VILNIUS UNIVERSITY

Kiril Lanevskij

ABSORPTION AND TISSUE DISTRIBUTION OF DRUG-LIKE COMPOUNDS: QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP ANALYSIS

Doctoral Dissertation

Physical Sciences, Biochemistry (04 P)

Vilnius, 2011

The dissertation work was carried out at the Department of Biochemistry and Biophysics, Vilnius University in collaboration with VšĮ „Aukštieji algoritmai" in 2007-2011.

Scientific supervisor:

dr. Remigijus Didžiapetris (VšĮ „Aukštieji algoritmai", Physical sciences, Biochemistry – 04 P)

VILNIAUS UNIVERSITETAS

Kiril Lanevskij

VAISTINIŲ JUNGINIŲ ABSORBCIJA IR PASISKIRSTYMAS AUDINIUOSE: KIEKYBINIO STRUKTŪROS IR AKTYVUMO RYŠIO ANALIZĖ

Daktaro disertacija

Fiziniai mokslai, biochemija (04 P)

Vilnius, 2011

Disertacija rengta 2007-2011 metais Vilniaus universiteto Gamtos mokslų fakulteto Biochemijos ir biofizikos katedroje bendradarbiaujant su VšĮ „Aukštieji algoritmai“.

Mokslinis vadovas:

dr. Remigijus Didžiapetris (VšĮ „Aukštieji algoritmai“, fiziniai mokslai, biochemija – 04 P)

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

ABBREVIATIONS

A – Overall solute hydrogen bonding acidity

AAG – α_1-Acid Glycoprotein

ABC – ATP Binding Cassette superfamily

ADME – Absorption, Distribution, Metabolism, Excretion

AFE – Average Fold Error

ANN – Artificial Neural Network

AUC – Area Under the Curve

B – Overall solute hydrogen bonding basicity

BBB – Blood-Brain Barrier

BUI – Brain Uptake Index

Caco-2 – human Colonic AdenoCarcinoma derived cell line

CNS – Central Nervous System

c_{SS} – Steady state concentration of drug

$c_{u,SS}$ – Unbound steady-state concentration of drug

E – Excess solute molar refraction

F – Oral Bioavailability

$f_{u,br}$ – Fraction unbound in brain

$f_{u,inc}$ – Fraction unbound in microsomal incubation

$f_{u,pl}$ – Fraction unbound in plasma

$f_{u,ti}$ – Fraction unbound in tissue

GALAS – Global, Adjusted Locally According to Similarity modeling methodology

GFA – Genetic Function Approximation

HIA – Human Intestinal Absorption

HPAC – High Performance Affinity Chromatography

HPLC – High Performance Liquid Chromatography

HSA – Human Serum Albumin

IAM – Immobilized Artificial Membrane

IV – Intravenous administration route

J – Transmembrane flux

K_a – absorption rate constant

kNN – k-Nearest Neighbors

K_p – Tissue to plasma partitioning coefficient

$K_p{}^u$ – Tissue to unbound plasma partitioning coefficient

$K_p{}^{uu}$ – Unbound tissue to unbound plasma partitioning coefficient

LDA – Linear Discriminant Analysis

$\log D_{o/w}$ – pH-dependent octan-1-ol/water distribution coefficient

$\log P_{o/w}$ – Octan-1-ol/water partitioning coefficient of neutral species

$\log BB$ – logarithm of Brain/Blood distribution ratio

log K_A^{HSA} – Human Serum Albumin affinity constant

log PS – logarithm of Permeability-Surface area product

MAE – Mean Absolute Error

MDCK – Madin-Darby Canine Kidney cell line

(M)LR – (Multiple) Linear Regression

N_{HA} – No. of hydrogen bond acceptors

N_{HD} – No. of hydrogen bond donors

NLS – Non-linear Least Squares regression procedure

N_{RB} – No. of rotatable bonds

OR – Oral administration route

p(S_{P-gp}) – Probability of the compound being a P-glycoprotein substrate

PAMPA – Parallel Artificial Membrane Permeability Assay

PCR – Principal Component Regression

P_e – Effective permeability in Caco-2 monolayers

P_{eff} – Effective jejunal permeability

P_F – Flow-limited permeability

PLS – Projection to Latent Structures by means of Principal Least Squares

P_{para} – Paracellular permeability

PPB – Plasma Protein Binding

P_{trans} – Transcellular permeability

P_{UWL} – Unstirred Water Layer permeability

Q^2 – Cross-validation determination coefficient

(Q)SAR – (Qualitative) Structure Activity Relationship

RBC – Red Blood Cells

RF – Random Forest

$RMSE$ – Residual Mean Square Error

ROC – Receiver Operating Characteristic

S – Solute polarity/polarizability parameter

S_{BBB} – CNS Access Score

SLC – Solute Carrier superfamily

SVM – Support Vector Machine

TEER – TransEpithelial Electrical Resistance

$(T)PSA$ – (Topological) Polar Surface Area

UWL – Unstirred Water Layer

V_d – Volume of distribution

V_{SS} – Steady-state volume of distribution

$V_{u,SS}$ – Unbound steady-state volume of distribution

V_x – McGowan characteristic volume

V_z – Terminal volume of distribution

WDI – World Drug Index database

INTRODUCTION

A major challenge in drug discovery projects is determination of ADME (Absorption, Distribution, Metabolism, Excretion) properties of new compounds. These properties have a profound influence on therapeutic efficacy of the drug, define the required dosing regimen, as well as the safety margins. Unfavorable pharmacokinetic profiles of lead compounds may result in costly failures at later stages[1,2]. The issues related to poor target access may be encountered in all therapeutic areas, but they are especially relevant in CNS drug projects since brain entry of many drug-like molecules is effectively restricted by blood-brain barrier[3]. A variety of experimental techniques have been developed for early identification of problematic leads. These are typically implemented in a screening cascade starting from readily automated and more cost efficient *in vitro* assays and moving to resource intensive animal models later on[1]. Use of *in silico* prediction methods, such as QSAR (Quantitative Structure Activity Relationship) modeling, has become an attractive alternative in early screening: these methods are cheaper, have very high throughput, do not require any samples of the analyzed compounds, and can be applied even prior to their actual synthesis.

Computational approaches used to predict ADME properties of drug-like compounds can be divided into two groups differing by their main principles. The first group represents statistical approaches attempting to achieve the best possible fit of experimental values to a matrix of molecular descriptors. A novel GALAS (Global, Adjusted Locally According to Similarity) methodology that combines statistical modeling with a special similarity-based routine has been developed and validated in our group[4,5]. These methods work well for simple physicochemical characteristics of drug molecules, such as lipophilicity or solubility, as well as for certain ADME and toxicity endpoints where relatively large amounts of experimental data are available. The requirement of a large and sufficiently diverse data set is crucial for extracting relevant information from the descriptor matrix and ensuring broad applicability limits of the model. Still, methods of this kind are often used for prediction of complex

multi-mechanism biological parameters including intestinal absorption, brain penetration, and partitioning into tissues. Data availability for these endpoints is very limited and existing sets of measurements clearly do not suffice for capturing the (often non-linear) effects of involved processes. Performance of the resulting models outside of the narrow region of chemical space covered by the training set remains highly questionable.

In mentioned cases the second approach – mechanistic physicochemical modeling – is much more useful. Here the property of interest is decomposed into simpler terms on the basis of common chemical and biological knowledge. Simple QSAR equations relating the latter terms to relevant physicochemical parameters are then derived. Applicability of mechanistic models is much less sensitive to the size of the training set since they rely on fundamental theoretical principles that are valid for all compounds regardless of the region of chemical space they represent. Also, the interpretation is easier because the used descriptors reflect molecular properties that have a clear meaning and are intuitively understandable to chemists. Consequently, mechanistic models may guide lead optimization efforts towards the desired levels of target properties.

Objectives of the study. The first step towards understanding the fate of a biologically active compound in the body is characterizing its delivery to the site of action. This aspect, covering the letters "A" and "D" in ADME, was the topic of the current work. The **main objective** was to perform detailed physicochemical QSAR analysis of commonly used drug properties related to their absorption and distribution in the body. In order to accomplish this goal, the following **specific tasks** have been set:

1. Evaluate the possibilities of using GALAS modeling methodology for obtaining accurate predictions of plasma protein binding strength (%*PPB*).
2. Determine what physicochemical properties of drug-like compounds govern their rate of passive diffusion across cellular permeability barriers and the extent of distribution into tissues.

3. Apply the reached conclusions to develop predictive models for the following endpoints:

   (i) permeability of drug-like compounds in several systems – *in situ* rodent blood-brain barrier, *in vivo* human small intestine, *in vitro* Caco-2 cell monolayers;

   (ii) two tissue distribution characteristics – rodent brain/plasma partitioning ratio and steady-state human volume of distribution.

4. Compare the ionization dependences of the considered endpoints.

5. Investigate the relationship between quantitative blood-brain transport parameters and qualitative data characterizing accessibility of drugs to central nervous system.

Practical value. The predictive models discussed here can be used in both academic research projects and pharmaceutical industry for evaluation of the respective properties of chemicals. A third-party research work using our blood-brain permeability model has already been described in the literature[6]. An important feature of physicochemical models is their capability to use either calculated or measured values of input parameters (log P or pK_a). Predictions performed for new compounds usually rely on calculated descriptors, but these can be substituted with experimental data if available. This approach, combining *in silico* predictions with *in vitro* measurements, (the so called *in combo* concept)[7] may help substantially improve prediction accuracy for novel drug classes where the uncertainty in physicochemical parameter values may lead to significant errors in estimation of more complex ADME properties. Another advantage of physicochemical models lies in their interpretability. The researcher can arbitrarily alter the model inputs and investigate the mechanistic relationship between properties as well as gain an insight on structural modifications that need to be introduced to the lead compound.

SCIENTIFIC NOVELTY

The majority of the topics considered in this work have already been extensively studied by QSAR methods. Nevertheless, most previous publications focus on purely statistical characterization of collected experimental data without proper analysis from biological point of view. This work raises an opposite objective – it presents an attempt to achieve better mechanistic understanding of the processes under consideration and to derive robust and readily interpretable models that would relate biological endpoints to simple physicochemical parameters.

A key element of this work is application of ion-specific partitioning concept to describe the ionization dependence of the modeled processes. Although the basic theoretical background of intestinal absorption and tissue distribution is already well established, this study extends the existing knowledge by combining known drug transport equations with ion-specific models of transepithelial permeability in the gut and non-specific lipid binding in tissues. Moreover, none of the previous QSAR studies of *in vitro* absorption in Caco-2 monolayers analyzed pH and stirring rate dependence of permeability on a large data set containing multiple data points per compound.

An important part of this study is dedicated to drug transport across blood-brain barrier (BBB). Here we present the first published model of BBB permeability (log *PS*) that is parameterized on more than 100 compounds and accounts for multiple physicochemical and physiological determinants of passive diffusion rate. Despite the vast number of existing models for predicting brain/blood partitioning ratio (log *BB*), the model described here is the first one that not only separates the contributions of compound binding to tissue and plasma constituents but also includes a proper data verification procedure enabling identification of erroneous values. Finally, it is demonstrated that a combination of the above parameters allows classifying drugs by CNS access. In contrast to earlier models, this classifier is theoretically grounded, and it is shown that calculated classification score well correlates with experimental brain exposure data.

STATEMENTS FOR DEFENSE

1. The predictive model for the extent of plasma protein binding (%*PPB*) derived using GALAS modeling methodology produces sufficiently accurate predictions for use as a descriptor in models of subsequent drug distribution processes.

2. Passive permeability of drug-like compounds across biological membranes can be described by the following physicochemical parameters: lipophilicity, ionization, hydrogen bonding potential, and molecular size.

3. Ionized species permeate through brain endothelium and intestinal epithelium 2 to 3 orders of magnitude slower than neutral molecules. The pattern of permeability-ionization relationship is independent of the cell type under consideration.

4. Tissue affinity of drug-like compounds well correlates with their lipophilicity, while ionization has a detrimental effect on non-specific binding to tissues only for acidic drugs.

5. Drug-like compounds can be classified by their access to central nervous system using a linear combination of quantitative transport parameters – brain/plasma equilibration rate and equilibrium brain/plasma partitioning ratio.

6. According to the proposed classifier, a compound can be considered accessible to central nervous system if its permeability is sufficient for reaching steady-state in brain.

1. LITERATURE OVERVIEW

Figure 1.1 presents a schematic view of the drug fate in the body after oral administration. In order to enter systemic circulation, the drug has to be absorbed in the gut and escape hepatic first-pass metabolism. In the central compartment the drug exists in equilibrium between free and protein bound form. Only free (unbound) molecules may cross cell membranes, partition into the tissues, and reach their target receptors. Accordingly, it is the unbound drug that can be extracted by liver and undergo glomerular filtration in kidneys, leading to irreversible elimination from the body. The overall picture becomes quite complicated if we consider that some drugs may undergo biliary excretion and re-enter gastrointestinal tract where they can be reabsorbed or excreted in feces. Finally, not only the parent compound but also its metabolites may contribute to therapeutic or undesired activity of the drug[8,9].

The scheme below highlights the critical importance of plasma binding for all subsequent pharmacokinetic processes. The major considerations related to plasma protein binding will be discussed in section 1.1, while sections 1.2-1.3 deal with general aspects of tissue distribution and unique features of blood-brain barrier that is responsible for significantly restricted drug delivery into the brain compared to other tissues. A detailed discussion of the processes involved in drug absorption and permeation across other cellular barriers will be presented in sections 1.4-1.5, whereas metabolism and excretion are out of the scope of the current study. Finally, section 1.6 will provide an overview of recent advances in QSAR modeling of the considered endpoints.
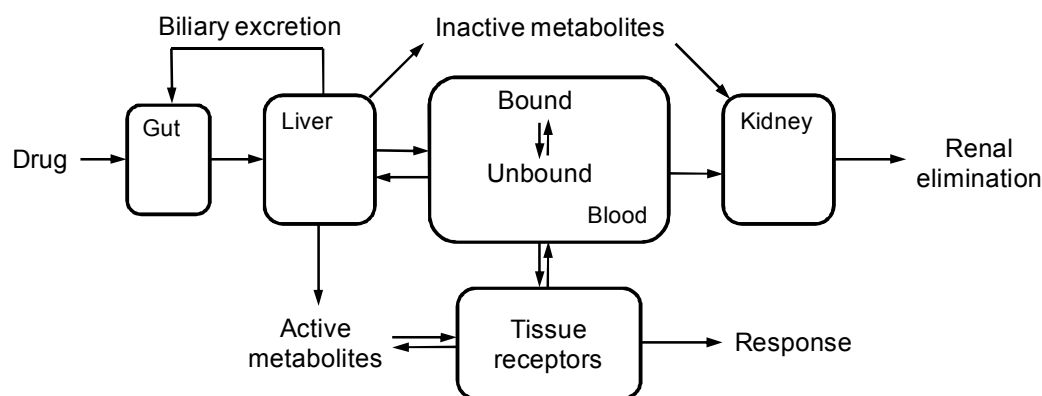
Figure 1.1 Overview of drug distribution in the body. Based on Refs.[9,10]

15

1.1. Plasma protein binding

In plasma most drug molecules reversibly bind to plasma proteins. The formation of drug-protein complex is a rapid process that can be expressed by the following mass-balance equation:

$$[D] + [P] \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [DP]$$

$$K_a = \frac{1}{K_d} = \frac{k_1}{k_{-1}} = \frac{[DP]}{[D] + [P]} \tag{1.1}$$
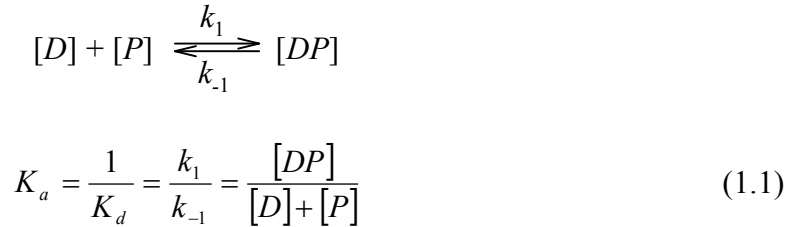
Here $[D]$, $[P]$, and $[DP]$ are concentrations of free drug, free protein, and drug-protein complex, respectively. Kinetic constants of both complex formation (k_1) and dissociation (k_{-1}) are usually very large, and equilibrium is reached within milliseconds time frame[11]. The interaction strength between drug and protein molecules can be characterized by an association (affinity) constant K_a or its inverse – dissociation constant K_d – according to Eq. (1.1). K_d can be viewed as the concentration of free drug that occupies half of available binding sites at equilibrium conditions[9].

