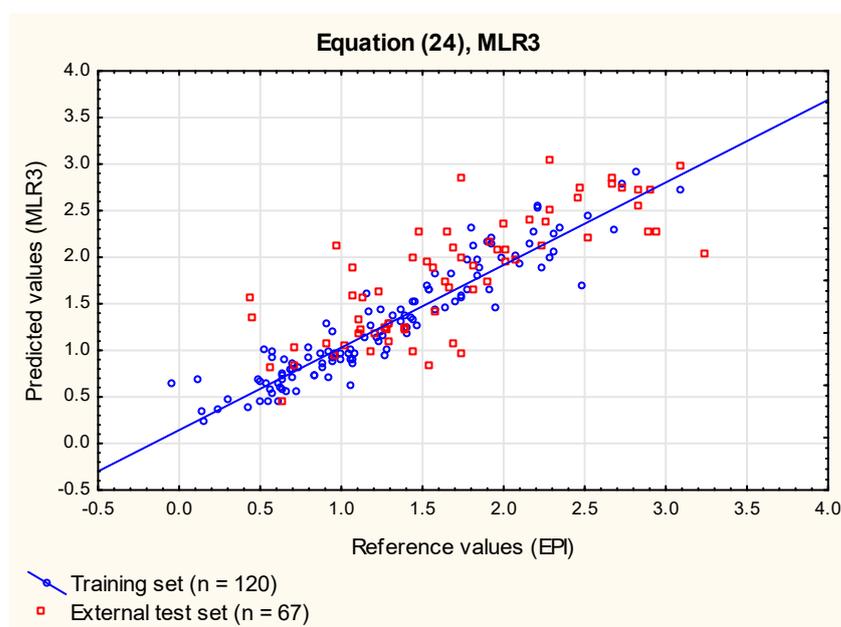


Figure 3. MLR3 model, Equation (24)—predicted vs. reference log *BCF* values.

($n = 120$, $R^2 = 0.89$, $R^2_{adj} = 0.88$, $RMSECV = 0.17$, $RMSEP_{ext} = 0.45$, $R^2_{vivo} = 0.83$, $RMSEP_{vivo} = 0.27$, $F = 147.3$, $p < 0.01$).

3.2. Partial Least Square (PLS) Models

In this study, the following PLS models were investigated (details to be found in Supplementary Materials):

- Models **PLS1** based on 16 independent variables—including those involved in MLR analysis and some other descriptors that were not included in MLR to avoid collinearity problems;
- Model **PLS2** based on a reduced set of independent variables.

PLS1 models based on the set of 16 independent variables and involving between 4 and 12 components were compared using $RMSEP_{ext}$, $RMSEP_{vivo}$ and $RMSECV$ values (Supplementary Materials). At a later step, multiple linear forward stepwise regression was also performed on the X-scores of all the possible 16 PLS components. Using these two approaches, it was established that the optimum number of components is six (Figure 4)—it led to a model that fitted the training dataset reasonably well, the model's predictive

potential was satisfying (i.e., the model was neither over-fitted or under-fitted) and all six PLS components selected by MLR were statistically significant.

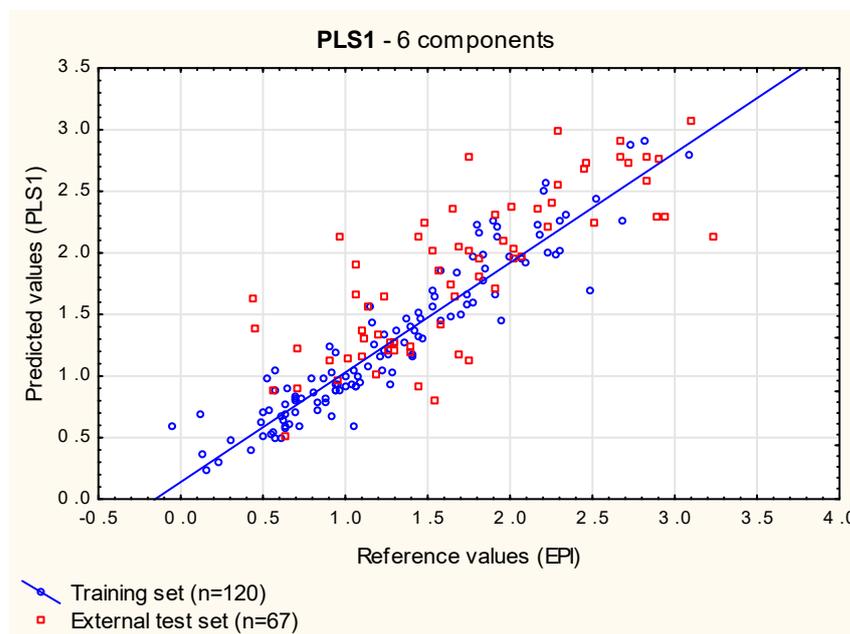


Figure 4. PLS1 model (six components)—predicted vs. reference log *BCF* values.

The importance of descriptors used in PLS models can be evaluated manually based on their variable importance in the projection (VIP) values calculated for the particular number of components (descriptors with $VIP < 1$ in a PLS model are excluded from the next one) [65] (Table 2). This procedure was applied to PLS1, and it was established that only two variables, $\log k_w^{IAM}$ and *MR* (a descriptor connected with polarizability of molecules, not selected in MLR) had a strong influence on log *BCF* (model PLS2, Figure 5). Surprisingly, the descriptors selected by stepwise multiple regression (apart from $\log k_w^{IAM}$, which is of utmost importance in all the models developed in this study) were of lesser importance in the PLS regression. Model PLS2, however, seemed excessively simplified, and its performance, evaluated using $RMSE_{CV}$, $RMSE_{ext}$ and $RMSE_{vivo}$, was slightly worse than that of PLS1 (Table 3).

Table 2. VIP values for independent variables, model PLS1.