The majority of drugs exhibit linear binding, which means that the fraction unbound remains independent of drug concentration within the therapeutic range. Non-linear behavior may be observed when drug concentration approaches K_d and the binding sites become saturated. Concentration-dependence is less likely to manifest if the compound binds to albumin, but it is more common in case of AAG binding due to relatively low amount of AAG in serum[11]. The degree of saturation, *i.e.*, the fraction of total binding sites occupied at a given drug concentration $[D]$ can be estimated using Eq. (1.2):

$$r = \frac{[DP]}{[D] + [DP]} = \frac{n \cdot K_a \cdot [D]}{1 + K_a \cdot [D]} \tag{1.2}$$

where n is the total number of binding sites on the protein. Note that all above calculations assume that the drug only interacts with binding sites of the same type. For molecules that can bind to multiple binding sites, the calculation be-

comes more complex as in this case several K_{ai} values for different sites have to be taken into account[9]:

$$r = \sum_{i=1}^{m} \frac{n \cdot K_{ai} \cdot [D]}{1 + K_{ai} \cdot [D]} \qquad (1.3)$$

Efficient binding lowers the fraction of the drug that can enter the tissues by transmembrane diffusion and ultimately exhibit pharmacological activity. At the same time, drug-protein complexes may serve as a depot prolonging drug action. The extent of plasma protein binding is, therefore, a key ADME property that needs to be determined at the earliest possible stages of drug discovery along with simple physicochemical parameters, such as lipophilicity, ionization, and solubility.

1.1.1. Proteins involved in drug binding

The main proteins that bind xenobiotics in human plasma are human serum albumin (HSA), α_1-acid glycoprotein (AAG), and lipoproteins. In a very simplified way, it can be stated that the preference towards a particular binding protein is largely determined by the ionization state of the drug – acids mostly interact with HSA, bases – with AAG, while unionized lipophilic compounds tend to bind to lipoproteins in a non-specific manner[11,12]. Steroid hormones can bind to both HSA and AAG[13], but many of them are transported by specific proteins such as Sex Hormone Binding Globulin (SHBG)[14], Corticosteroid Binding Globulin (CBG)[15], transcortin, and others[11]. Table 1.1 summarizes the common properties of drugs that bind to aforementioned proteins, while the two major carriers are briefly reviewed below.

Human Serum Albumin (HSA) is the most abundant carrier protein in human plasma at concentrations of 35-50 g/L (about 0.6 mM)[11,16]. Its natural function is transport of fatty acids and some other endogenous molecules (bilirubin, eicosanoids). The albumin molecule consists of a single polypeptide chain that contains 585 amino acid residues and has molecular weight of 66.5 kDa. Albumin exhibits broad ligand specificity and has numerous binding sites for endogenous and exogenous molecules. For example, palmitate ion can

bind to at least seven distinct sites[17]. Yet, the majority of xenobiotics bind to one or two high-affinity sites identified by Sudlow et al.[18] (Figure 1.2, (A-C)).

Site I, also called warfarin-binding site, is a large hydrophobic cleft in subdomain II; its entrance is aligned with basic residues. This binding site is flexible and can accommodate many structurally diverse compounds, mostly dicarboxylic acids and bulky heterocyclic molecules with a negatively charged moiety in the middle[17].

The general organization of Site II (benzodiazepine site) located in subdomain IIIA is remarkably similar, but this site is narrower and more rigid, leading to higher ligand selectivity[17]. Apparently, the presence of negative charge

Table 1.1 Human plasma proteins participating in drug binding, their ligand binding sites, and examples of compounds interacting with these sites.

Protein	Binding site	Common ligand classes	Example ligands
Human Serum Albumin	I	Large heteroaromatic acids, sulfamides	Warfarin, phenbutazone, indomethacin, tolbutamide, furosemide, valproic acid, bilirubin, eisosanoids
	II	Medium-chain fatty acids, arylpropionic acids, benzodiazepines	Ibuprofen, naproxen, diazepam, etodolac octanoic acid, hippuric acid, monooleyl glycerols
	Other sites	Long chain fatty acids, basic drugs	Amitriptyline, probenecid, debrisoquine, propofol, halothane, palmitic acid, myristic acid
α1-Acid Glycoprotein	F1*S variant	bulky hydrophobic molecules	Dipyridamole, warfarin, binedalin, prazosin
	A variant	basic aromatic drugs	Imipramine, desipramine, maprotiline, diphenhydramine
Lipoproteins		Triglycerides, neutral hydrophobic molecules	Cholesterol, retinol, etretinate, orlistat, reserpine
SHBG		Androgenic and estrogenic hormones	Testosterone, DHT, estradiol, levonorgestrel
CBG		Corticosteroid hormones	Hydrocortisone, prednisolone, prednisone

is not absolutely required for efficient binding – a typical ligand is diazepam, which is a weak base, mostly non-ionized at physiological conditions. Binding does not necessarily occur exclusively at a single binding site – in crystal structures a number of drugs have been found associated with lower affinity secondary sites in addition to the primary Sudlow sites. Albumin can also interact with some molecules bearing a positive charge – in this case the interaction often takes place at one of the secondary sites[19].

Serum albumins are highly conserved across species, yet some significant interspecies differences are evident. Kosa et al. studied drug-drug interactions between several typical albumin site I and site II ligands in various species[20]. Their studies revealed that the binding of warfarin and phenbutazone to both human and rat albumins was competitive, indicating that the binding regions of these drugs overlap (site I). In the case of ibuprofen and diazepam, a similar competitive character of binding was only observed for human albumin, while in rats diazepam was very weakly bound, even though the important site II residues are almost fully conserved. The authors attribute this discrepancy to smaller cavity size in rat albumin[20].

α_1-Acid Glycoprotein (AAG), or orosomucoid, is another protein that significantly contributes to drug transport in the circulation. Human AAG molecule consists of a single 183 residue polypeptide chain with 5-6 attached N-linked sialylated glycans. AAG is an unusual glycoprotein due to its high carbohydrate content (45% of the overall 41-43 kDa mass) and acidic surface (pI = 2.8-3.8)[13]. Its plasma levels are normally much lower than those of albumin (0.4-1 g/L), but they can markedly increase in case of inflammation or infection as AAG is one of the major acute phase proteins[11]. It is considered a natural immunomodulatory agent and also a carrier for steroids and certain drug compounds. Due to its negatively charged surface, AAG preferentially interacts with molecules bearing a positive charge – one high-affinity binding site for basic drugs is usually discerned. In most individuals two distinct genetic variants of AAG – A and F1*S – are expressed, each encoded by a separate gene. Their amino acid sequences differ by 22 residues, leading to slightly dif-

ferent properties of the binding sites[21]. F1*S-variant contains a large hydrophobic pocket that accommodates a large variety of drugs in a mostly non-specific manner. In fact, this isoform does not show a clear preference towards a particular ionization state and can even bind acidic compounds such as warfarin, phenobarbital, or retinoic acid. In A-variant the binding site is smaller and exhibits higher ligand specificity. A pharmacophore describing its typical ligands could be defined as a basic ionization center with two aromatic rings connected by short aliphatic linker(s)[21,22], as depicted in Figure 1.2, (D, E).
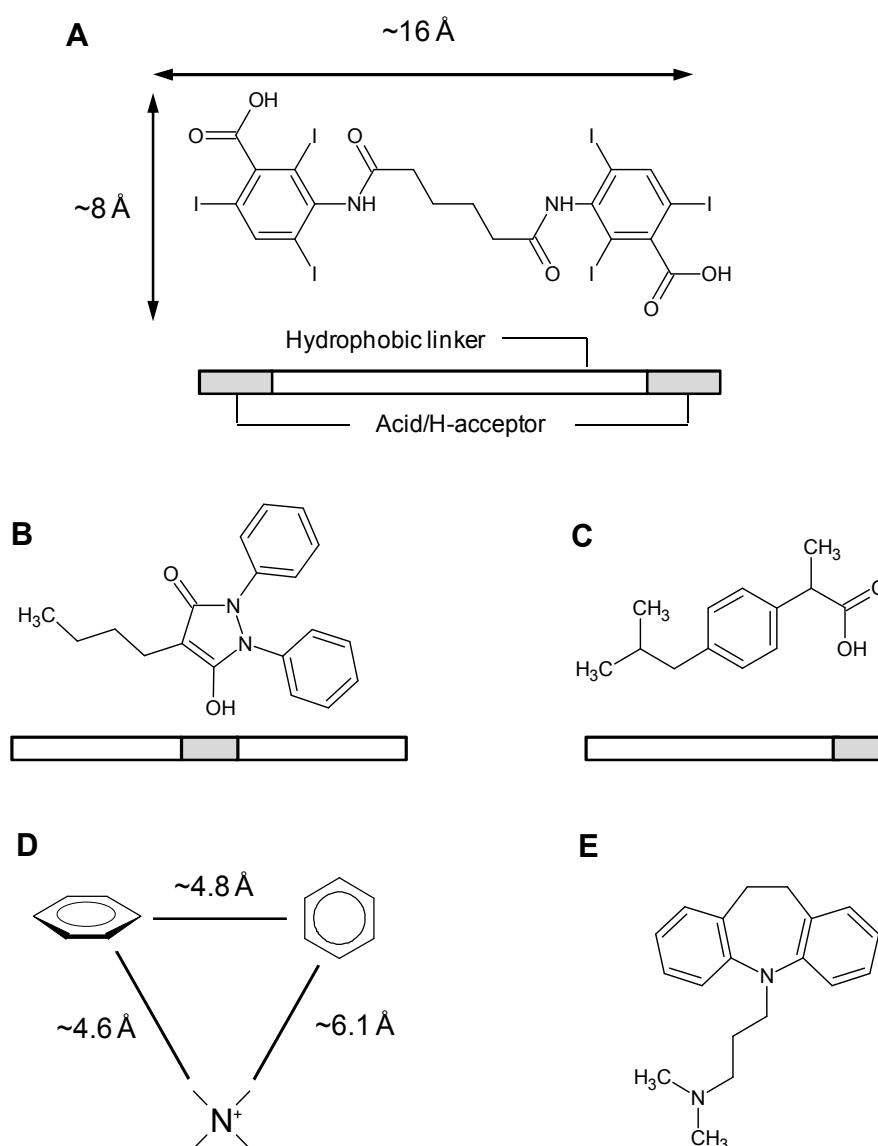
Figure 1.2 Typical ligands of HSA and AAG, their orientation within the binding sites, and common structural features: (A) iophenoxate (HSA Site I); (B) phenylbutazone (HSA Site I); (C) ibuprofen (HSA Site II); (D) generalized pharmacophore of A-variant AAG drug binding site; (E) imipramine, an example drug containing the AAG pharmacophore.

20

1.1.2. Experimental determination of plasma binding

The extent of plasma protein binding is commonly reported either as an overall percentage fraction of the bound compound (%*PPB*) or as a free (unbound) fraction $f_{u,pl}$:

$$\%PPB = \frac{c_{bound}}{c_{total}} \cdot 100\% = \left(1 - f_{u,pl}\right) \cdot 100\% \tag{1.4}$$

Additionally, studies focusing on the interactions with a particular protein report association constants K_a. The aforementioned parameters can be determined by a variety of methods that can be divided into two broad-categories: separative methods relying on physical separation of free and bound drug molecules and non-separative methods that measure the degree of association indirectly[9]. Techniques that are most often used for plasma protein binding determination in practice, especially in early drug discovery, belong to the category of separative methods. These are briefly overviewed below.

Equilibrium dialysis (Figure 1.3, (A)) remains the reference method for measuring unbound drug concentrations in plasma. The experiments are conducted in a device consisting of two compartments separated by a semi-permeable membrane. It restricts the passage of macromolecules, such as proteins and drug-protein complexes, but is freely diffusible to small molecular weight solutes. One compartment contains a plasma sample or a protein solution, while a buffer solution of the test chemical is placed into the other compartment. After a certain amount of time, the free drug concentrations on both sides of the membrane equalize and equilibrium is established. The extent of protein binding can then be quantified after sampling the drug from both compartments. Equilibrium dialysis is attractive by its simplicity, yet this method has a range of issues that may lead to significant %*PPB* determination errors. Non-specific adsorption on the compartment cells and dialysis membrane reduces the apparent concentration of free drug, especially in case of highly lipophilic molecules. Moreover, due to osmotic pressure, water flows from the buffer compartment to the protein compartment and dilutes the protein solution. If this volume shift exceeds 10%, a special correction has to be introduced. Other

issues include overestimation of free concentrations due to possible protein leakage through the membrane and uneven distribution of ionic species between the compartments in case of weakly bound ionizable drugs (Donnan effect). Finally, equilibration is usually slow (12-48 h), and it can take even longer for large tightly bound molecules. Furthermore, the equilibration time is *a priori* unknown and additional studies are needed to determine the required duration of the experiment. Fortunately, the analysis can be automated and accelerated with modern 96-well format systems[9,23,24].

Ultrafiltration (Figure 1.3, (A)) is another simple technique that differs from equilibrium dialysis by two main aspects: (i) here both drug and protein are placed into the same compartment; (ii) equilibration of free drug across the semi-permeable membrane is accelerated by applying pressure or centrifugal force of around 2000 g. Ultrafiltration suffers from the same major disadvantages, namely, non-specific binding, Donnan effect, and protein leakage; but this method reduces the possible errors arising from these effects since the experiment is completed much faster. Due to its higher throughput, ultrafiltration is the method of choice for ranking large numbers of compounds at the earliest stages of drug discovery[9,24].

In **Ultracentrifugation** (Figure 1.3, (B)) the separation of free drug from protein and drug-protein complexes is based on the application of very large gravitational force (> 600,000 g). Centrifugation is performed until macromolecules sediment on the bottom of the test tube (very low density lipoproteins, denoted VLDL in the figure, form a thin layer floating on top). The unbound drug can then be sampled from supernatant. Since the ultracentrifugation setup does not involve dialysis membranes, this technique overcomes most issues associated with the methods described above and is particularly suitable for the analysis of drugs prone to non-specific adsorption. However, the results of ultracentrifugation measurements are not always consistent with equilibrium dialysis. Substantial errors may arise for high molecular weight drugs due to sedimentation of the drug itself. Also, this method has low throughput and requires high cost equipment, limiting its potential use[9,24].

| A Ultrafiltration and equilibrium dialysis | B Ultracentrifugation |
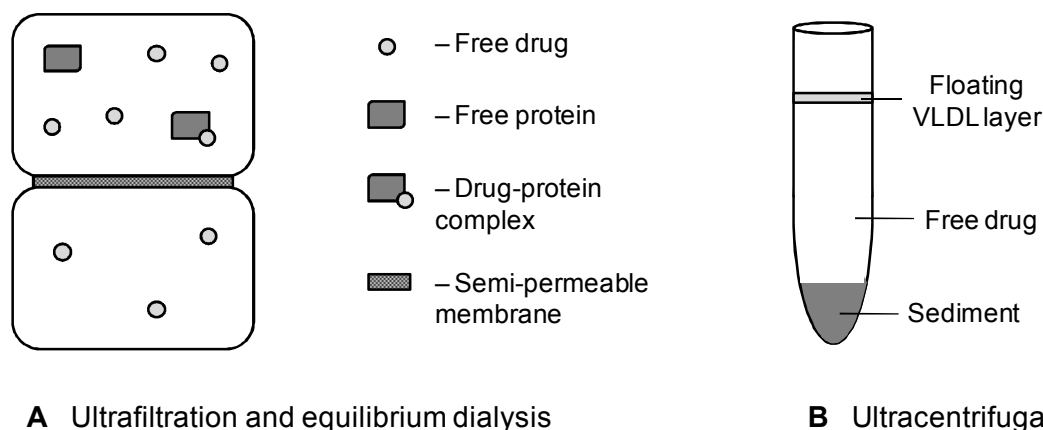
Figure 1.3 Separation of bound and unbound drug in commonly used assays for determination of the extent of plasma protein binding.