Variable	VIP	Importance
$\log k_w^{IAM}$	2.53	1
<i>MR</i>	1.08	2
<i>#HvAt</i>	0.97	3
M_w	0.97	4
<i>HD</i>	0.92	5
<i>DipPCh</i>	0.88	6
E_t	0.84	7
<i>FRB</i>	0.84	8
<i>DipS</i>	0.83	9
<i>TPSA</i>	0.76	10
<i>#ArHvAt</i>	0.72	11
<i>DipH</i>	0.71	12
E_{HOMO}	0.69	13
<i>HA</i>	0.64	14
E_{LUMO}	0.48	15
F_{Csp3}	0.32	16

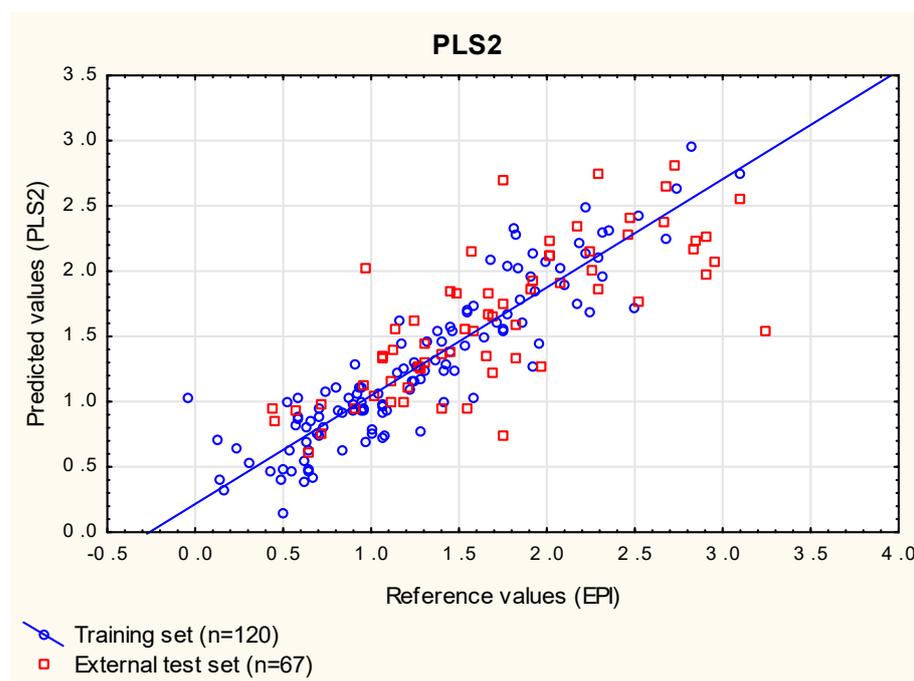


Figure 5. PLS2 model—predicted vs. reference log BCF values.

Table 3. Summary of MLR, PLS and ANN models developed in this study.

	MLR1	MLR2	MLR3	PLS1	PLS2	ANN14	ANN43	ANN44
RMSECV	0.30	0.25	0.17	0.26	0.29	-	-	-
RMSEP _{ext}	0.35	0.42	0.45	0.45	0.46	0.47	0.47	0.47
RMSEP _{vivo}	0.35	0.27	0.27	0.27	0.31	0.28	0.28	0.30
R ² _{vivo}	0.74	0.83	0.83	0.83	0.77	0.81	0.82	0.79

3.3. Artificial Neural Networks

Artificial neural networks are widely used to predict drugs' bioavailability [66] or properties such as affinity for phospholipids using IAM chromatography and calculated descriptors [67]. The great advantages of neural networks compared to MLR are the possibility of utilizing both linear and non-linear relationships between input data and a predicted parameter and the ability of ANNs to learn these relationships directly from the data being modeled.

In this study, the ANN models were built for the same group of compounds (**1** to **120**) that was used as the training set in the MLR and PLS analyses. This group of compounds was randomly assigned to three subgroups: train (70%), test (15%) and validation (15%)—the latter two groups were needed to optimize the ANNs as they were being created. Similarly to the MLR and PLS analyses presented in this study, the compounds **121** to **240** were used as an additional, external test set. At this point, 1000 networks were generated, and 50 with the smallest error were retained for further examination in search of those that give the results in the closest agreement with the reference data (log BCF_{EPI}) for compounds **121** to **187** (RMSEP_{ext}) and with the experimental data (log BCF_{vivo}) for a subgroup of 40 cases, whose experimental log BCF values were available (R²_{vivo}, RMSEP_{vivo}). The selection of the best exemplary networks generated in this study (ANN14, ANN43 and ANN44, Figures 6–8) was based on their ability to predict new cases (RMSEP_{ext}) and to obtain the results in the closest possible agreement with the experimental data (R²_{vivo}, RMSEP_{vivo}) rather than on their ability to fit the training data (Supplementary Materials).

ANNs make it possible to process a large number of descriptors that can be easily obtained using readily available software. The selection of ANN input data is an important

step because, if the number of parameters is excessive considering the number of cases, models are over-fitted. The importance of independent variables can be evaluated using a tool known as global sensitivity analysis (GSA), which rates the importance of the models' input variable by computing sums of squared residuals for the model when the respective predictor is eliminated compared to the full model. When an input variable scores 1 or less than 1 in GSA, it means that this particular network is likely to perform better without this variable; however, in the networks generated in this study, the majority of GSA scores were at least slightly above this threshold.