HPLC. Affinities of drugs to a particular protein can also be determined by the means of high performance liquid chromatography (HPLC). Several chromatographic techniques can be used to study drug-protein interactions. These include size-exclusion chromatography where both interacting species are free in the solution and affinity chromatography where the protein is immobilized on the column as a stationary phase[9]. However, the former technique is ineffective and inconvenient, and most studies dealing with drug binding to human serum albumin use high performance affinity chromatography (HPAC)[25,26]. In these experiments a solution of the analyzed drug is injected into a HSA column, and the affinity is estimated from the retention time – high affinity ligands elute later than weak binders. The major advantages of HPAC technique are the possibility to reuse the same protein preparation for multiple experiments, ease of automation, and high throughput of measurements (often 5-10 min is sufficient for a single experiment). The obtained results are in good agreement with values determined by equilibrium dialysis and other solution-based methods. Recently, a number of successful attempts to measure association strength using HPAC columns with immobilized AAG or high density lipoproteins have also been reported[10].

1.2. Partitioning into tissues. Volume of distribution

Both pharmacological and toxicological potential of a drug as well as its ultimate fate in the body to a large extent depends on the efficiency of distribution into various tissues[27]. In most tissues drug molecules freely diffuse within the aqueous environment of blood and interstitial fluid through the leaky pores in capillary walls. The notable exceptions are restrictive cellular barriers such as the blood-brain barrier or placental barrier. At steady-state the unbound concentrations on both sides of the capillary wall become equal, whereas the concentrations of the bound drug depend on its relative affinities to plasma and tissue constituents. Compounds highly bound to albumin or AAG are retained in blood, while tissue accumulation mostly takes place due to non-specific binding to lipid components (neutral lipids, phospholipids, etc.)[28]. In practice partitioning between plasma and tissues is usually evaluated using a special parameter termed volume of distribution (V_d).

Volume of distribution is a theoretical concept that relates the amount of drug in the body (A) at time t to its concentration in circulation (c):

$$V_d = \frac{A(t)}{c(t)} \tag{1.5}$$

V_d does not have a strict physiological meaning, and the values beyond 0.6-0.8 L/kg do not represent any particular physical volume. Several different distribution volumes can be found in the literature, each of them having a somewhat different meaning and usage (see Table 1.2 for an overview).

1. Central volume of distribution (V_c) is defined as the ratio of the administered dose (D) to the initial concentration in circulation (c_o):

$$V_c = \frac{D}{c_o} \tag{1.6}$$

Such definition implicitly assumes that the drug is instantaneously distributed within the central compartment and no subsequent redistribution to other compartments occurs. It is an idealistic parameter that corresponds to an apparent volume from which hepatic and renal elimination occurs but does not provide any information regarding drug affinity to tissues and is at best of limited use[29].

Table 1.2 Different volume of distribution (V_d) terms found in the literature.

Notation	Description	Conditions	Administration route
V_c	Central V_d	Reflects c_{pl} at time point $t = 0$	Intravenous
V_{SS}	V_d at steady-state	c_{pl} = const (equilibrium)	Intravenous
V_z	Terminal V_d	c_{pl} reached pseudo-equilibrium	Intravenous
V_z/F	Extravascular terminal V_d	c_{pl} reached pseudo-equilibrium	Extravascular

2. Steady-state volume of distribution (V_{SS}) signifies the relationship between the total amount of drug (A_{SS}) and plasma concentration (c_{SS}) at steady-state conditions, which may be observed in case of continuous IV infusion:

$$V_{SS} = \frac{A_{SS}}{c_{SS}}$$ (1.7)

V_{SS} reflects the situation when constant drug concentrations are maintained and clearance is apparently null. Given that it represents equilibrium conditions and is unaffected by side processes, V_{SS} is the best available measure of the balance between plasma and tissue binding strength of a drug. V_{SS} can be determined not only during infusion but also when steady-state conditions are reached after multiple drug dosage[29]. There are different approaches for estimating V_{SS} from raw experimental data including the commonly used method of statistical moments illustrated by Eq. (1.8)[29,30]:

$$V_{SS} = \frac{D \cdot AUMC}{AUC^2} = CL \cdot MRT$$ (1.8)

Here AUC corresponds to the area under the plasma concentration-time curve ($\int_0^\infty c_{pl} \cdot dt$); $AUMC$ is the area under the first moment of the disposition curve ($\int_0^\infty c_{pl} \cdot t\, dt$). The ratio of dose to AUC yields the overall plasma clearance (CL), and $AUMC/AUC$ ratio defines MRT – mean residence time in the system[31]. Note that here and further in the text the term AUC implies integration over the entire curve (time interval from 0 to ∞), except where the interval is explicitly denoted in the superscript (e.g., $AUC^{0 \to T}$).

3. Terminal volume of distribution (V_z or V_{area}). The plasma concentration-time curve observed after a single bolus injection of a drug can be divided into distribution and elimination phases. During the distribution phase, drug disappearance from plasma can be mostly attributed to partitioning into tissues rather than elimination. Ultimately, pseudo-equilibrium conditions are reached when the unbound tissue to plasma ratio approaches unity ($c_{ti}{}^{u} = c_{pl}{}^{u}$), and the overall drug concentrations in all compartments start to decrease due to total clearance from the body. At this point, volume of distribution reaches its maximal value, designated V_z (or V_{area}), which can be calculated as follows[29,30]:

$$V_z = \frac{D}{\lambda_z \cdot AUC} = \frac{CL}{\lambda_z} \qquad (1.9)$$

Here the additional parameter λ_z represents the first-order elimination rate constant. Terminal volume of distribution is typically slightly higher than V_{SS} due to the difference between equilibrium and pseudo-equilibrium conditions. In most cases the small discrepancy between V_z and V_{SS} can be neglected. It may become unacceptably large if a significant fraction of the drug is eliminated earlier than pseudo-equilibrium conditions are reached. Since the expression of V_z contains the elimination rate term, it is important to consider that terminal volume of distribution of drugs may be overestimated in patients with renal insufficiency[29]. The difference between V_{SS} and V_z determination is illustrated in Figure 1.4.
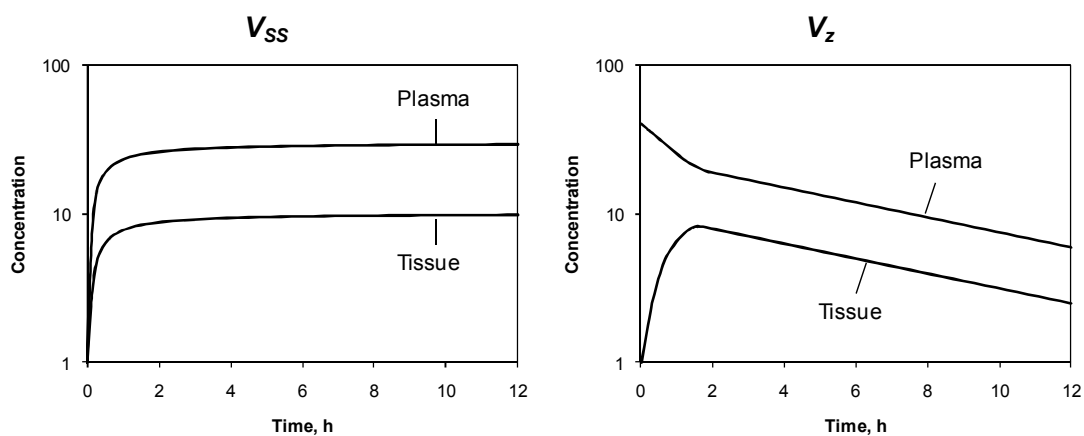
Figure 1.4 Concentration-time curves in plasma and tissues and the corresponding distribution volumes obtained after different administration routes (adapted from Ref.[29]): V_{SS} – from IV infusion data; V_z – from single IV bolus dose data.

4. Volume of distribution after extravascular administration (V_z/F). The above definition of V_z implies that the drug is administered intravenously. If the drug enters systemic circulation by an extravascular route (intraperitoneal, subcutaneous, or most typically – oral), accurate calculation of V_z becomes problematic. The value obtained from Eq. (1.9) then represents not the actual V_z, but rather the ratio of V_z to bioavailability (F):

$$\frac{V_z}{F} = \frac{CL}{\lambda_z}$$ (1.10)

Even if the fraction reaching systemic circulation is known, in certain scenarios V_z/F may not reflect true distribution volume. To ensure correct estimation, the terminal phase must be a pure elimination phase. If the drug undergoes slow (even if complete) absorption, the terminal phase overlaps with the absorption phase (the so called "flip-flop" process), and terminal distribution volume is overestimated by the calculated V_z/F value[29].

Clinical significance of V_d. The clinical importance of the distribution volume of a drug may be inferred from Eq. (1.9). Rearranging this equation and expressing λ_z in time units (1.11) yields Eq. (1.12) for estimation of the drug's half-life ($\tau_{1/2}$) in the body:

$$\tau_{1/2} = \frac{\ln 2}{\lambda_z}$$ (1.11)

$$\tau_{1/2} = \ln 2 \cdot \frac{V_d}{CL}$$ (1.12)

Thereby, both V_d and clearance directly influence duration of drug effect, which in turn controls the required dosing regimen. Larger volume of distribution leads to longer half-life and allows for longer intervals between doses[32].

In case of repeated dosage (*i.e.*, administration of daily maintenance doses), drug concentrations in plasma and tissues level off, plasma concentration profile becomes smoother, and ultimately steady-state conditions are reached[29,32]. The magnitude of concentration fluctuations between the concurrent doses is determined by V_{SS}. Larger V_{SS} results in smaller difference between minimal and peak plasma concentrations, which may be important for drugs

having a narrow therapeutic window[29]. The time needed to reach steady-state is independent of the dosing regimen and is predefined by terminal half-life. Drug levels close to steady-state are observed only after 3-5 half-lives after the first dose[33]. This may be unacceptable for certain medications used in emergency situations, and a loading dose (D_L) may be necessary in the beginning of the treatment:

$$D_L = \frac{V_{SS} \cdot C_{SS}}{F}$$ (1.13)

The purpose of administering a loading dose is rapid achievement of desired therapeutic levels represented by C_{SS} in Eq. (1.13), which otherwise would take a very long time[29]. The required loading dose is a function of V_{SS} and bioavailability.

Volume of distribution and drug toxicity. Although drug accumulation in the tissues associated with large V_d values is often considered advantageous from purely pharmacokinetic point of view[32], it can also be related to toxic effects. Firstly, volume of distribution is highly correlated with lipophilicity, which is the main determinant of non-specific (baseline) toxicity. Baseline toxicity defines the minimum toxicity level of a compound and is usually attributed to disruption of the functions of biological membranes[34]. There is also a specific phenomenon – phospholipidosis – that is closely related to high tissue concentrations of drugs. Phospholipidosis is a lysosomal storage disorder characterized by intracellular accumulation of phospholipids enclosed within lamellar bodies[35-37]. It is caused by inhibition of lysosomal phospholipase, enhanced phospholipid biosynthesis, and other processes induced by repeated administration of cationic amphiphilic compounds[38]. Although there is currently no direct evidence of a link between phospholipidosis and clinical adverse effects, affected tissues may exhibit inflammatory reactions and histopathological changes. Therefore, excessive phospholipid storage is considered a factor of increased risk with respect to potential tissue damage[35]. Notably, common properties of known inducers – presence of a lipophilic substructure and a basic ionizable center[37] – are basically the same as those needed for efficient par-

titioning into tissues. In a recent QSAR study, it was shown that V_d alone, as a parameter reflecting the compounds' residence in the tissues, allows correctly identifying a large number of drugs that induce phospholipidosis[39].

1.3. Blood-brain barrier

Brain tissue is separated from systemic circulation by one of the most effective physiological barriers – the blood-brain barrier (BBB)[40]. BBB is a complex biological formation consisting of a dense network of tight junctions interconnecting adjacent brain capillary endothelial cells, a variety of metabolic enzymes, and carrier proteins[40,41]. BBB maintains brain homeostasis, limits entry of various endogenous compounds, and protects the brain from xenobiotics[42]. Because of pronounced barrier properties of cerebral vasculature, designing new central nervous system (CNS) drugs remains a challenging task with a higher attrition rate compared to any other therapeutic area[43]. Brain penetration is of great importance not only for CNS-targeted pharmaceuticals but also for peripheral drug candidates as in this case permeable compounds can cause side effects in brain[44].

1.3.1. Anatomy and functions

Blood-brain barrier is present in almost all brain regions, but it is not the only formation connecting neural tissue and blood. The others are the choroid plexus epithelium constituting blood-cerebrospinal fluid barrier (BCSFB) and avascular arachnoid epithelium completely enclosing the CNS[45,46]. Yet, it is the BBB that has the largest surface area (between 150 and 200 cm^2/g tissue in humans) and is the main interface of blood-brain exchange[46]. Endothelial cells forming the walls of cerebral microcapillaries serve as a physical, transport, and metabolic barrier simultaneously[46,47]. Physical barrier functionality is accomplished by junctional complexes between adjacent cells that physically restrict the passage of molecules though the intercellular cleft. In order to enter the brain, chemicals have to diffuse by transcellular route where they encounter the transport barrier – a number of carrier proteins expressed in endothelium mediating specific uptake or efflux of certain drugs. Finally, the metabolic bar-

rier is represented by extracellular peptidases and nucleases metabolizing macromolecules and ATP as well as intracellular monoamine oxidase and cytochrome P450 proteins responsible for biotransformation of many small molecules passing through the cells[47].

In mature cerebrovasculature neural tissue is separated from the capillary wall by several additional cell layers. The endothelial layer is surrounded by basal lamina, which is 30-40 nm thick and includes several classes of extracellular matrix proteins (structural – collagen, elastin; specialized – fibronectin, laminin; proteoglycans), Cell Adhesion Molecules (CAM), and signaling proteins. Most of these proteins are synthesized in pericytes that are enclosed within the lamina and associated with the outer surface of the capillaries[45,48]. Pericytes are found exclusively at small capillaries where they cover about 20 to 30% of endothelial cell surface[45], while in larger vessels they are replaced by a continuous layer of smooth muscle. From the other side, basal lamina is in close contact with astrocytes and microglial cells[46,47]. The supporting cells perform BBB induction, maintenance, and regulatory functions, whereas the permeability barrier function is primarily attributed to the endothelium[45].
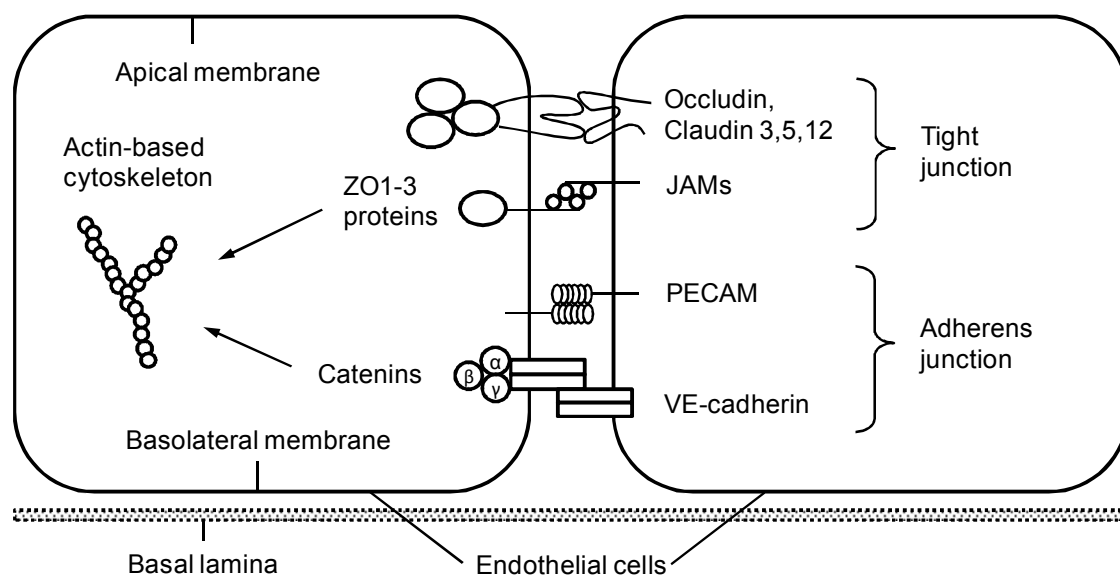
Figure 1.5 Intercellular junctions between brain capillary endothelial cells and their molecular composition (adapted from Refs.[46,47]).

Intercellular junctions. The extremely dense junctional network between the endothelial cells includes two types of cellular junctions – adherens junctions (*zonula adherens*) and tight junctions (*zonula occludens*). Figure 1.5 provides a schematic view of two adjacent endothelial cells and the macromolecules participating in the formation of both junction types.