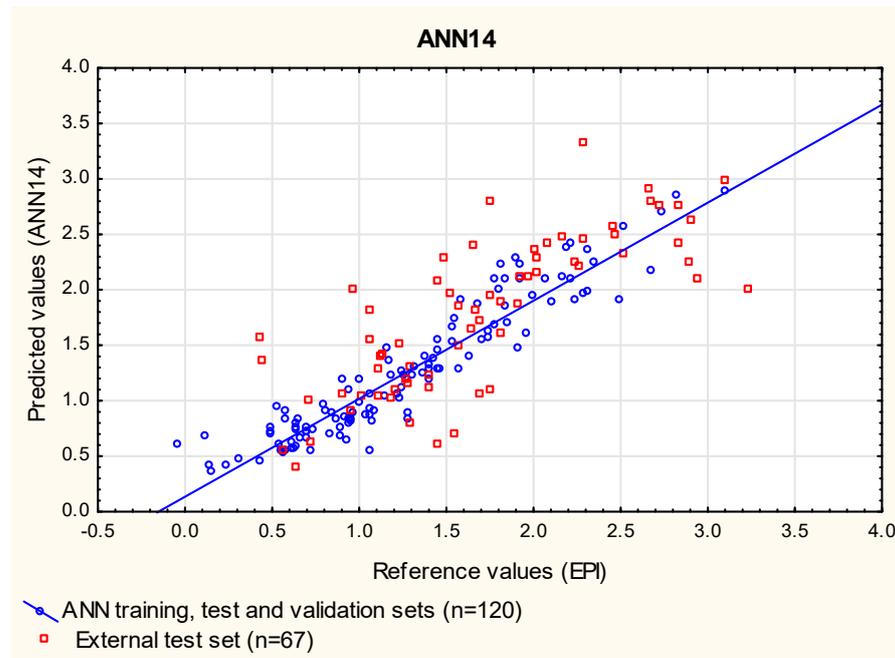


Figure 6. ANN14 model—predicted vs. reference log *BCF* values.

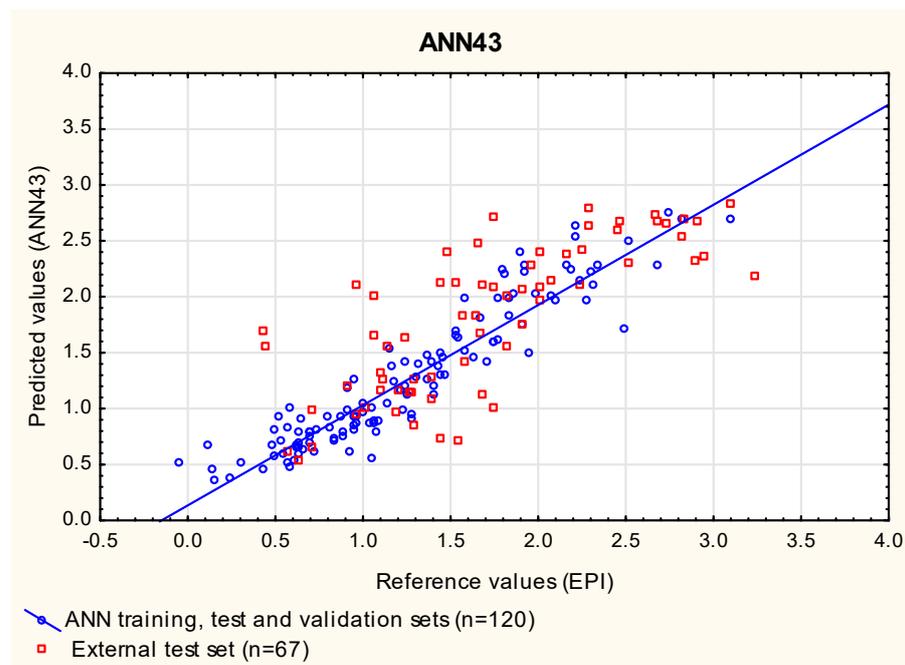


Figure 7. ANN43 model—predicted vs. reference log *BCF* values.

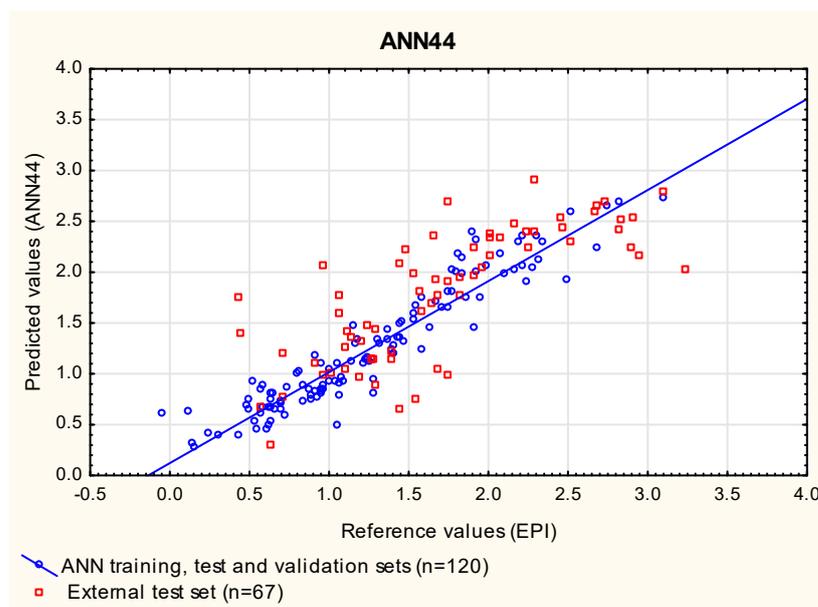


Figure 8. ANN44 model—predicted vs. reference log BCF values.

Log k_w^{IAM} is an important predictor accounting for 80% of log BCF variability. It encodes the molecule's properties responsible for its ability to cross biological membranes—lipophilicity and size (molecular weight, heavy atom count), Table 4—and, when additional descriptors are incorporated, it leads to efficient BCF models. In this study, the models were generated using log k_w^{IAM} values obtained directly for aqueous mobile phases. Using the external test group of solutes, it was demonstrated, however, that log k_w^{IAM} values obtained by extrapolation of log k^{IAM} values to zero concentration of organic modifiers in the mobile phase were sufficient to give reasonable predictions—although, since log k_w^{IAM} is the most important descriptor in all the models, imperfections of this variable in the external test dataset always had some influence on the $RMSEP_{ext}$ values.

Models MLR2, PLS1 and ANN43 were finally compared (Figure 9) by plotting the predicted log BCF values against the experimental ones (log BCF_{vivo}), and it was confirmed that their ability to model the experimental log BCF data was similar.

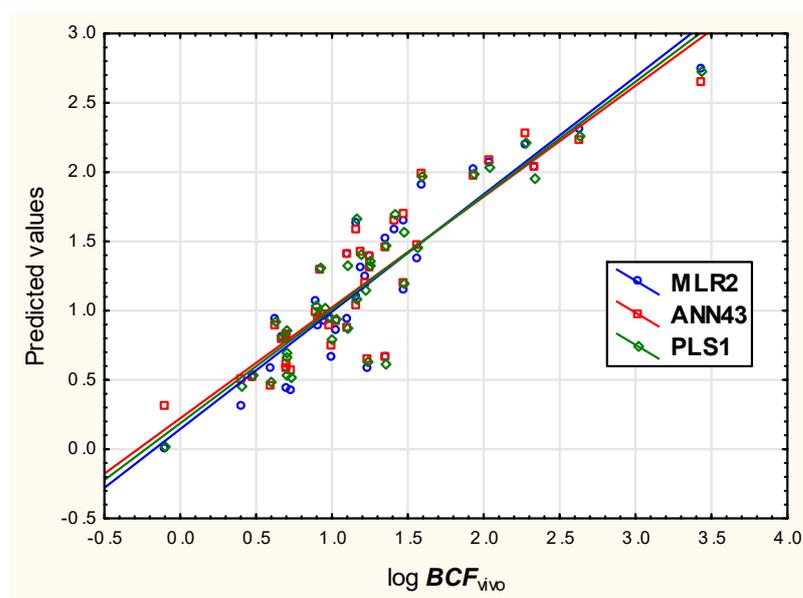


Figure 9. Predicted vs. experimental log BCF values for models MLR2, PLS1 and ANN43.

Table 4. Correlations between descriptors analyzed in this study.

	$\log k_w^{IAM}$	M_W	#HAt	#ArHAt	F_{Csp3}	FRB	HA	HD	MR	TPSA	E_t	E_{HOMO}	E_{LUMO}	DipPCh	DipH	DipS	$\log K_{ow}$
$\log k_w^{IAM}$	1.00	0.51	0.51	0.39	−0.07	0.34	0.09	−0.15	0.57	−0.08	−0.44	0.38	−0.17	−0.09	0.05	−0.04	0.84
M_W	0.51	1.00	0.98	0.50	0.07	0.62	0.78	0.38	0.98	0.64	−0.86	0.49	−0.42	0.45	0.52	0.49	0.32
#HAt	0.51	0.98	1.00	0.52	0.06	0.63	0.78	0.37	0.99	0.64	−0.87	0.53	−0.39	0.46	0.52	0.49	0.32
#ArHAt	0.39	0.50	0.52	1.00	−0.58	0.16	0.27	0.02	0.53	0.17	−0.44	0.56	−0.52	0.13	0.40	0.16	0.29
F_{Csp3}	−0.07	0.07	0.06	−0.58	1.00	0.27	0.09	0.06	0.08	0.00	−0.01	−0.29	0.59	−0.01	−0.07	−0.01	−0.04
FRB	0.34	0.62	0.63	0.16	0.27	1.00	0.59	0.31	0.63	0.45	−0.52	0.31	−0.06	0.26	0.33	0.31	0.25
HA	0.09	0.78	0.78	0.27	0.09	0.59	1.00	0.53	0.70	0.86	−0.75	0.25	−0.39	0.61	0.54	0.61	−0.11
HD	−0.15	0.38	0.37	0.02	0.06	0.31	0.53	1.00	0.33	0.67	−0.35	0.26	−0.14	0.26	0.52	0.28	−0.27
MR	0.57	0.98	0.99	0.53	0.08	0.63	0.70	0.33	1.00	0.57	−0.84	0.56	−0.36	0.41	0.51	0.44	0.38
TPSA	−0.08	0.64	0.64	0.17	0.00	0.45	0.86	0.67	0.57	1.00	−0.63	0.21	−0.43	0.67	0.55	0.68	−0.28
E_t	−0.44	−0.86	−0.87	−0.44	−0.01	−0.52	−0.75	−0.35	−0.84	−0.63	1.00	−0.43	0.42	−0.48	−0.42	−0.49	−0.25
E_{HOMO}	0.38	0.49	0.53	0.56	−0.29	0.31	0.25	0.26	0.56	0.21	−0.43	1.00	−0.23	0.13	0.44	0.19	0.26
E_{LUMO}	−0.17	−0.42	−0.39	−0.52	0.59	−0.06	−0.39	−0.14	−0.36	−0.43	0.42	−0.23	1.00	−0.39	−0.27	−0.38	−0.05
DipPCh	−0.09	0.45	0.46	0.13	−0.01	0.26	0.61	0.26	0.41	0.67	−0.48	0.13	−0.39	1.00	0.33	0.97	−0.28
DipH	0.05	0.52	0.52	0.40	−0.07	0.33	0.54	0.52	0.51	0.55	−0.42	0.44	−0.27	0.33	1.00	0.44	−0.11
DipS	−0.04	0.49	0.49	0.16	−0.01	0.31	0.61	0.28	0.44	0.68	−0.49	0.19	−0.38	0.97	0.44	1.00	−0.25
$\log K_{ow}$	0.84	0.32	0.32	0.29	−0.04	0.25	−0.11	−0.27	0.38	−0.28	−0.25	0.26	−0.05	−0.28	−0.11	−0.25	1.00