Adherens junctions are located closer to the abluminal (brain-facing) side. Their major components are Ca^{2+}-regulated transmembrane glycoproteins VE-cadherins that are linked with cytoskeleton by cytoplasmic scaffolding proteins α-, β-, and γ-catenins. In a similar way, homotypic adhesion is mediated by platelet-endothelial cell adhesion molecules (PECAM). Adherens junctions form a continuous "adhesion belt" that gives the tissue structural support and is essential for the formation of tight junctions.

Tight junctions are localized on the luminal (blood-facing) side of the endothelial cells. Here the intercellular contacts are maintained by occludin-claudin complexes as well as Junctional Adhesion Molecules (JAM1-3). Cytoplasmic scaffolding function and interaction with actin cytoskeleton is mediated by accessory proteins Zonula Occludens (ZO1-3) and cingulin[45-48]. Tight junctions form a dense network of parallel interconnected strands that ensure particularly high electrical resistance of *in vivo* blood-brain barrier[49-51]. Furthermore, tight junction strands inhibit lateral diffusion of lipids and proteins within the membrane bilayer, thereby, constructing a "fence" that effectively separates luminal and abluminal regions of the membrane and precludes mixing of the components between these regions[52]. This fact is important to consider when studying permeability of drug molecules as it means that the compounds have to traverse through the cytoplasm in order to get to the other side of the cell. On the other hand, in the absence of such "fence", ampholytes could circumvent BBB by lateral diffusion with apolar parts of their molecules submerged in the bilayer and polar parts facing the interstitial fluid.

1.3.2. Quantitative blood-brain transport parameters

Blood to brain transport efficiency of a compound is usually evaluated by two quantitative parameters – log PS and log BB, characterizing the rate and extent of BBB penetration, respectively.

PS (Permeability-Surface area product) is a kinetic parameter representing the rate of diffusion across endothelial cell membranes. This term is defined as the compound's permeability coefficient P (in units of cm/s) multiplied by cerebral microcapillary surface area S, which is a characteristic of the test species (about 100-130 cm^2/g in rat)[53,54] and is constant for all considered drugs. By its physiological meaning, PS is equivalent to the unidirectional influx rate constant K_{in} corrected for cerebrovascular blood flow rate Q[55]:

$$K_{in} = Q\left(1 - e^{-PS/Q}\right) \qquad (1.14)$$

The main advantage of PS (or log PS in the form appropriate for QSAR) as a measure of blood-brain transport is that it reflects the compounds' intrinsic capability to diffuse through BBB membranes. This value is not biased by other processes such as binding to plasma proteins or further redistribution within the brain that includes lipid binding or sink effect of cerebrospinal fluid.

log BB (logarithm of Brain/Blood distribution ratio) is a thermodynamic parameter that represents the extent of brain delivery at steady-state conditions. It is expressed as the ratio of steady-state concentrations (c^{SS}) of the drug in brain tissue and plasma or, alternatively, as the ratio of areas under the concentration-time curves (AUC) in the respective compartments[56]:

$$\log BB = \log \frac{c_{br}^{SS}}{c_{pl}^{SS}} = \log \frac{AUC_{br}}{AUC_{pl}} \qquad (1.15)$$

Oppositely to a common misconception, log BB does not show how rapidly the drug would permeate across BBB but instead describes the partitioning equilibrium governed by drug binding to brain and plasma constituents. Both log PS and log BB parameters are needed for a comprehensive evaluation of blood-brain transport potential of new compounds, whereas any single property would not suffice to guide decision making in drug discovery.

1.3.3. Experimental determination of blood-brain transport

BBB permeability (log *PS*) can be measured using a variety of assays ranging from the oldest known indicator dilution method[57] through more advanced techniques such as quantitative autoradiography or intracerebral microdialysis to modern non-invasive methods including positron emission tomography and magnetic resonance imaging[55-58]. Non-invasive techniques are particularly advantageous since they are applicable to humans, yet they are costly and thus rarely used in drug discovery[56]. log *PS* measurements are usually conducted in rodents using one of the popular methods: intravenous injections, brain uptake index, and *in situ* brain perfusion[55].

Intravenous injections. Experiments involving intravenous administration of test compound to rodents represent one of the easiest methods of log *PS* determination[59]. The simplest single-time point analysis is performed as follows: the arterial blood of the animal is sampled at regular intervals after bolus injection in order to obtain the plasma concentration-time curve. At the terminal time point (*T*), the animal is sacrificed, and a sample of brain tissue is taken to determine the brain concentration c_{br}^{T}. If the assumption of unidirectional uptake is valid, the initial transport rate is given by:

$$K_{in} = \frac{AUC_{pl}^{0 \to T}}{c_{br}^{T}} \qquad (1.16)$$

PS can be calculated from K_{in} according to Kety-Renkin-Crone equation (1.14) where two boundary conditions can be identified: (i) $K_{in} \approx PS$ when $PS << Q$; (ii) $K_{in} \approx Q$ when $PS >> Q$. For accurate estimation of permeabilities in the intermediate range, it is necessary to know flow rate *Q*, which is usually estimated using freely diffusible compounds (butanol, diazepam) as markers[55].

Despite its simplicity, the method is sensitive enough as it allows for multiple passages of molecules through the brain capillary bed. However, this also leads to several serious pitfalls. At sufficiently late time points, back flux becomes significant, and the assumption of unidirectional transport is no longer valid. Biased results may also be obtained for metabolically labile substances and for molecules highly bound to albumin or other plasma proteins[55,56].

BUI and *in situ* perfusion. These techniques were designed to overcome the shortcomings of IV injection approach. Brain Uptake Index (BUI) suggested by Oldendorf[60] is a single pass method involving a rapid intracarotid bolus injection of labeled test compound followed by decapitation within 5-15 s from the start of experiment. *BUI*% is then defined as the relative first-pass extraction of test compound compared to that of a permeable marker (usually tritiated water). *PS* products are calculated from the absolute extraction values (*E*) according to the modified Kety-Renkin Crone equation:

$$E = 1 - e^{-PS/Q} \tag{1.17}$$

BUI technique allows freely manipulating the composition of the injectate buffer, avoiding metabolism, and minimizing the consequences of buffer mixing with endogenous plasma. These advantages come at the cost of much lower sensitivity compared to simple intravenous injection experiments. Therefore, BUI method is only applicable to compounds having at least moderate permeability[55-57].

Takasato et al.[61] introduced the concept of *in situ* rat brain perfusion. In these experiments the common carotid artery of the anesthetized rat is ligated and perfused with oxygenated physiological buffer for up to 10 min. Such experimental setup not only ensures high sensitivity of measurements but also offers full control over the perfusate composition, flow rate, and duration of perfusion. The concentration of the test compound in the perfusion fluid is maintained constant over the time course of the experiment, simplifying the calculation of K_{in} and *PS* values. Smith et al.[62] improved the method by introducing an additional step that helps to minimize perfusate mixing with plasma. In their modified approach, the endogenous blood flow is stopped by severing the heart ventricles prior to the start of perfusion. Dagenais et al.[63] adapted the method for use with mice. This provided the possibility to obtain accurate passive diffusion rates of P-gp substrates by conducting the measurements in mdr1a knockout mice.

To summarize, it can be stated that *in situ* perfusion is the best established log *PS* determination method that retains the advantages of both IV injection

and BUI techniques and addresses their major issues. Nevertheless, it requires a rather complex experimental setup and involves a surgical procedure[55-57]. As a result, many research groups prefer not to measure BBB permeation rate but rather evaluate blood-brain transport of drug only in terms of log *BB*.

log *BB* determination. The two alternative expressions of log *BB* given in Eq. (1.15) are closely related to the method of determination. The steady-state drug concentrations in both compartments can be measured directly if the drug is administered to the test animal by the means of constant IV infusion maintained for several hours until blood concentrations reach a plateau[3,64]. More often, however, brain/plasma partitioning ratio is determined using multiple time point method. The animals receive a single bolus IV injection, and several animal groups are sacrificed at different time intervals postdose. Whole blood and brains are then collected and analyzed for presence of test compound and its associated metabolites. Once the drug concentrations in brain and plasma are known at several time points, it is possible to derive the respective concentration-time curves and estimate the brain/plasma partitioning ratio from *AUC* values[3,56]. Single time point data are of limited use since the observed concentration ratios are time-dependent[56]. Still, in some cases such data can be retained in further analysis if brain and plasma concentrations have been determined at a sufficiently late time point, and it can be assumed that pseudo-equilibrium has been reached. In general, estimation of log *BB* values using these methods closely resembles determination of steady-state and terminal volume of distribution discussed in previous section.

Obviously, intravenous infusion yields the most precise log *BB* values, while the multiple time point method introduces a higher degree of uncertainty. Unfortunately, no standard protocol exists for determination of brain/plasma ratio. Investigators use different drug doses and administration routes (intravenous, oral, intraperitoneal, subcutaneous), and there is no uniformly accepted number of sampling points[56]. As a result, the overall quality of available data is somewhat compromised in comparison with better standardized endpoints.
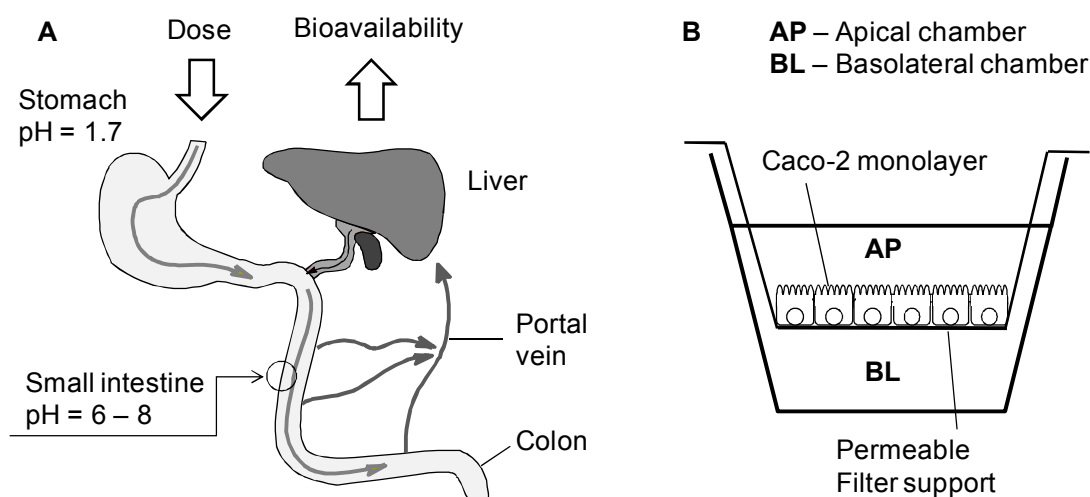
Figure 1.6 Overview of the considered intestinal permeability systems: (A) human gastrointestinal tract; (B) typical experimental setup for Caco-2 permeability study.

1.4. Gastrointestinal barrier

Drug administration *per os* is the easiest, safest, and therefore most popular administration route[65]. Ensuring efficient delivery of oral drugs remains an important challenge for medicinal chemists. Poor oral bioavailability raises various issues such as the need for increased dosing, development of complex formulations, or forcing drug administration by intravenous route[65,66].

Figure 1.6, (A) illustrates the path of the orally administered drug in human gastrointestinal tract. First, the active component of the formulation has to withstand acidic pH of stomach. After food ingestion, stomach pH can promptly increase to 7, but within three hours it returns to the fasted state value of 1.7. After the gastric emptying, the drug moves down the intestine where it can be absorbed[66-68]. The two major factors that affect the absorption efficiency of a compound are its solubility and permeability. Before the compound can be absorbed, it has to dissolve in the aqueous environment – poorly soluble drugs may exhibit solubility-limited absorption. Another possibility is that the administered amount of drug does not exceed its solubility, but dissolution is slow, and the compound cannot fully dissolve during the gastrointestinal transit time. In such cases dissolution rate-limited absorption is observed[69]. It should also be noted that solubility of lipids and other lipophilic molecules in the intestine is better than in the pure water due to the solubilizing effect of bile micelles[69,70].

The permeability barrier function is mediated by mucosal epithelium covering the gut wall, which will be briefly described below. About 99% of the total absorptive surface is located in jejunum and ileum. The transit time through these regions of the intestine is 3-5 hours[66-68].

Sietsema defined the extent of human intestinal absorption (%HIA) as the amount of drug passing from the lumen into the tissue of the gastrointestinal tract[71]. Once in the tissue, the drug is considered absorbed, although here it can be metabolized by gut wall enzymes. The fraction of the drug that remains unchanged at this point gets into the portal vein and reaches liver where it can undergo first-pass metabolism. Non-absorbed drug is retained in colon and subsequently excreted in feces. The fraction of orally administered drug that manages to escape intestinal and hepatic first-pass metabolism, and ultimately enters systemic circulation is designated oral bioavailability (%F)[72].

1.4.1. Mucosal epithelium

The internal surface of the intestines is lined by an epithelial cell layer, which in turn is covered by heavily glycosylated and hydrated gel – mucus layer[73]. The cell layer mainly consists of columnar epithelial cells (enterocytes) responsible for selective permeability to certain compounds but also includes lymphocytes and specialized epithelial cells (gastric foveolar mucus cells and goblet cells) that produce mucins – the primary components of the mucus layer[67,68,73]. Mucins are very large ($MW \approx 2$ MDa) glycoproteins with about 90% carbohydrate content rich in sialic acids residues[67,68]. Mucus layer precludes bacteria and other large particles from contacting directly with the epithelium. The consequences of defective mucus formation are evident in diseases such as cystic fibrosis[73]. Intestinal mucosa does not prevent passage of small molecules, but the bulk flow is significantly slowed down in the vicinity of the epithelium. Diffusion across this viscous fluid devoid of convective mixing forces, referred as aqueous boundary layer or *unstirred water layer* (UWL), may be the rate limiting step in the absorption of lipophilic molecules[73,74].

The small intestine has ridged surface resulting in a three-fold higher surface area than could be expected in case of a simple smooth tube of the same radius. This area is further increased up to ten times by multiple microvillous structures forming in the apical membranes of enterocytes. The microvilli are distributed unevenly throughout the intestines with their highest density found in duodenum, jejunum, and ileum (hence, the larger surface area in these regions), and lower density – in proximal colon[66-68]. The overall absorptive surface of human gastrointestinal tract is estimated at > 200 m^2 (Ref.[75]).

The organization of the permeability barrier is similar to BBB. Enterocytes are connected by various junctional complexes including adherens and tight junctions as well as desmosomes providing strong adhesive bonds between the cells and serving as attachment points for keratin filaments[76]. The major difference from BBB is that the tight junctions in the intestinal epithelium are much leakier than those in brain endothelium. The increased leakiness is explained by the presence of aqueous pores ranging in size from about 7 to 9 Å in jejunum and 3-4 Å in ileum[67,68]. Small solutes can traverse these pores taking the paracellular route across the epithelium instead of transcellular diffusion *via* the enterocyte.

1.4.2. Experimental determination of absorption

Quantitative human intestinal absorption data are usually obtained either from oral bioavailability or from urinary/fecal excretion data[77].

Bioavailability. If we compare the factors influencing the extent of absorption and bioavailability, the latter can be defined as "absorption + gut wall and hepatic first-pass metabolism". Therefore, %*HIA* is always $\geq$ %*F*, and if a drug has high bioavailability (%*F* > 80%), its first-pass metabolism is apparently minimal and fraction absorbed can be considered roughly equivalent to %*F*. Bioavailability can be calculated as the ratio of *AUC* values obtained after oral and intravenous administration of the drug[78]:

$$\%F = \frac{AUC_{OR}}{AUC_{IV}} \tag{1.18}$$

If %*F* is low, it may significantly underestimate absorption because a fraction of the drug may not reach systemic circulation due to extensive first-pass metabolism despite having passed the intestinal barrier.

Excretion in urine and feces. This method is applicable if the absorbed drug is predominantly excreted from the body in urine and it can be completely recovered from urine and feces. In this case %*HIA* can be determined as the fraction of drug-related material including parent drug and its metabolites undergoing urinary excretion after oral administration. This approach may produce unreliable results for compounds with urinary excretion < 80% since part of the absorbed drug could still be excreted in feces. Alternatively, it is possible to express %*HIA* as 100% – % excreted in feces. Yet, there is less fecal excretion data available in the literature compared to urinary excretion data[77,78].

Excretion in urine after oral and intravenous administration. If the data regarding drug fate after intravenous administration are available, the fraction absorbed may be evaluated more accurately than from oral administration data alone. %*HIA* is then expressed as the ratio of cumulative urinary excretion of drug and its metabolites obtained from OR and IV routes[77,78]. The major advantage of this method is the lack of requirements for urinary excretion being the main elimination route and complete recovery from urine and feces. Therefore, it has broader applicability limits and yields larger errors only when the fraction of the drug excreted in urine is small after both OR and IV administration. Yet, obtaining IV data is problematic for poorly water soluble drugs. Other issues that preclude reliable determination of %*HIA* using this approach are dose- or formulation-dependent absorption and gut wall metabolism[77].

***In situ* intestinal perfusion.** In addition to the aforementioned methods for estimating %*HIA*, intestinal permeability may be measured directly by jejunal perfusion[79-81]. In these experiments a perfusion tube is inserted into human jejunum, and drug solution is allowed to pass through a 10 cm long occluded segment of the intestine. The measured difference between "in" and "out" concentrations of a drug yields its effective jejunal permeability coefficient P_{eff} that relates to %*HIA* by kinetic equations. Jejunal perfusion is a useful tech-

nique, but since it is an *in vivo* assay involving humans, it is rarely performed, and P_{eff} data available up to date are very limited[66].

1.4.3. Permeability in Caco-2 monolayers

A variety of *in vitro* approaches have been employed for preclinical evaluation of gastrointestinal absorption. Several cell-based assays such as 2/4/A1, immortalized Madin-Darby Canine Kidney (MDCK) cells, and human colon adenocarcinoma derived cell line (Caco-2) have been developed to screen for permeability[82]. There is an approximate sigmoidal relationship between flux through such cells and the extent of absorption[83]. Experiments with cell monolayers are relatively easy to conduct, require smaller quantities of compounds for analysis, and may be used in earlier stages of drug development compared to *in vivo* methods. A need stills exists for *in silico* prediction of permeability of drug-like compounds in various *in vitro* systems. Computational methods could be useful for validation of experimental results and identification of potentially biased measurements, studies of transport mechanisms, as well as for better understanding of the relationship between *in vivo* and *in vitro* permeability. Caco-2 cell assay has become a gold standard for *in vitro* evaluation of intestinal permeability and absorption[84,85]; therefore, it was selected as one of the objects of the current study.

Experimental determination of permeability coefficients in Caco-2 cell line is a two-step procedure where the first step involves cultivating Caco-2 cell monolayers and the transport experiment itself is conducted in the second step[84]. The cells are grown on permeable filter supports for 3-4 weeks in advance of the experiment to ensure that monolayers are well-differentiated and express relevant brush border enzymes. Then the filters are transferred to the wells with two separated chambers shown in Figure 1.6, (B). At the beginning of the experiment, the buffer solution of the test compound adjusted to appropriate pH is placed into the apical chamber (donor compartment). The buffer from the basolateral chamber (receiver compartment) is then sampled at 4-5 time points to quantify the compound passing across the monolayer and calcu-

late its permeability coefficient in AB (apical to basolateral) direction. Donor and receiver compartments may be reversed in order to obtain permeability in BA direction, which provides an insight on the absorption mechanism – AB/BA or BA/AB permeability ratio > 3 is a clear indication of the contribution of carrier-mediated processes. The experiments can be repeated at other concentrations of test compound to determine if the transport is saturable and to establish kinetic parameters (K_M and V_{max})[84].

The major advantage of absorption studies using *in vitro* cell lines, such as Caco-2, is the possibility to manipulate the experimental conditions[86]. Adjusting pH of the buffer in both donor and acceptor compartments enables the researcher to assess the ionization dependence of permeation rate. In addition, one can evaluate what effect the unstirred water layer has on permeability. In order to reduce the UWL thickness, the solution can be stirred, which is usually accomplished by insertion of magnetic stirrer or shaking the apparatus[87,88].

A common problem with Caco-2 permeability measurements is their interlaboratory variability. As outlined by Artursson & Carlsson, Caco-2 cells form heterogeneous populations, so that the properties of monolayers may differ between laboratories and even within the same laboratory when the cultures from different time periods are compared[83]. This variability contributes to the overall uncertainty in measured permeation rates when the data from different sources are aggregated together.

1.5. Membrane permeation mechanisms

As briefly mentioned in previous sections, drug molecules may cross the intestinal epithelium either by paracellular route through aqueous pores in tight junctions or by transcellular route passing through the cell. Possible transport routes and mechanisms are illustrated in Figure 1.7 and discussed below.

Paracellular pathway is the main transport route for small hydrophilic solutes. Although the intercellular space is sealed by junctional complexes, in most cells types some flux through the pores between occludin and claudin-based protein structures can still occur. Paracellular transport can be viewed as

passive molecular size-restricted diffusion through water-filled chambers in the tight junctions connecting neighboring cells. The diffusion rate depends on a variety of factors including the morphology of the epithelium and properties of the solute. Epithelium can be characterized by its porosity ε (fraction occupied by pores) and pore tortuosity factor δ (path length along the paracellular channel)[74,89]. Avdeef defined the ratio of these two parameters as the pore capacity factor ε/δ[90,91].

Another determinant of the epithelium leakiness is the size of the pores. In recent publications two types of pores are discerned: (i) smaller high capacity charge-selective pores with average pore radius (R) in different cell types ranging from 4 to 15 Å; (ii) larger low capacity non-selective pores[76,82]. The origin of the smaller pores is attributed to specific claudin expression patterns, while the second type is possibly related to temporary irregularities or "breaks" in the otherwise intact tight junction barriers[76]. Leaky pores permit some limited flux of proteins and bacterial lipopolysaccharides[73], but due to their low capacity they do not significantly contribute to the overall transport rate of most small molecules and can be neglected in the analysis. Charge selectivity of small pores arises due to an apparent potential drop across the epithelium. As a result, cations are driven from lumen not only by concentration gradient but also by electrochemical gradient, whereas electrochemical contribution is unfavorable for diffusion of anions[74,89].
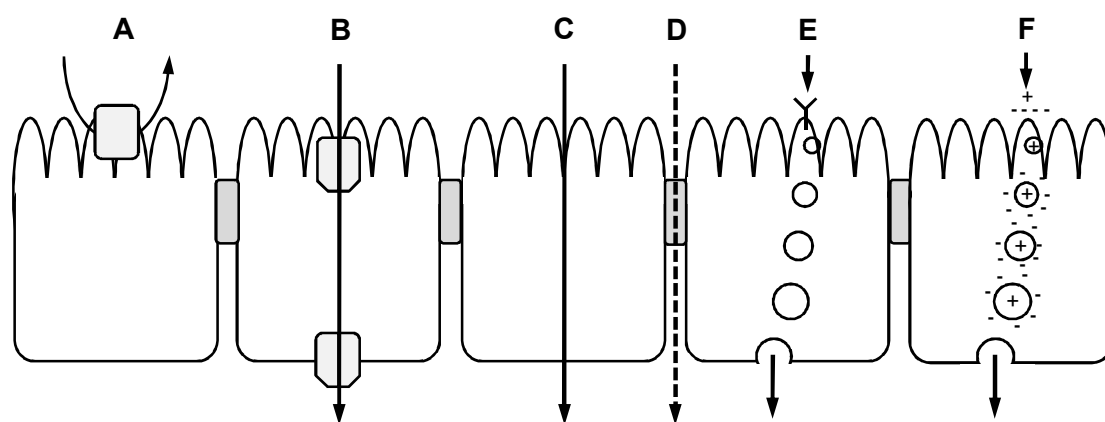
Figure 1.7 Mechanisms of permeation across cellular barriers (adapted from Ref.[46]): (A) carrier-mediated efflux; (B) carrier-mediated influx; (C) passive transcellular diffusion; (D) passive paracellular diffusion; (E) receptor mediated transcytosis; (F) adsorption mediated transcytosis.

Relationship between paracellular diffusion rate of a compound and its molecular size is usually approximated by complex empirical functions such as Renkin molecular sieving function (see Section 3.2 for details). In general, the rate of diffusion through the pores quickly decreases as hydrodynamic radius of the solute approaches R. Therefore, paracellular transport is only significant for small molecules having MW < 250-300. The epithelium in human small intestine is considered leakier than colonic epithelium. This statement has been confirmed by the measurements of transepithelial electrical resistance (TEER = 50-100 $\Omega \cdot cm^2$ in small intestine *vs*. 300-400 $\Omega \cdot cm^2$ in colon)[91]. Similarly to colon and colon-derived Caco-2 cell line, the paracellular pathway is more tightly sealed in MDCK cells, while leakier 2/4/A1 monolayers are thought to better mimic this particular aspect of gastrointestinal barrier[82]. Paracellular route is essentially eliminated at BBB due to the extremely dense junctional network in brain endothelium (TEER = 1800-2000 $\Omega \cdot cm^2$) forcing the molecules to take the transcellular pathway[49-51].

Passive transcellular diffusion (or simple diffusion) is the predominant transport mechanism for the majority of larger and sufficiently lipophilic compounds. To pass through the epithelial or endothelial monolayer by this route, the molecules have to cross two phospholipid membranes on the apical (luminal) and basolateral (abluminal) sides separated by a cytoplasmic layer[92]. Getting to the other side of the lipid bilayer portion is considered the rate limiting step[79]. An important characteristic of passive diffusion is that it is a completely non-specific process governed by general physicochemical laws. Since carrier proteins are not involved, transport by this mechanism is non-saturable – solute flux across the membrane linearly depends on its concentration gradient[74].

The main determinant of passive diffusion rate of a compound is its lipid solubility, which depends on several factors[93]. Affinity towards hydrophobic core of the membrane is reflected by the compound's lipophilicity. It is usually modeled by partitioning ratios between various organic solvents and water – a widely used example is 1-octanol/water partitioning system[74]. Furthermore, transfer from the aqueous phase into lipid phase is highly energetically unfa-

vorable for charged species. This fact is reflected in pH-partition hypothesis stating that only uncharged molecules permeate, and the diffusion rate is determined by the compound's neutral form fraction at the pH of the environment[94]. A number of recent studies demonstrated that pH-partition hypothesis cannot satisfactorily explain higher than expected permeability of charged drugs in certain systems indicating that the actual relationship between permeability and ionization is more complex[94-97]. Another factor contributing to the energy barrier is the loss of hydration layer. The molecules have to desolvate before they can pass the membrane – hydrophilic compounds with a large number of hydrogen bonding sites are poorly absorbed. Various authors outlined that in order to cross BBB, drug-like structures should have polar surface area $\leq$ 70-120 Å^2 or a number of N + O atoms not larger than the numerical value of log $P_{o/w}$[98].

Theoretically, the rate of passive diffusion could vary in different cell types due to the differences in membrane composition. However, there are no conceptual differences in permeation mechanism, and similar trends could be expected at blood-brain barrier, intestinal epithelium, Caco-2 cell monolayers, hepatocytes, and other cell types[74].

Carrier-mediated transport. Membrane permeation mechanisms that involve participation of transporter proteins can be differentiated according to the transport direction and the energy source. Drug efflux from brain and intestinal epithelium mainly takes place through the members of ATP-binding cassette (ABC) superfamily that mediate primary active transport using energy generated from ATP hydrolysis. Influx carriers expressed at BBB and enterocyte membranes belong to the Solute Carrier (SLC) superfamily. These are either facilitative uniporters, that can transport their substrates in both directions depending on the concentration gradient, or secondary active transporters, that utilize energy accumulated in the form of ion gradients created by ion pumps such as Na^+/K^+-ATPase. A list of transporter proteins expressed in the considered absorptive systems is presented in Table 1.3, whereas their membrane localization and substrate transport directions are shown in Figure 1.8.

Table 1.3 Selected carrier proteins participating in drug transport across small intestine and BBB along with their typical substrates. Data aggregated from Refs[46,99,100].

Protein(s)	Expression		Transport Mechanism	Example substrates
	BBB	Enterocyte		
ABC superfamily				
P-gp (ABCB1)	+	+	Primary active (ATPase)	Amitriptyline, colchicine, doxorubicin, erythromycin, etoposide, paclitaxel, ritonavir
BCRP (ABCG2)	+	+	Primary active (ATPase)	Flavonoids, mitoxantrone, nitrofurantoin, zidovudine
MRP1 (ABCC1)	+		Primary active	Aflatoxins, daunorubicin
MRP2 (ABCC2)	+	+	(ATPase)	p-Aminohippurate, vincristine
MRP3 (ABCC3)		+		Taurocholate, M3G, M6G
MRP4 (ABCC4)	+			Adefovir, DHEAS, urate
MRP5 (ABCC5)	+			cAMP, 6-mercaptopurine
SLC superfamily				
PEPT1		+	H^+ symport	Di- and tripeptides,
PEPT2		+	H^+ symport	β-lactams (cefadroxil, cephalexin)
ASBT		+	Na^+ symport	Bile acids
OSTα,β		+	Facilitative	Bile acids
SGLT		+	Na^+ symport	Glucose
GLUT1	+		Facilitative	Glucose
GLUT2		+	Facilitative	Glucose
ENT1-2	+	+	Facilitative	Nucleobases, nucleosides,
CNT1-2	+	+	Na^+ exchange	nucleotides (adenosine, inosine)
MCT1	+	+	H^+ exchange	Small monocarboxylic acids
OATs	+	+	Ion exchange	Organic anions (allopurinol, estrone-3-sulfate, indoxyl sulfate, DHEAS, fluvastatin, valsartan, prostaglandins)
OATPs	+	+	Ion exchange	
OCTs	+	+	Facilitative	acetylcholine, cimetidine, ranitidine, metformine, spermidine, amantadine
OCTNs	+	+	Various	L-carnitine, pyrilamine, quinidine, TEA, verapamil
CTL1	+		Facilitative	Choline
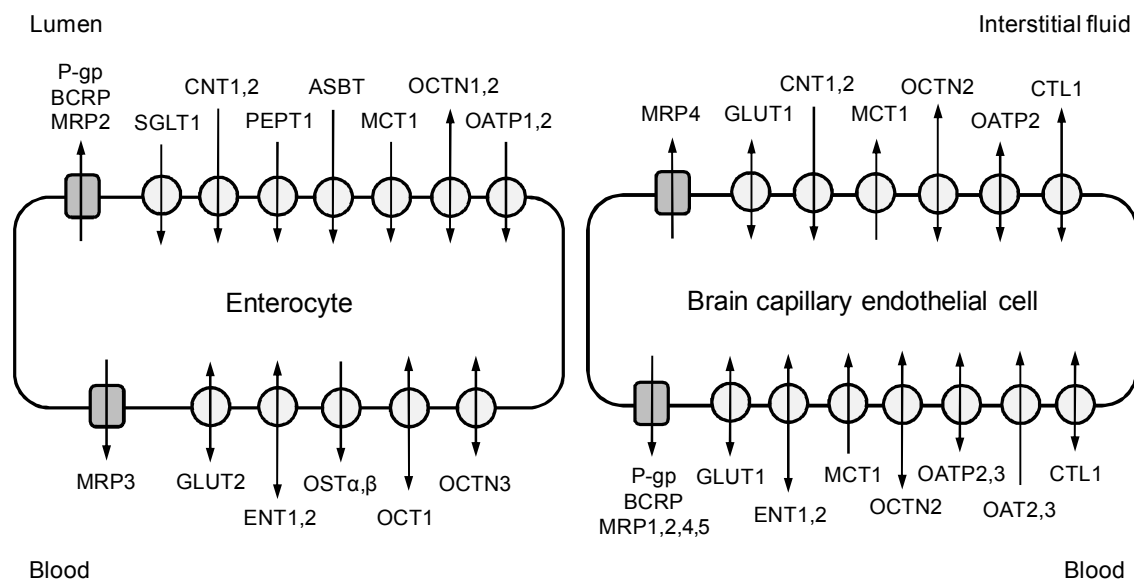
Figure 1.8 Membrane localization of key efflux and influx transporters expressed in mammalian enterocytes and brain capillary endothelium. Data aggregated from Refs.[46,99,100]

ABC superfamily. Efflux carriers belonging to the ABC superfamily are ubiquitously expressed membrane spanning proteins that participate in numerous absorption, distribution and elimination processes. Their primary function is protection of tissues from toxic substances and other potentially harmful compounds of both endogenous and exogenous origin[42]. Moreover, ABC transporters along with drug metabolizing enzymes play a major role in drug-drug interactions. Although they are of great importance for many healthy tissues, particularly high levels are found in cancer cells. A key feature of these transporters is very broad substrate specificity, which together with high expression levels is responsible for poor susceptibility of cancer to treatment by many drugs (multidrug resistance phenomenon). Although there are 48 known ABC transporter genes constituting seven families (A to G), only 9 members from 3 families are pharmacologically significant[101,102]:

- **P-gp** (P-glycoprotein), also denoted ABCB1 or MDR1, is the first discovered and currently best studied member of ABC superfamily. It is expressed in liver, kidney, intestine, blood-brain and placental barriers[103]. P-gp extrudes a wide variety of structurally diverse compounds but demonstrates a preference towards large ($MW > 400$) molecules having good

hydrogen bond accepting capabilities but no acidic ionizable centers[104]. Its substrates include numerous anti-cancer drugs, antihypertensive and anti-arrhytmic agents, antimicrobials, tricyclic antidepressants, HIV protease inhibitors.