4. Conclusions

The ability of compounds to bioconcentrate in aquatic organisms is strongly related to their affinity for phosphatidylcholine-based immobilized artificial membranes (IAM), and other physico-chemical parameters of a molecule are less important in this process. QSAR models of $\log BCF$ involving the IAM chromatographic retention factor and other descriptors were built using multiple linear regression, partial least square regression and artificial neural networks. The MLR approach is a powerful technique with the great advantage of simplicity—models generated using this technique usually involve a relatively small number of independent variables (parameters), whose physical meaning and contribution towards an dependent variable can be easily understood. In this study, the selected MLR, PLS and ANN models gave fairly comparable results in terms of their ability to predict new cases ($\log BCF_{ext}$), and the results obtained using these models were in similar agreement with experimental data ($\log BCF_{vivo}$) (surprisingly, simple MLR equations based on a relatively small number of independent variables seemed to perform slightly better than more complex ANN or PLS models). Generally speaking, PLS regression deals with the collinearity of independent variables, and the ANN approach is especially useful in the case of non-linear relationships, but, in this study, linear equations (especially Equation (23), MLR2) gave satisfying prediction results, and they were more intuitive. All the models reported above can be easily applied during the early steps of the drug discovery process concurrently with IAM chromatographic pharmacokinetic studies and, as described earlier, in the studies of compounds' mobility in the soil–water compartment [6]. In lieu of $\log k_w^{IAM}$ obtained directly using aqueous mobile phases, extrapolated values can be used, although, in such situations, the quality of BCF predictions is slightly impaired. The models proposed in this study are applicable to compounds over a relatively wide range of lipophilicity, with the exception of very lipophilic molecules ($\log K_{ow} > ca. 7$), whose retention times on the IAM chromatographic support are very long and $\log k_w^{IAM}$ cannot be conveniently measured. This limitation of the applicability domain of the models presented in this study, however, is not a major drawback—very lipophilic compounds, as demonstrated by some authors, do not bioconcentrate or bioaccumulate easily [16,27,28,34], which is either a direct result of their hydrophobicity or, indirectly, an effect of the larger molecular size of highly lipophilic molecules [32]. Above a certain lipophilicity threshold ($\log K_{ow} > ca. 7$), the bioconcentration factor becomes inversely proportional to lipophilicity and decreases rapidly. On the other hand, a large proportion of compounds released to the environment by agriculture or the pharmaceutical industry (e.g., pesticides or drugs) meets the criteria of optimum intestinal, transdermal or lung absorption [19,68–72]. Such compounds are usually moderately lipophilic ($\log K_{ow}$ rarely higher than 7, in the majority of cases, between 0 and 5), so quantitative studies of their bioconcentration using the models discussed above are feasible.

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/membranes12111130/s1> as an Excel file: Supplementary Materials.xlsx.

Funding: This research was supported by an internal grant from the Medical University of Lodz, Poland no. 503/3-016-03/503-31-001.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this manuscript.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Pidgeon, C.; Venkataram, U.V. Immobilized Artificial Membrane Chromatography: Supports Composed of Membrane Lipids. *Anal. Biochem.* **1989**, *176*, 36–47. [CrossRef]
2. Sobanska, A.W.; Brzezinska, E. Phospholipid-Based Immobilized Artificial Membrane (IAM) Chromatography: A Powerful Tool to Model Drug Distribution Processes. *Curr. Pharm. Des.* **2017**, *23*, 6784–6794. [CrossRef]