- **BCRP** (Breast Cancer Resistance Protein), also known as ABCG2 or MXR (Mitoxantrone Resistance Protein), is sometimes referred as placenta specific antigen since its largest levels are observed at placental barrier. Nevertheless, BCRP is functional in other cell types, among them – brain capillary endothelium and enterocytes, where it restricts absorption of various flavonoids, quinolone antibiotics, antihypertensives, antivirals, statins. Its substrate specificity partially overlaps with P-gp[100,105].

- Several members of ATP Binding Cassette C-subfamily also play an important role in drug efflux. These are denoted **MRP** (Multidrug Resistance associated Protein). A common feature of MRP proteins is their preference to anionic substrates in contrast to P-gp[100].

All described ABC transporters except several MRP isoforms are expressed in brush border membranes of small intestine and in luminal membranes of brain capillary endothelium. According to the current knowledge supported by crystallographic data, these transporters "capture" their ligands from the inner leaflet of the lipid bilayer and extrude them through a cavity formed by α-helices in the transmembrane domain of the protein[106]. Actually, these proteins do not remove xenobiotics from the cells but prevent their cell entry.

SLC superfamily is considerably larger with *ca.* 360 known genes and 30 xenobiotic transporters[102]. Many solute carriers listed in Table 1.3 mediate facilitated diffusion and in principle can transfer their ligands in both directions. At physiological conditions the concentration gradient is usually directed towards net transport from blood to brain (BBB) or lumen to blood (intestine). Mechanisms of carrier-mediated influx of some common nutrients and xenobiotics are summarized below:

- PEPT1 and PEPT2 proteins are proton-coupled carriers for small oligopeptides consisting of at most three residues, as well as certain β-lactam antibiotics (mainly cephalosporins). The proton gradient needed for transport is maintained by Na^+/H^+-ATPase. Both PEPT1 and PEPT2 are expressed in luminal membranes of enterocytes. PEPT2 is also expressed in choroid plexus epithelium and mediates peptide transport across BCSFB but is not functional at BBB[100,107].

- Intestinal mucosa exhibits carrier-mediated influx of bile acids and conjugated steroids. These compounds are absorbed into the enterocytes by Apical Sodium-dependent Bile acid Transporter (ASBT)[100], while diffusion out of the cells across basolateral membrane is aided by Organic Solute Transporters OSTα and OSTβ[108].

- D-Glucose, its derivatives such as glucosamine, and certain other hexose monosaccharides enter the brain *via* facilitative GLUT1 carrier expressed in both luminal and abluminal membranes[46]. In the intestine glucose is pumped from lumen against its concentration gradient by Na^+/glucose co-transporter SGLT and then rapidly leaves the enterocytes by facilitative GLUT2 carrier[75].

- The mechanism of nucleoside influx is similar in the intestine and BBB. It is mediated by active concentrative (CNT) and passive equilibrative (ENT) nucleoside transporters. In both systems ENT proteins are found on the blood-facing side.

- Small (*MW* < 200) acidic drugs may follow the same transport route as endogenous metabolites such as lactate, pyruvate, or acetate. In many cell types, these are taken up by MonoCarboxylic acid Transporter MCT1[109].

- Both endothelial and epithelial cell membranes contain a variety of transporters specific to α-, β-, and γ-amino acids of different size and charge state. In order to keep this overview concise, amino acid transport systems were not included in Figure 1.8 and Table 1.3. For a thorough review of amino acid uptake and efflux systems, one can refer to publications by Ohtsuki & Terasaki[99] and Hawkins et al.[110]

Although some drug-like compounds share the nutrient transport systems, such as amino acid carriers or GLUTs, the majority of drugs subject to carrier-mediated influx interact with multispecific transporters that can be divided into two groups according to the preferred charge state of their substrates:

- Organic Cation Transporters (OCTs) are responsible for higher than expected membrane permeabilities of some basic drugs[111]. However, transport inhibition experiments have shown that certain groups of cationic substrates do not mutually inhibit brain uptake of each other. This means that several different transport systems for cationic compounds exist at BBB. One of these is associated with Organic Cation/Carnitine Transporters (OCTNs)[109]. Recent studies suggest that facilitative choline transporter CTL1 may also be involved in cation uptake at BBB[46].

- Tissue accumulation of acidic compounds, in particular those with steroidal/peptidic backbone, can be mediated by Organic Anion Transporters (OATs) or related Organic Anion Transporting Polypeptides (OATPs), that function as ion exchangers with partially overlapping specificities[42,100].
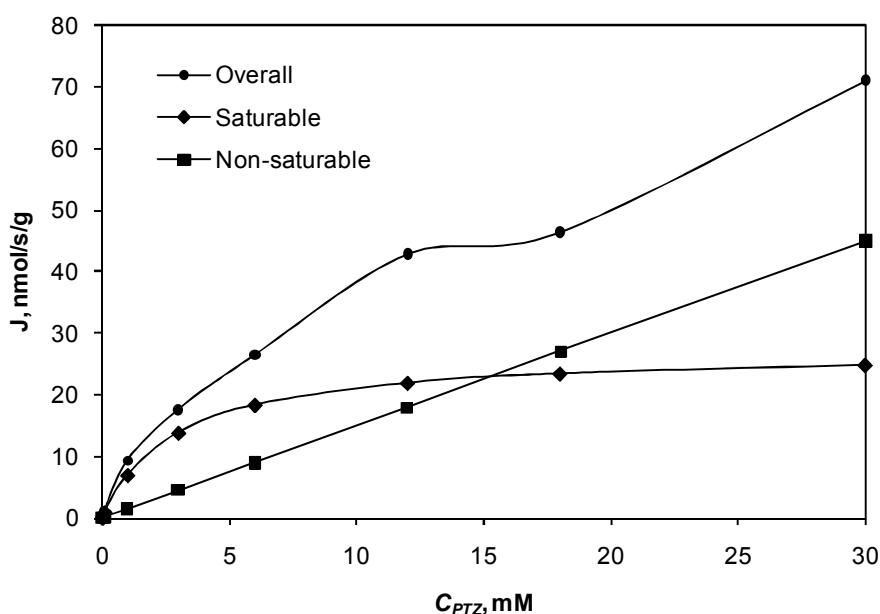
Figure 1.9 Contributions of saturable and non-saturable components to blood-brain flux of pentazocine at different concentrations of unlabeled substrate. Reproduced from Ref.[112]

A key difference from passive diffusion that allows detecting the presence of carrier-mediated component is its concentration dependence. Transporter proteins become saturated with increasing substrate concentration, and carrier-mediated permeation rate asymptotically approaches its maximal value (V_{max}) following typical Michaelis-Menten kinetics. Sugano pointed out that passive and carrier-mediated processes coexist at cellular barriers[74]. All compounds diffuse passively to a certain extent proportional to their lipophilicity. If a compound can also undergo carrier-mediated influx, the dominant transport mechanism depends on several factors including the compound's passive permeability, its concentration, and affinity to the transporter given by K_M. As an example, the concentration dependence of pentazocine flux (J) across BBB is shown in Figure 1.9. At low substrate concentrations (C), carrier-mediated component prevails, but if the compound is lipophilic enough and its concentrations are high, transporter becomes saturated and net transport rate starts increasing linearly. The overall flux can be decomposed into the contributions of particular mechanisms by fitting the data points to Eq. (1.19):

$$J = C \cdot PS + \frac{C \cdot V_{max}}{K_M + C}$$

(1.19)

A number of blood-brain log PS constants for substrates of influx transporters determined in this manner were included in our data set. Prior to such decomposition, the overall permeation rates (J/C) affected by transporter contributions do not reflect true PS values and are of little interest for mechanistic analysis. In case of active efflux, the contribution of passive component can be determined in several ways. Firstly, the uptake rate can be measured in presence of efflux inhibitors or a large excess of unlabeled substrate saturating the transporter. This approach is not always reliable as the efflux carriers may not be fully inhibited, yielding the values not completely free of unwanted effect. A better solution is to conduct experiments in animals lacking the respective transporter gene. P-gp deficient mice (mdr1a(–/–)) are now widely used for this purpose[63,113], whereas recently derived bcrp(–/–)[114] and mrp(–/–)[115] mouse lines enable the investigators to study drug efflux mediated by other members

of ABC superfamily. Notably, the described strategies of eliminating the active efflux contribution are suitable not only for determination of transmembrane permeation rates but also for obtaining unbiased tissue distribution data, *e.g.*, log *BB* values.

Transcytosis. Certain molecules may cross cellular barriers by the means of vesicular transport. Receptor mediated transcytosis (Figure 1.7, (E)) is initiated by ligand binding to specific receptors expressed in the membrane. In the case of adsorption mediated transcytosis (Figure 1.7, (F)), formation of endocytic vesicles is triggered in a non-specific manner by adsorption of cationic proteins, such as histones, on cell surface[46,47]. Both mechanisms are typical for macromolecules that cannot enter the cells by other routes and, therefore, will not be considered in the current study.

1.6. Overview of earlier QSAR studies

1.6.1. Distribution

Plasma protein binding. QSAR approaches for predicting plasma protein binding strength (Table 1.4, top and middle) may be separated into two broad categories according to the considered endpoint: (i) models that calculate the affinity constant to human serum albumin (log K_A^{HSA}); (ii) models for estimating the overall fraction unbound in plasma ($f_{u,pl}$). The former endpoint is much easier to predict as it involves drug interactions with a single protein, even though albumin has multiple binding sites. However, the situation is complicated by very limited literature data. Most new studies[116-118] reanalyze the same data set containing only 94 compounds published by Colmenarejo back in 2003[119]. Given the complexity of modeling methods and used descriptors (topological, quantum chemical, etc.), the test set of 10 molecules provided in Ref.[119] is insufficient for drawing any definitive conclusions about the performance of these models. A somewhat larger data set of 151 log K_A^{HSA} constants was analyzed by Kratochwil et al.[16]. For 76 of those molecules, the authors also determined octanol/water log D at pH 7.4 and correlated log D *vs.*

$\log K_A^{HSA}$. Lipophilicity alone explained half of the variation in HSA affinities of acidic drugs, while no correlation was observed for bases.

In the context of the current study, it is necessary to know the overall extent of binding to all plasma proteins, not to albumin alone. Although accurate prediction of plasma unbound fraction is an immensely difficult task, a useful insight can be obtained from simple physicochemical models. For example, Yamazaki & Kanaoka[12] related $f_{u,pl}$ to lipophilicity using NLS method:

$$\frac{1}{f_{u,pl}} = k_1 e^{\log D_{o/w}} + k_2 \tag{1.20}$$

Two different models were built for acids and non-acids (bases and neutrals) reflecting preferential binding of these drug classes to HSA and AAG, respectively. Both models allowed for a good fit to experimental values, but in case of acids, this could only be achieved after restricting the data set to compounds possessing a typical albumin site I/II pharmacophore (hydrophobic/aromatic center + H-bond donor at 4-5 Å distance). Unfortunately, simple relationship between HSA/AAG binding and lipophilicity did not hold for zwitterionic compounds.

Several important conclusions have been reached in two newer publications by Rodgers et al.[120] and Weaver & Gleeson[121]. Both studies focused on temporal evaluation of model predictive power, *i.e.*, how the models perform on new data sets acquired over time. In Ref.[120] the authors employed a two-step approach involving a global PLS model based on more than 3000 compounds and automatic correction libraries used to compensate for the bias introduced by PLS predictions by comparing predicted and observed values for similar compounds. When model development was finished, 20 test set sets were collected, each representing in-house data measured over a period of one month. Analysis of these monthly sets, ranging in size from 11 to 83 compounds, revealed a statistically significant difference in *RMSE* of predictions made by PLS model alone and after library-based correction, the latter being consistently smaller. *RMSE* of model predictions for a final test set containing 512 compounds steadily decreased from 0.59 to 0.47 when the monthly sets were added

to the correction library in a stepwise manner. The approach employed in Ref.[120] is very similar to GALAS modeling methodology developed in our group[4,5]. Since the results presented above demonstrate the effectiveness of automatic correction libraries, GALAS was the method of choice for building plasma protein binding models in the current study.

In Ref.[121] the authors validated their PLS models using three external test sets. The first two sets consisting of 385 and 132 compounds also represented in-house data from different time periods (< 6 months and 6-12 months after the initial study, respectively). One can clearly see from Table 1.4 that the predictive ability of the model gradually decreased as in-house projects moved to different regions of chemical space over time. Even worse results were obtained for additional 324 compounds collected from the literature. The authors highlighted the importance of model applicability domain assessment for evaluating the accuracy of predictions and demonstrated the consistence between the observed prediction errors and the compounds' distance to the model estimated on the basis of molecular fingerprints.

Volume of distribution. When previous QSAR studies for predicting volume of distribution are considered (Table 1.4, bottom), a key reference point is the physicochemical analysis of neutral and basic drugs performed by Lombardo and coworkers[122,123]. They converted experimental V_{SS} data to tissue unbound fractions ($f_{u,ti}$) according to Øie-Tozer equation (see Section 3.3) and fitted the resulting values to a log-linear function of octanol/water partitioning coefficient, ionized form fraction, and fraction unbound in plasma by the means of principal component regression (PCR). The real value of these publications is in highlighting the importance of decomposing volume of distribution to the individual contributions of plasma and tissue binding and demonstrating the utility of physiological Øie-Tozer equation for accomplishing this task. The proposed approach allowed predicting V_{SS} slightly outside of the 2-fold error margin, but their model is applicable only to neutral and basic compounds.

Table 1.4 A summary of selected models for predicting plasma protein binding and volume of distribution.

References	Method	Descriptors	N^a	Statistical performance R^2	RMSE
log K_A^{HSA} models					
Kratochwil et al.[16]	LR	log $D_{o/w}$	76	0.49 (acids)	n.r.[b]
				0.09 (bases)	n.r.
	PLS	Topological pharmacophores	151	0.72 ($Q^2 = 0.48$)	0.62
Wichmann et al.[117]	MLR	Quantum-chemical	84	0.66 ($Q^2 = 0.61$)	0.33
			$10^{(t)}$	0.80	0.32
%*PPB* models					
Yamazaki & Kanaoka[12]	NLS	log $D_{o/w}$ (neutrals and bases)	84	0.803	*MAE*=10.4%
		log $P_{o/w}$ (acids with pharmacophores)	44	0.786	*MAE*=5.5%
		2 models combined	$20^{(t)}$	0.830	*MAE*=11.6%
Rodgers et al.[120]	PLS + local correction	2D, 3D, and electronic	$11\text{-}83^{(e)}$	n.r.	0.44-0.66
			$512^{(e)}$	n.r.	0.47-0.59
Weaver & Gleeson[121]	PLS	1D and 2D descriptors	685	0.56	0.55
			$210^{(t)}$	0.58	0.54
			$385^{(e)}$	0.51	0.57
			$132^{(e)}$	0.44	0.53
			$324^{(e)}$	0.34	1.05
V_{SS} models					
Lombardo et al.[123]	PCR + Øie-Tozer equation	log $D_{o/w}$, base pK_a, $f_{u,pl}$	120	0.87 ($Q^2 = 0.85$)	0.37 *AFE* = 2.08
			$18^{(t)}$	n.r.	*AFE* = 2.26
Hollosy et al.[124]	MLR	log k_{IAM}, log K_A^{HSA}, acid & base pK_a	149	0.82	0.29
			$30^{(t)}$	0.72	0.38
Rodgers & Rowland[125]	Tissue composition	$f_{u,pl}$, K_{pu}^{RBC}	140	0.74	0.31 *AFE* = 1.80
Sui et al.[126]	MLR + Øie-Tozer equation	log k_{IAM}, $f_{u,pl}$, acid & base pKa	97	0.83 ($Q^2 = 0.81$)	0.33 *AFE* = 1.72
			$24^{(t)}$	n.r.	*AFE* = 1.39
Berellini et al.[127]	PLS, RF	MOE and Volsurf+ descriptor sets	669	n.r.	*AFE* = 1.8
			$29^{(t)}$	n.r.	*AFE* = 1.9
Poulin et al.[128,129]	Tissue composition	$f_{u,pl}$, K_{pu}^{RBC}	45	0.85	*AFE* = 1.57
			45	0.86	*AFE* = 1.38

[a] N – number of compounds. Superscripts denote data sets used for validation purposes: (t) – test set, (e) – external validation set.
[b] n.r. – statistic not reported.