3. Tsopelas, F.; Stergiopoulos, C.; Tsantili-Kakoulidou, A. Immobilized Artificial Membrane Chromatography: From Medicinal Chemistry to Environmental Sciences. *ADMET DMPK* **2018**, *6*, 225–241. [[CrossRef](#)]
4. Tsopelas, F.; Stergiopoulos, C.; Tsakanika, L.-A.; Ochsenkühn-Petropoulou, M.; Tsantili-Kakoulidou, A. The Use of Immobilized Artificial Membrane Chromatography to Predict Bioconcentration of Pharmaceutical Compounds. *Ecotoxicol. Environ. Saf.* **2017**, *139*, 150–157. [[CrossRef](#)] [[PubMed](#)]
5. Stergiopoulos, C.; Makarouni, D.; Tsantili-Kakoulidou, A.; Ochsenkühn-Petropoulou, M.; Tsopelas, F. Immobilized Artificial Membrane Chromatography as a Tool for the Prediction of Ecotoxicity of Pesticides. *Chemosphere* **2019**, *224*, 128–139. [[CrossRef](#)]
6. Sobańska, A.W. Immobilized Artificial Membrane Chromatographic and Computational Descriptors in Studies of Soil-Water Partition of Environmentally Relevant Compounds. *Environ. Sci. Pollut. Res.* **2022**. [[CrossRef](#)]
7. Fent, K.; Zenker, A.; Rapp, M. Widespread Occurrence of Estrogenic UV-Filters in Aquatic Ecosystems in Switzerland. *Environ. Pollut.* **2010**, *158*, 1817–1824. [[CrossRef](#)]
8. Gago-Ferrero, P.; Díaz-Cruz, M.S.; Barceló, D. UV Filters Bioaccumulation in Fish from Iberian River Basins. *Sci. Total Environ.* **2015**, *518–519*, 518–525. [[CrossRef](#)]
9. Saunders, L.J.; Hoffman, A.D.; Nichols, J.W.; Gobas, F.A.P.C. Dietary Bioaccumulation and Biotransformation of Hydrophobic Organic Sunscreen Agents in Rainbow Trout. *Environ. Toxicol. Chem.* **2020**, *39*, 574–586. [[CrossRef](#)]
10. Schneider, S.L.; Lim, H.W. Review of Environmental Effects of Oxybenzone and Other Sunscreen Active Ingredients. *J. Am. Acad. Dermatol.* **2019**, *80*, 266–271. [[CrossRef](#)]
11. Vidal-Liñán, L.; Villaverde-de-Sáa, E.; Rodil, R.; Quintana, J.B.; Beiras, R. Bioaccumulation of UV Filters in *Mytilus Galloprovincialis* Mussel. *Chemosphere* **2018**, *190*, 267–271. [[CrossRef](#)] [[PubMed](#)]
12. Burkhard, L.P. Evaluation of Published Bioconcentration Factor (BCF) and Bioaccumulation Factor (BAF) Data for Per- and Polyfluoroalkyl Substances Across Aquatic Species. *Environ. Toxicol. Chem.* **2021**, *40*, 1530–1543. [[CrossRef](#)] [[PubMed](#)]
13. Li, H.; Ran, Y. Distribution and Bioconcentration of Polycyclic Aromatic Hydrocarbons in Surface Water and Fishes. *Sci. World J.* **2012**, *2012*, 632910. [[CrossRef](#)] [[PubMed](#)]
14. Carvalho, I.T.; Santos, L. Antibiotics in the Aquatic Environments: A Review of the European Scenario. *Environ. Int.* **2016**, *94*, 736–757. [[CrossRef](#)]
15. Weisbrod, A.V.; Burkhard, L.P.; Arnot, J.; Mekenyan, O.; Howard, P.H.; Russom, C.; Boethling, R.; Sakuratani, Y.; Traas, T.; Bridges, T.; et al. Work group Report: Review of Fish Bioaccumulation Databases Used to Identify Persistent, Bioaccumulative, Toxic Substances. *Environ. Health Perspect.* **2007**, *115*, 255–261. [[CrossRef](#)]
16. Arnot, J.A.; Gobas, F.A. A Review of Bioconcentration Factor (BCF) and Bioaccumulation Factor (BAF) Assessments for Organic Chemicals in Aquatic Organisms. *Environ. Rev.* **2006**, *14*, 257–297. [[CrossRef](#)]
17. Arnot, J.A.; Gobas, F.A.P.C. A Generic QSAR for Assessing the Bioaccumulation Potential of Organic Chemicals in Aquatic Food Webs. *QSAR Comb. Sci.* **2003**, *22*, 337–345. [[CrossRef](#)]
18. Grisoni, F.; Consonni, V.; Villa, S.; Vighi, M.; Todeschini, R. QSAR Models for Bioconcentration: Is the Increase in the Complexity Justified by More Accurate Predictions? *Chemosphere* **2015**, *127*, 171–179. [[CrossRef](#)]
19. Chmiel, T.; Mieszkowska, A.; Kempnińska-Kupczyk, D.; Kot-Wasik, A.; Namieśnik, J.; Mazerska, Z. The Impact of Lipophilicity on Environmental Processes, Drug Delivery and Bioavailability of Food Components. *Microchem. J.* **2019**, *146*, 393–406. [[CrossRef](#)]
20. Costanza, J.; Lynch, D.G.; Boethling, R.S.; Arnot, J.A. Use of the Bioaccumulation Factor to Screen Chemicals for Bioaccumulation Potential. *Environ. Toxicol. Chem.* **2012**, *31*, 2261–2268. [[CrossRef](#)]
21. Gissi, A.; Nicolotti, O.; Carotti, A.; Gadaleta, D.; Lombardo, A.; Benfenati, E. Integration of QSAR Models for Bioconcentration Suitable for REACH. *Sci. Total Environ.* **2013**, *456–457*, 325–332. [[CrossRef](#)] [[PubMed](#)]
22. Isnard, P.; Lambert, S. Estimating Bioconcentration Factors from Octanol-Water Partition Coefficient and Aqueous Solubility. *Chemosphere* **1988**, *17*, 21–34. [[CrossRef](#)]
23. van Gestel, C.A.M.; Otermann, K.; Canton, J.H. Relation between Water Solubility, Octanol/Water Partition Coefficients, and Bioconcentration of Organic Chemicals in Fish: A Review. *Regul. Toxicol. Pharmacol.* **1985**, *5*, 422–431. [[CrossRef](#)]
24. Neely, W.B.; Branson, D.R.; Blau, G.E. Partition Coefficient to Measure Bioconcentration Potential of Organic Chemicals in Fish. *Environ. Sci. Technol.* **1974**, *8*, 1113–1115. [[CrossRef](#)]
25. Veith, G.D.; DeFoe, D.L.; Bergstedt, B.V. Measuring and Estimating the Bioconcentration Factor of Chemicals in Fish. *J. Fish. Res. Board Can.* **1979**, *36*, 1040–1048. [[CrossRef](#)]
26. Mackay, D. Correlation of Bioconcentration Factors. *Environ. Sci. Technol.* **1982**, *16*, 274–278. [[CrossRef](#)]
27. Connell, D.W.; Hawker, D.W. Use of Polynomial Expressions to Describe the Bioconcentration of Hydrophobic Chemicals by Fish. *Ecotoxicol. Environ. Saf.* **1988**, *16*, 242–257. [[CrossRef](#)]
28. Bintein, S.; Devillers, J.; Karcher, W. Nonlinear Dependence of Fish Bioconcentration on *n*-Octanol/Water Partition Coefficient. *SAR QSAR Environ. Res.* **1993**, *1*, 29–39. [[CrossRef](#)]
29. Devillers, J.; Bintein, S.; Domine, D. Comparison of BCF Models Based on LogP. *Chemosphere* **1996**, *33*, 1047–1065. [[CrossRef](#)]
30. Spacie, A.; Hamelink, J.L. Alternative Models for Describing the Bioconcentration of Organics in Fish. *Environ. Toxicol. Chem.* **1982**, *1*, 309–320. [[CrossRef](#)]
31. Jonker, M.T.O.; van der Heijden, S.A. Response to “Comment on ‘Bioconcentration Factor Hydrophobicity Cutoff: An Artificial Phenomenon Reconstructed’”. *Environ. Sci. Technol.* **2008**, *42*, 9451–9452. [[CrossRef](#)]