Recently, the method of Lombardo was modified by Sui et al.[126]. The authors replaced $\log D_{o/w}$ with $\log k_{IAM}$ (partitioning ratio between water and immobilized artificial membranes) and included all classes of electrolytes in the analysis. The reported statistical characteristics are superior to those achieved by Lombardo, but the model relies on phospholipid binding strength characterized by k_{IAM}. These values are not readily available and need to be determined experimentally. Another questionable aspect is the description of ionization. In this model the ionization state of drugs is expressed by two descriptors – $\log f_o$ and $\log f_i$ (logarithms of neutral and ionized form fractions, respectively). Hence, all ionic species are represented by a single parameter $\log f_i$, and there is no discrimination between acids, bases, and zwitterions.

In contrast to blood-brain partitioning, which is often modeled without any regard to its relationship with drug binding strength in different compartments, almost all authors publishing in the field of volume of distribution QSAR acknowledge the influence of plasma protein binding. Consequently, the dependent value is usually recalculated to yield either $\log f_{u,ti}$ (as in Refs.[122,123,126]) or the so called "unbound volume of distribution": $V_{u,SS} = V_{SS}/f_{u,pl}$. Yet, Hollosy et al.[124] chose a minimalistic approach by correlating $\log V_{SS}$ against $\log k_{IAM}$ and $\log K_A^{HSA}$, thought to represent tissue and plasma binding, respectively. While attractive by its simplicity, this model does not reflect true physiological relationship between parameters and only takes into account drug binding to albumin in plasma. Moreover, existing methods for predicting albumin affinity are not reliable enough, so that its use as a descriptor implies the necessity of *in vitro* measurements.

Berellini et al.[127] compiled a large dataset containing experimental V_{SS} values for 669 compounds and tried building predictive models based on a formal statistical approach. Several models were generated by the means of PLS and Random Forest (RF) methods using a set of 95 descriptors. Individual models tended to perform better for certain drug classes, while the consensus model yielded slightly smaller than 2-fold average error for both training and test set. This is not a significant improvement over simple physicochemical equations

derived by Lombardo et al., and given the lack of interpretability of such complex models, it can be concluded that mechanistic approach is preferable for a QSAR study of volume of distribution.

Quite different route was taken by Rodgers & Rowland[125]. They described the steady-state unbound volume of distribution by the multiple tissue affinity method:

$$V_{u,SS} = \frac{V_P}{f_{u,pl}} + \sum_i V_{T,i} K_{p,i}^u \qquad (1.21)$$

Here V_P is the plasma volume, $V_{T,i}$ are the volumes of particular tissues, and $K_{p,i}^u$ – unbound tissue to plasma partitioning coefficients. Volumes and composition of different tissues were taken from the literature, and $K_{p,i}^u$ were calculated by deterministic equations describing drug dissolution into tissue water, lipid/water partitioning, electrostatic interactions with acidic phospholipids (approximated by blood to plasma partitioning), and extracellular protein binding. Initially, the analysis involved 13 tissues (adipose, bone, brain, gut, heart, kidney, liver, lung, muscle, pancreas, skin, spleen, and thymus), but in their second approach the authors investigated the possibility to simplify the model by substituting all specific tissue affinity data with the affinities to muscle, which is the most abundant tissue in human body. Interestingly, the two approaches produced virtually identical $V_{u,SS}$ prediction errors indicating that volume of distribution can be successfully modeled using the "generalized" tissue to plasma partitioning concept, and if different tissues are not considered separately, this is not too much of an oversimplification.

Similar models were derived by Poulin & Theil[128,129], who were the first to suggest the idea of *a priori* prediction of tissue/plasma partitioning from tissue composition data back in 2000[130]. In their new works, instead of using fully deterministic expressions for $K_{p,i}^u$, they relied on empirical correlation between $K_{p,i}^u$ and *in vitro* red blood cell (RBC) binding strength. This idea is based on the assumption that both RBC and tissue partitioning of basic drugs is governed by interactions with acidic phospholipids (primarily phosphatidylserine). Initially[128], the modified multiple tissue affinity approach yielded

slightly lower *AFE* of prediction compared to Ref.[125], though in a follow-up study[129], the model was improved by addition of a separate regression equation for certain drug classes (local anesthetics and cardiovascular drugs) that were overpredicted by the original model. Unfortunately, their latest model is still applicable only for basic molecules and requires knowledge of RBC affinities. Despite all recent progress in volume of distribution modeling, which is undoubtedly more significant than in other QSAR fields overviewed here, there are still no published models that would use a large data set covering a wide range of drug-like structures, and at same time maintain mechanistic interpretability and rely on simple, readily available physicochemical parameters.

1.6.2. Blood-brain transport

Numerous attempts to predict drug transport across BBB using *in silico* methods have been reported up to date. However, there is sharp contrast between the two commonly used quantitative characteristics – the overwhelming majority of earlier studies focus on modeling blood/brain partitioning ratio expressed as log *BB*, while the number of studies predicting log *PS* is very limited. This can probably be explained by differences in data availability. As outlined in Section 1.3.3, log *PS* experiments require a more complex setup and involve a surgical procedure. For this reason, there are much less published log *PS* values compared to log *BB*, and the latter parameter is usually preferred for modeling.

log *PS*. The top part of Table 1.5 summarizes the most notable QSAR models dealing with log *PS* prediction. The studied data sets rarely contained more than 20–30 compounds, and this lack of high-quality data restricted the possibility of understanding what factors govern BBB permeability and led to heavily oversimplified considerations. The first model of this kind was published over 30 years ago by Levin[131], who related log P_c (a close analog of log *PS*) to a simple linear function of hydrophobicity and molecular size. The overall effect was represented by a composite descriptor: $\log (P \cdot MW^{-1/2}) = \log P - \frac{1}{2} \log MW$.

Several other authors[132-134] proposed similar equations, whereas Bodor and Buchwald[54] obtained a reasonable correlation between experimental $\log P_c$ and $\log P_{o/w}$ without any correction for molecular size (for electrolytes $\log D_{o/w}$ measured around physiological pH was used instead of $\log P_{o/w}$ if available). Luco and Marchevsky[135] revised previously published QSAR studies and reanalyzed the data sets provided therein using a similar approach. All these models mostly focus on improving Levin's results rather than achieving a better understanding of BBB permeation mechanisms.

M.H. Abraham's group employed a different approach based on Abraham's solvation equation[136]. Solvation parameters can describe various H-bonding effects more accurately than octanol/water $\log P$, yet this model disregards ion partitioning effects and is applicable only to non-electrolytes. In a later publication[53], the data set was extended to 30 compounds, although the modeling method remained the same. Finally, in his newest model[137], the author incorporated additional descriptors J^+ and J^- related to ion-solvent interactions and obtained a much more useful model that is applicable for ionizable molecules.

Liu et al.[138] replaced $\log P_{o/w}$ and solvation parameters with pH-dependent $\log D_{o/w}$ and a combination of surface descriptors – vsa_base (van der Waals surface area of basic atoms) and TPSA. The use of $\log D_{o/w}$ implies an assumption that ionic species have the same affinity to phospholipid membrane as to octanol, which is quite disputable (see Section 3.1).

Several notable studies followed our publication describing the $\log PS$ model developed in the current study[139]. One of these is the latest study by Abraham[137] discussed above, while another novel model was proposed by Dagenais et al.[140] The authors performed a thorough analysis of available literature data and determined brain permeation rates in P-gp knockout mice for additional 19 compounds. Their initial data set contained 507 data points, but only a subset of 182 $\log PS$ values (significantly overlapping with our data) was considered to represent passive diffusion across BBB.

Table 1.5 A summary of published studies for predicting quantitative BBB transport parameters.

References	Method	Descriptors	N^a	Statistical performance	
				R^2	$RMSE$
log *PS* models					
Levin[131]	LR	$\log (P_{o/w} \cdot MW^{-1/2})$	22	0.83	n.r.[b]
Abraham et al.[53,136,137]	MLR	Solvation parameters (A, B, E, S, V_x)	18	0.95	0.48
			30	0.87	0.52
		Same as above $+ J^+, J^-$	88	0.81	0.53
Bodor and Buchwald[54]	LR	$\log P_{o/w}$ (log $D_{o/w}$ in some cases)	58	0.90	0.62
Liu et al.[138]	MLR	$\log D_{o/w}$, *PSA*, vsa_base	23	0.74	0.50
Luco & Marchevsky[135]	MLR	$\log P_{o/w}$, log *MW*, MW^2	7-37	0.80-0.96	n.r.
Dagenais et al.[140]	MLR	$\log P_o^{PAMPA}$, solvation parameters (A, B)	$130^{(t)}$	0.80	0.51
			$52^{(e)}$	0.59	0.67
log *BB* models					
Young[64]	MLR	$\Delta \log P$	20	0.69	n.r.
Garg & Verma[141]	ANN	cLogP, *MW*, *TPSA*, N_{HD}, N_{HA}, N_{RB}, p($S_{P\text{-}gp}$)	132	0.82	0.30
			50	0.80	0.32
Abraham et al.[142,143]	MLR	$\log P$, $\Delta \log P$, solvation parameters (A, B, E, S, V_x)	302	0.75	0.30
			160	0.57	0.30
Fu et al.[144]	MLR	*MW*, MW^2, No. of polar atoms	78	0.74 ($Q^2 = 0.71$)	0.37
			$25^{(t)}$	0.50	0.53
Zhang et al.[145]	kNN, SVM	MolConnZ, 2D MOE, & DRAGON descriptors	144	0.91	*MAE*=0.21
			$15^{(t)}$	0.80	*MAE*=0.29
Kortagere et al.[146]	SVM	$\log P_{o/w}$, log *S*, *TPSA*, N_{RB}, etc.	78	0.70	n.r.
			$100^{(t)}$	0.65	n.r.
Chen et al.[147]	ANN	$\log P$, *PSA*, N_{HD}, N_{HA}, N_{RB}, molar volume, %*PPB*, p($S_{P\text{-}gp}$)	120	0.67	0.25
			$21^{(t)}$	0.52	0.22
Fan et al.[148]	GFA	*PSA*, No. of aromatic rings, electrotopological indices, QikProp descriptors	193	0.74 ($Q^2 = 0.72$)	n.r.
			$81^{(t)}$	0.62	n.r.
			$147^{(e)}$	0.65	n.r.

[a] N – number of compounds. Superscripts denote data sets used for validation purposes: (t) – test set, (e) – external validation set.
[b] n.r. – statistic not reported.

The model was derived using *in combo* approach based on *in vitro* PAM-PA permeability and *in silico* solvation descriptors to account for various H-bonding effects. In order to utilize the full potential of Abraham solvation parameters, that only work well for non-ionized molecules, the authors converted original permeability data to "intrinsic" values of neutral form according to pH-partition hypothesis. Overall, this approach yielded a very good model, but its use in computer-aided drug design is limited by its reliance on PAMPA permeabilities that need to be determined experimentally.

log *BB*. Due to a very large of number of publications dealing with log *BB* predictions, Table 1.5 lists only a few notable older articles and the studies published in the last few years that are not yet covered by recent reviews. For a more detailed discussion of earlier work one can refer to Refs.[142,145,149]. Proposed models vary significantly in terms of methodological approaches ranging from simple regression equations that describe log *BB* as a linear combination of selected physicochemical properties to complex models using sophisticated statistical techniques and large pools of theoretical descriptors. Yet, their most common drawbacks are basically the same as those of BBB permeability models. First, the number of data points used to parameterize the models is usually quite small. For several decades, starting from the classical physico-chemical model of Young[64] till the publications by Garg & Verma[141] and Abraham[142], most authors had been using less than 100 compounds. Second, little effort is made to ensure that only high-quality data are used for modeling. A common approach is to collect the data from several previously published articles without raising any concerns regarding the possible involvement of different BBB transport mechanisms or performing any other verification procedures. Third, log *BB* is usually fitted to selected descriptors "as is" disregarding the complex nature of this parameter and its relationship with plasma protein binding. Only a few of the presented modeling approaches account for fraction unbound in plasma[147] and probability of P-gp mediated efflux[141,147], albeit by simply including these properties in the descriptor matrix. Finally, some models are likely to be over-fitted because reported prediction errors (expressed as

RMSE or *MAE*) are smaller than the experimental log *BB* determination error, which is at least 0.3 log units[142,150].

In general, if the performance of the models summarized in Table 1.5 is considered, it can be noted that most authors report very good prediction errors but the overall correlation between experimental and observed log *BB* varies considerably with R^2 values spanning the range from about 0.5 to 0.9. Better R^2 is usually obtained when the data set used for modeling includes a significant amount of very large or very small log *BB* values. Such extreme data points may not adequately reflect passive equilibration across BBB as they might be affected by carrier-mediated transport or other issues. Finally, the majority of above-mentioned models except the most recent work by Fan and coworkers[148] have not been tested on an external validation set, and their actual predictive power remains questionable.

Table 1.6 A summary of earlier classification models of BBB penetration.

References	Method	Descriptors	N^a	Statistical performance	
				Accuracy	*AUC* or U^b
Adenot & Lahana[151]	LDA, PLS-DA	PhysChem, surface areas, topological	1605 $82^{(t)}$	97% 91%	n.r.[c] n.r.
Zhao et al.[152]	BPLS	$N_{HD} + N_{HA}$	1593 $397^{(t)}$	91.3% 78.8%	n.r. n.r.
		Chain-based fragments	1593 $397^{(t)}$	97.0% 74.6%	n.r. n.r.
Vilar et al.[153]	LDA	2D MOE (log *P*, TPSA, no. of acidic and basic atoms)	307 307	80.4% 78.4%	$U = 0.70$ $U = 0.84$
		log *BB* ≥ 0.3	1457	n.r.	$AUC = 0.95$
		log *BB* < -1	1457	85.6%	$AUC = 0.97$
Fan et al.[148]	LDA	log *BB* cut-off	1403	82%	$AUC = 0.9$

[a] *N* – number of compounds. Superscript (t) denotes test set compounds.
[b] In addition to accuracy of classification, some authors report other statistical characteristics: *AUC* – Area Under ROC Curve (see text for details), *U* – Wilk's statistic. Both parameters range from 0 to 1, but an ideal fit is characterized by *AUC* = 1; *U* = 0.
[c] n.r. – statistic not reported.

Classification models. Some modeling studies do not raise the objective to predict quantitative BBB penetration parameters but instead are dedicated to the development of classification models. The binary endpoint can be defined as BBB+/– or CNS+/–, indicating what particular aspect – crossing the BBB or CNS activity – the authors focus on. Predicting drugs' activity against CNS targets is out of the scope of the current study, and only those approaches that deal with the ability of drug compounds to cross BBB will be considered here. Notably, many publications use the same data set of about 1500 drugs from WDI database compiled by Adenot & Lahana[151]. It is a diverse and well-constructed data set attempting to overcome the difficulties with unambiguous assignment of BBB+/– categories from therapeutic activity data alone. At the same time, it is the largest source of qualitative BBB penetration data published up to date. This data set was also used in the current study for evaluation of the presented approach for classifying drugs by CNS access.

Several publications analyzing the data from Ref.[151] are listed in Table 1.6. The actual number of compounds used in these studies slightly varies due to different cleaning and filtering procedures performed by the authors. Adenot & Lahana themselves employed LDA and PLS-Discriminant Analysis methods to select relevant descriptors from a pool of 67 parameters and subsequently model the data. While their best model produced an almost ideal fit to the training set, more realistic performance numbers were obtained for external validation set compounds (91% overall accuracy).

Several interesting conclusions followed from the work by Zhao et al.[152] The authors built a number of binomial PLS models using either chain-based fragmental descriptors (linear, branched, or cyclic atom chains of the specified length) or a combination of hydrogen bonding parameters. The relatively large number of chain fragments produced an apparently over-fitted model that showed a significant drop in performance when going from training to the validation set. Interestingly enough, more than 90% of training set data could be accurately classified by simple H-bond donor and acceptor counts while giving better external predictivity compared to the fragmental model. These results

demonstrate that the compounds from Ref.[151] constitute a rather "modeling-friendly" data set. The two drug classes are reasonably well discriminated even by quite simple classifiers, whereas increasing model complexity does not necessarily improve the predictive power.