32. Jonker, M.T.O.; van der Heijden, S.A. Bioconcentration Factor Hydrophobicity Cutoff: An Artificial Phenomenon Reconstructed. *Environ. Sci. Technol.* **2007**, *41*, 7363–7369. [[CrossRef](#)]
33. Yang, Z.-Y.; Zeng, E.Y. Comment on “Bioconcentration Factor Hydrophobicity Cutoff: An Artificial Phenomenon Reconstructed”. *Environ. Sci. Technol.* **2008**, *42*, 9449–9450. [[CrossRef](#)] [[PubMed](#)]
34. Garg, R.; Smith, C.J. Predicting the Bioconcentration Factor of Highly Hydrophobic Organic Chemicals. *Food Chem. Toxicol.* **2014**, *69*, 252–259. [[CrossRef](#)] [[PubMed](#)]
35. Dimitrov, S.D.; Dimitrova, N.C.; Walker, J.D.; Veith, G.D.; Mekenyan, O.G. Predicting Bioconcentration Factors of Highly Hydrophobic Chemicals. Effects of Molecular Size. *Pure Appl. Chem.* **2002**, *74*, 1823–1830. [[CrossRef](#)]
36. Dimitrov, S.; Dimitrova, N.; Parkerton, T.; Comber, M.; Bonnell, M.; Mekenyan, O. Baseline Model for Identifying the Bioaccumulation Potential of Chemicals. *SAR QSAR Environ. Res.* **2005**, *16*, 531–554. [[CrossRef](#)]
37. Lombardo, A.; Roncaglioni, A.; Boriani, E.; Milan, C.; Benfenati, E. Assessment and Validation of the CAESAR Predictive Model for Bioconcentration Factor (BCF) in Fish. *Chem. Cent. J.* **2010**, *4*, S1. [[CrossRef](#)]
38. Lunghini, F.; Marcou, G.; Azam, P.; Patoux, R.; Enrici, M.H.; Bonachera, F.; Horvath, D.; Varnek, A. QSPR Models for Bioconcentration Factor (BCF): Are They Able to Predict Data of Industrial Interest? *SAR QSAR Environ. Res.* **2019**, *30*, 507–524. [[CrossRef](#)]
39. Miller, T.H.; Gallidabino, M.D.; MacRae, J.I.; Owen, S.F.; Bury, N.R.; Barron, L.P. Prediction of Bioconcentration Factors in Fish and Invertebrates Using Machine Learning. *Sci. Total Environ.* **2019**, *648*, 80–89. [[CrossRef](#)]
40. Petoumenou, M.I.; Pizzo, F.; Cester, J.; Fernández, A.; Benfenati, E. Comparison between Bioconcentration Factor (BCF) Data Provided by Industry to the European Chemicals Agency (ECHA) and Data Derived from QSAR Models. *Environ. Res.* **2015**, *142*, 529–534. [[CrossRef](#)]
41. Meylan, W.M.; Howard, P.H.; Boethling, R.S.; Aronson, D.; Printup, H.; Gouchie, S. Improved Method for Estimating Bioconcentration/ Bioaccumulation Factor from Octanol/ Water Partition Coefficient. *Environ. Toxicol. Chem.* **1999**, *18*, 664–672. [[CrossRef](#)]
42. United States Environmental Protection Agency. EPISuite™- Estimation Program Interface | USEPA. Available online: <https://www.epa.gov/tsca-screening-tools/epi-suite-estimation-program-interface> (accessed on 10 August 2022).
43. Zhao, C.; Boriani, E.; Chana, A.; Roncaglioni, A.; Benfenati, E. A New Hybrid System of QSAR Models for Predicting Bioconcentration Factors (BCF). *Chemosphere* **2008**, *73*, 1701–1707. [[CrossRef](#)] [[PubMed](#)]
44. Hu, H.; Xu, F.; Li, B.; Cao, J.; Dawson, R.; Tao, S. Prediction of the Bioconcentration Factor of PCBs in Fish Using the Molecular Connectivity Index and Fragment Constant Models. *Water Environ. Res.* **2005**, *77*, 87–97. [[CrossRef](#)] [[PubMed](#)]
45. Lu, X.; Tao, S.; Cao, J.; Dawson, R.W. Prediction of Fish Bioconcentration Factors of Nonpolar Organic Pollutants Based on Molecular Connectivity Indices. *Chemosphere* **1999**, *39*, 987–999. [[CrossRef](#)]
46. Sabljčić, A. The Prediction of Fish Bioconcentration Factors of Organic Pollutants from the Molecular Connectivity Model. *Z. Gesamte Hyg. Grenzgeb.* **1987**, *33*, 493–496.
47. Park, J.H.; Lee, H.J. Estimation of Bioconcentration Factor in Fish, Adsorption Coefficient for Soils and Sediments and Interfacial Tension with Water/ or Organic Nonelectrolytes Based on the Linear Solvation Energy Relationships. *Chemosphere* **1993**, *26*, 1905–1916. [[CrossRef](#)]
48. Sahu, V.K.; Singh, R.K. Prediction of the Bioconcentration Factor of Organic Compounds in Fish. *CLEAN—Soil Air Water* **2009**, *37*, 850–857. [[CrossRef](#)]
49. Hong, H.; Wang, L.; Han, S.; Zou, G. The Estimation of Bioconcentration Factors of Aromatic Hydrocarbons by High Performance Liquid Chromatography. *Toxicol. Environ. Chem.* **1996**, *56*, 185–191. [[CrossRef](#)]
50. Guo, R.; Liang, X.; Chen, J.; Zhang, Q.; Wu, W.; Kettrup, A. Using HPLC Retention Parameters to Estimate Fish Bioconcentration Factors of Organic Compounds. *J. Liq. Chromatogr. Relat. Technol.* **2004**, *27*, 1861–1873. [[CrossRef](#)]
51. Sobanska, A. RP-18 TLC Retention Data and Calculated Physico-Chemical Parameters as Predictors of Soil-Water Partition and Bioconcentration of Organic Sunscreens. *Chemosphere* **2021**, *279*, 130527. [[CrossRef](#)]
52. Sprunger, L.; Blake-Taylor, B.H.; Wairegi, A.; Acree, W.E.; Abraham, M.H. Characterization of the Retention Behavior of Organic and Pharmaceutical Drug Molecules on an Immobilized Artificial Membrane Column with the Abraham Model. *J. Chromatogr. A* **2007**, *1160*, 235–245. [[CrossRef](#)] [[PubMed](#)]
53. Neri, I.; Laneri, S.; DiLorenzo, R.; Dini, I.; Russo, G.; Grumetto, L. Parabens Permeation through Biological Membranes: A Comparative Study Using Franz Cell Diffusion System and Biomimetic Liquid Chromatography. *Molecules* **2022**, *27*, 4263. [[CrossRef](#)] [[PubMed](#)]
54. Russo, G. Study of the Mechanisms of Drug Passage through Biological Barriers Aimed to Optimize Bioavailability and/or Blood-Brain Barrier Permeation. Ph.D. Thesis, University of Naples Federico II, Naples, Italy, 2016. Available online: <http://www.fedoa.unina.it/10986/1/Giacomo%20Russo.pdf> (accessed on 11 August 2022).
55. Soczewiński, E.; Wachtmeister, C.A. The Relation between the Composition of Certain Ternary Two-Phase Solvent Systems and RM Values. *J. Chromatogr. A* **1962**, *7*, 311–320. [[CrossRef](#)]
56. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, *7*, 42717. [[CrossRef](#)]

57. Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A.K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V.V.; Tanchuk, V.Y.; et al. Online Chemical Modeling Environment (OCHEM): Web Platform for Data Storage, Model Development and Publishing of Chemical Information. *J. Comput Aided Mol. Des.* **2011**, *25*, 533–554. [[CrossRef](#)]
58. Fliieger, J.; Tatarczak-Michalewska, M.; Kowalska, A.; Rządowska, M.; Szacoń, E.; Kaczor, A.A.; Matosiuk, D. Fragmental Method KOWWIN as the Powerful Tool for Prediction of Chromatographic Behavior of Novel Bioactive Urea Derivatives. *J. Braz. Chem. Soc.* **2016**, *27*, 2312–2321. [[CrossRef](#)]
59. Liu, C.; Zhang, X.; Nguyen, T.T.; Liu, J.; Wu, T.; Lee, E.; Tu, X.M. Partial Least Squares Regression and Principal Component Analysis: Similarity and Differences between Two Popular Variable Reduction Approaches. *Gen. Psychiatry* **2022**, *35*, e100662. [[CrossRef](#)]
60. Garthwaite, P.H. An Interpretation of Partial Least Squares. *J. Am. Stat. Assoc.* **1994**, *89*, 122–127. [[CrossRef](#)]
61. Næs, T. (Ed.) . *A User-Friendly Guide to Multivariate Calibration and Classification*; NIR Publications: Chichester, UK, 2004; ISBN 978-0-9528666-2-6.
62. Clark, D.E. Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 1. Prediction of Intestinal Absorption. *J. Pharm. Sci.* **1999**, *88*, 807–814. [[CrossRef](#)]
63. Clark, D.E. Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 2. Prediction of Blood–Brain Barrier Penetration. *J. Pharm. Sci.* **1999**, *88*, 815–821. [[CrossRef](#)]
64. Sobańska, A.W.; Brzezińska, E. IAM Chromatographic Models of Skin Permeation. *Molecules* **2022**, *27*, 1893. [[CrossRef](#)] [[PubMed](#)]
65. Zhang, Y.-H.; Xia, Z.-N.; Yan, L.; Liu, S.-S. Prediction of Placental Barrier Permeability: A Model Based on Partial Least Squares Variable Selection Procedure. *Molecules* **2015**, *20*, 8270–8286. [[CrossRef](#)] [[PubMed](#)]
66. Carracedo-Reboredo, P.; Liñares-Blanco, J.; Rodríguez-Fernández, N.; Cedrón, F.; Novoa, F.J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernandez-Lozano, C. A Review on Machine Learning Approaches and Trends in Drug Discovery. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4538–4558. [[CrossRef](#)] [[PubMed](#)]
67. Ciura, K.; Kovačević, S.; Pastewska, M.; Kapica, H.; Kornela, M.; Sawicki, W. Prediction of the Chromatographic Hydrophobicity Index with Immobilized Artificial Membrane Chromatography Using Simple Molecular Descriptors and Artificial Neural Networks. *J. Chromatogr. A* **2021**, *1660*, 462666. [[CrossRef](#)]
68. Lipinski, C.A. Rule of Five in 2015 and beyond: Target and Ligand Structural Limitations, Ligand Chemistry Structure and Drug Discovery Project Decisions. *Adv. Drug Deliv. Rev.* **2016**, *101*, 34–41. [[CrossRef](#)]
69. Lipinski, C.A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov. Today: Technol.* **2004**, *1*, 337–341. [[CrossRef](#)]
70. Chedik, L.; Mias-Lucquin, D.; Bruyere, A.; Fardel, O. In Silico Prediction for Intestinal Absorption and Brain Penetration of Chemical Pesticides in Humans. *Int. J. Environ. Res. Public Health* **2017**, *14*, 708. [[CrossRef](#)]
71. Ibrahim, M.; Garcia-Contreras, L. Mechanisms of Absorption and Elimination of Drugs Administered by Inhalation. *Ther. Deliv.* **2013**, *4*, 1027–1045. [[CrossRef](#)]
72. Eixarch, H.; Haltner-Ukomadu, E.; Beisswenger, C.; Bock, U. Drug Delivery to the Lung: Permeability and Physicochemical Characteristics of Drugs as the Basis for a Pulmonary Biopharmaceutical Classification System (PBCS). *J. Epithel. Biol. Pharmacol.* **2010**, *3*, 1–14.