In a few recent publications, the authors try to assign quantitative BBB+/– categories on the basis of a certain log *BB* threshold. Vilar et al.[153] collected log *BB* data for 307 compounds and derived two LDA models that estimate whether the drugs have log $BB \geq 0.3$ or log $BB < -1$. The models were parameterized using physicochemical descriptors calculated by MOE software. Running predictions against Adenot & Lahana data set revealed that lower log *BB* threshold is more suitable for classification. If the compounds predicted to have log $BB > -1$ are considered BBB positive, about 86% of all data points are classified correctly, albeit with a high percentage of false negatives. The models for both thresholds demonstrated very high predictive power when evaluated using ROC (Receiver Operating Characteristic) curves – areas under the curve (*AUC*) were close to 1, which corresponds to an almost ideal fit.

Fan et al.[148] used a slightly different approach. They developed a model predicting quantitative log *BB* values (listed in Table 1.5) and then used calculated log *BB* as a linear classifier function to assign BBB penetration categories. The optimal classifier with log *BB* cut-off set at -0.52 was able to achieve 82% classification accuracy and *AUC* of 0.9 on the validation set compounds.

Despite the impressive ROC profiles, log *BB*-based models do not introduce a significant improvement in separation of BBB+ and BBB– classes over the simple classifier relying on H-bonding potential presented by Zhao et al. This observation supports the statement that log *BB* alone does not suffice to deduce if a compound would penetrate the brain effectively. In the current study, classification of drugs by CNS access was performed using a similar approach, yet involving both quantitative parameters discussed here – log *BB* and log *PS*, thus accounting not only for the extent but also for the rate of brain penetration.

1.6.3. Intestinal permeability and absorption

A variety of approaches have been employed for evaluating gastrointestinal absorption of drugs. As mentioned in Section 1.4.3, several cell-based assays have been developed to screen for permeability, and Caco-2 cell line is one of the most widely used[83]. Although there is an approximate sigmoidal relationship between flux through such cells and human intestinal absorption, predictions on the steep part of the curve are often not reliable. Alternative experimental approaches such as permeation through artificial membranes (PAMPA)[68] and empirical *in silico* models have also been derived to estimate passive diffusion-driven permeability. Discussions of the earlier QSAR studies can be found in several reviews[154,155], and similarly to the previous subsection, only the most recent QSAR studies dealing with prediction of intestinal absorption are briefly summarized in Table 1.7.

In general, absorption models suffer from the same major issues as the models for predicting blood-brain transport. Some of the problems are further exaggerated by the specifics of the absorption system. It is well-known that predictivity of all data driven models heavily relies on the quality of the training set used for model generation. This is particularly true for *%HIA* models where sigmoid relationship between fraction absorbed and permeability is very sensitive to small changes in compounds' properties on the steep part of the curve. Another important aspect to consider is the involvement of multiple absorption mechanisms. Even if actively transported compounds are omitted from analysis, passive diffusion across intestinal membranes may take place by transcellular and paracellular pathways. In this situation, only mechanistic models that bring together physicochemical representation of the analyzed molecules and mathematical description of the fundamental biophysical processes can be applicable for a wide range of chemical structure classes[156].

Sugano and coworkers have taken such mechanistic route towards the absorption prediction and combined a theoretical model of paracellular absorption pathway with transcellular permeation constants, either determined in a PAMPA experiment[157] or estimated from octanol-water partition coeffi-

cients[158]. However, little progress has been made since then and most recent *in silico* prediction models[159-163] do not differentiate between absorption mechanisms. They rather adopt a formal statistical approach focusing on automated selection of „best" descriptors from a large pool and evaluating which statistical method yields the best predictive power judged by purely numerical criteria. The lack of biological interpretation becomes particularly evident when complex molecular descriptors are used. These are often difficult to understand by synthetic chemists, and as a consequence – difficult to translate into drug design. On the contrary, use of simple physicochemical parameters would be helpful for adoption of recently advocated *in combo* concept where predictions are improved by combining results from *in vitro* screens with *in silico* tools[7].

The current situation around predictive models of *in vitro* permeability in Caco-2 cells in even more complicated. Caco-2 transport experiments can be conducted at widely varying conditions, particularly with respect to pH, that enables determination of permeability dependence on ionization. Still, in most QSAR articles[164-167], the used data sets are provided as simple lists of $\log P_e$ values without any reference to pH at which they were measured or other experimental details. This raises the question whether the authors specifically selected the data corresponding to the same pH (*i.e.*, 7.4) or completely disregarded this aspect and used the values from all sources without verification. Unsurprisingly, the predictive power of such models is generally poor, and despite obtaining good statistical parameters on training set data, external validation reveals their inherent limits. For example, very poor correlation between observed and predicted $\log P_e$ was observed in Ref.[165] with external R^2 being as low as 0.16. In Ref.[167], the model performance steadily decreased when going from training set and cross-validation results (Q^2) to internal, and finally external test set, clearly indicating the over-fitting issue.

Table 1.7 A summary of recent models for predicting the extent of human intestinal absorption and permeability in Caco-2 cell monolayers.

References	Method	Descriptors	N^a	Statistical performance	
				R^2	$RMSE^b$
%*HIA* models					
Iyer et al.[159]	MI-QSAR	ClogP, *PSA*, other dissolution and	188	0.73 ($Q^2 = 0.67$)	n.r.[c]
		solvation descriptors, solute-	117	0.78 ($Q^2 = 0.74$)	n.r.
		membrane interaction descriptors	106	0.82 ($Q^2 = 0.79$)	n.r.
			21[(t)]	0.5	0.39 log unit
Deconinck et al.[160,161]	MLR, PLS, PCR	Various theoretical and chromatographic descriptors	67	0.93	7.2%
			140	0.90	9.9%
Hou et al.[162]	GFA	log $D_{o/w}$ at pH = 6.5, *TPSA*, N_{HD}, no. of	455	0.84	7.2%
		"rule of 5" violations	98[(t)]	0.90	9.5%
Talevi et al.[163]	MLR	0D-3D Dragon theoretical descriptors	90	0.80	0.18 log unit
			30[(t)]	0.66	0.21 log unit
Caco-2 permeability models					
Di Fenza et al.[165]	GA-NN	Volsurf descriptors, ACD/pK_a	79	$Q^2 = 0.52$	n.r.
			50[(t)]	0.16	n.r.
		Volsurf descriptors	155	$Q^2 = 0.57$	n.r.
			50[(t)]	0.37	n.r.
Castillo-Garit et al.[166]	MLR	Non-stochastic linear indices	77	0.67 ($Q^2 = 0.60$)	0.47
			23[(t)]	0.61	*MAE* = 0.45
		Stochastic linear indices	77	0.63 ($Q^2 = 0.56$)	0.50
			23[(t)]	0.64	*MAE* = 0.39
Gozalbes et al.[167]	MLR	MLogP, substructures defined by	97	0.82 ($Q^2 = 0.69$)	0.44
		MACCS keys, theoretical descriptors	41[(t)]	0.52	0.66
			21[(e)]	0.24	0.75

---

2. DATA & METHODS

2.1. Experimental data

2.1.1. Plasma protein binding and volume of distribution

Experimental data on the extent of binding to plasma proteins and volume of distribution in human body were collected from reference pharmacokinetic tabulations, drug prescription information, and original research articles dealing with determination of these parameters. The main data sources were well-known pharmacokinetic books[168-172]. The key research articles providing protein binding data were several studies focusing on chromatographic determination of drugs' affinity to HSA[25,26] and the work of Hall et al.[173] where %PPB values were reported for a range of β-lactam antibiotics. The starting point for the compilation of V_d data was the series of publications by Lombardo et al.[122,123] discussed in Section 1.6.1. When a range of values was provided for a compound or different values were published in different sources, arithmetic mean was recorded in the database in most cases. Several such entries were discarded due to unacceptably large variability in reported %PPB or V_d values.

It was pointed out in Section 1.2 that V_d may be given as steady-state distribution volume V_{SS}, terminal phase distribution volume V_z, or as V_z/F (V_{SS}/F) determined after oral administration. Since V_{SS} values represent the highest quality data, these were preferred wherever available. Furthermore, the physiological model of drug distribution employed in the current study (see Section 3.3) is strictly applicable only to V_{SS}. Consequently, V_z data were evaluated more critically, though most values that did not contradict general physico-chemical trends were kept in analysis preferring larger diversity of training set compounds to obtaining slightly better statistical characteristics of the model. V_z/F is potentially the most problematic type of data – these values were included in the database only if bioavailability was known or the compounds could be safely assumed to have complete absorption and had no known issues such as P-gp efflux or extensive first-pass metabolism.

Table 2.1 Plasma protein binding and volume of distribution data set sizes.

Data set	Property	
	%PPB	*V_d*
Training set	1162	346
Internal test set	291	150
External validation sets	n/a	352
		90

Overall, the compiled *%PPB* data set consisted of 1453 compounds, that were split into training and test sets at 80% to 20% ratio. *V_d* data set contained 848 molecules, 496 of which overlapped with *%PPB* database (Table 2.1). In order to avoid uncertainty in plasma free fraction, which is used as a descriptor in *V_{SS}* model (see Section 3.3), only this subset, that could be supplemented with experimental *$f_{u,pl}$*, was utilized for modeling. The remaining 352 *V_d* values were reserved for external validation purposes. The described approach when experimental *$f_{u,pl}$* data are used for model development but only predicted *$f_{u,pl}$* are available for the validation set allows evaluating the quality of both *$f_{u,pl}$* and *V_{SS}* predictors simultaneously.

Another validation set was collected from a recent publication by Obach et al.[174]. The data analyzed therein included 670 drug compounds with measured *V_{SS}* values, and a subset of 90 molecules was not present in our database. For some of these molecules experimental *%PPB* data were also given, while for remaining compounds *$f_{u,pl}$* values were calculated using our predictive model, as was the case with the first validation set. The respective 90 molecules, provided in the Appendix (Table A1), were used as an independent test set to demonstrate that the model parameterized on *V_d* data determined by different methods would perform well on high-quality *V_{SS}* data.

2.1.2. Blood-brain transport

Three different endpoints were considered in QSAR analysis of blood to brain transport of drugs: two quantitative characteristics (rate and extent of BBB penetration) listed in Table 2.2 and qualitative CNS access categories (BBB+ or BBB−).

BBB permeability. Quantitative BBB permeation rates expressed as permeability-surface area products (log *PS*) were compiled from various experimental works (mainly one compound per article) and earlier QSAR studies. Only those QSARs where the experimental data were determined in the same laboratory but not obtained from secondary sources were selected for this purpose. Among the most notable sources of data was one of the first QSAR studies of blood-brain permeability published by Levin[131] and a few more recent publications by M. H. Abraham's group[53,136]. The considered experimental works used one of the methods presented in Section 1.3.3 – intravenous administration[55], brain uptake index[55,60], or *in situ* perfusion[55,61]. Historically, most studies were conducted with rats, but a significant amount of high quality data on mouse brain uptake were published since Dagenais et al. adapted *in situ* perfusion method for use with mice[63]. Although rats and mice are two congeneric rodent species and their blood-brain barriers could be expected to have similar permeabilities to chemicals, it was necessary to evaluate whether log *PS* values determined in rats and mice correlate well and could be used together in a single model. Analysis of 25 drugs with experimental log *PS* determined in both species revealed that permeabilities were indeed almost identical with a constant shift in log *PS* reflecting the interspecies difference in brain capillary surface area. The following equation was used to bring the data on a single scale (please refer to Ref.[139] for more details):

$$\log PS \,(\text{rat}) = \log PS \,(\text{mouse}) + 0.4 \tag{2.1}$$

Only the values reflecting passive diffusion across BBB were included in the analysis. The final data set after all verification procedures contained log *PS* values for 178 compounds. These were used for model development and internal testing. An additional set of measured BBB permeability coefficients for 50 marketed drugs taken from a recent publication of Summerfield and coworkers[175] was reserved for external validation purposes. The internal test set is available in Ref.[139], while the external validation set compounds are listed in the Appendix (Table A2).

Table 2.2 Data set sizes for quantitative properties characterizing drug transport across blood-brain barrier.

log PS		log BB	
Data set	N	Data set	N
Training set	125	Training set	329
Internal test set	53	Internal test set	141
External test set[175]	43	p$f_{u,br}$ validation set (1)[176]	30
		p$f_{u,br}$ validation set (2)[177]	21

Blood-brain distribution. Similarly to log PS, quantitative blood-brain distribution data (log BB) were collected from original experimental articles and earlier modeling works (notably Refs.[141,142]), the latter being rechecked in the original source wherever possible. The data were verified to ensure that they represent passive equilibration across BBB and are not biased by other issues (*e.g.*, compound not reaching steady-state). The possible ways to obtain passive diffusion data for actively transported compounds were briefly over-viewed in Section 1.5, while a detailed discussion of various data analysis aspects can be found in our earlier publication[112]. The mentioned procedures led to a final set of 470 compounds with quantitative log BB values. After that, brain/blood partitioning ratios were corrected for the extent of protein binding in plasma yielding fractions unbound in brain tissue (designated $f_{u,br}$). As explained in Section 3.3, the latter parameter in its logarithmic form (p$f_{u,br}$) is preferable for modeling. The resulting values were randomly split into training (70%) and internal test set (30%).

The performance of the obtained model was additionally tested on two external data sets representing another type of experimental data. The external sets consisted of $f_{u,br}$ values directly measured in rodent brain homogenates by equilibrium dialysis technique rather than recalculated from log BB, as was the case with the compounds used in modeling. The first set extracted from a recent publication by Kalvass and coworkers[176] contained unbound fractions in mouse brain for 34 compounds. The second validation set of 21 $f_{u,br}$ values determined in rats was taken from the work of Summerfield et al.[177]. The external validation sets were preprocessed prior to the analysis, and four compounds

were removed from Kalvass set (digoxin, doxorubicin, ivermectin 1a, and pac-litaxel). The reason for their exclusion was large discrepancy between $f_{u,pl}/f_{u,br}$ ratios calculated from the data provided in Ref.[176] and $\log BB$ values in mdr1a(–/–) mice reported in other studies, even despite the absence of P-gp effect in both assays. No entries were excluded from Summerfield set. Observed and predicted $\text{p}f_{u,br}$ along with physicochemical property values for compounds comprising the internal and external validation sets are available in Ref.[178].

CNS access. 1696 drugs with assigned BBB penetration categories, as published by Adenot & Lahana[179], were used for classifying chemicals as accessible or inaccessible to CNS. This data set was constructed from World Drug Index (WDI) database after filtering out the records considered not relevant by the authors (local, topical applications, industrial chemicals, etc.). BBB+ category was assigned to the compounds used in neuropsychiatry (ATC code "N0" – Nervous system) except for those with a peripheral mechanism of action or having an ambiguous status regarding BBB penetration. Importantly, the authors acknowledge that CNS activity is not equivalent to BBB passage, and for many non-CNS drugs it is unclear whether they are unable to enter the brain or simply do not interact with any CNS targets. Therefore, only those drug classes for which there is enough evidence of being restricted from brain (*e.g.*, various antibiotics) were included in BBB– subset. Since the current study deals with simple physicochemical models of processes governed by passive diffusion, 81 BBB– compounds denoted as P-gp efflux substrates in Ref.[179], as well as 25 data points in BBB+ subset (several compound classes known to enter the brain by carrier-mediated influx – amino-acids, nucleosides, vitamins), were removed yielding a smaller data set of 1581 compounds (1311 BBB+ and 270 BBB– molecules).

2.1.3. Intestinal permeability

The current study involved QSAR analysis of two systems characterizing permeability of drugs across intestinal epithelium: *in vivo* human small intes-

tine and Caco-2 monolayers, a system serving as an *in vitro* surrogate of intestinal absorption (see Table 2.3). *In vivo* absorption measurements are commonly reported as fraction absorbed after oral administration, designated %*HIA* in the percentage form. %*HIA* database was taken from an earlier study by our group that focused on classification of drugs as absorbed/non-absorbed[180]. The data used therein were mainly collected from reference publications[168-171]. Another key data source was the compilation by Zhao et al.[77] where the authors carefully analyzed the absorption data from original articles and prepared a list of "preferred" %*HIA* values. In general, if several different fractions absorbed were reported for the same compound, the value from Ref.[77] as the most reliable data source was included in our database. The previously analyzed data set contained over 1000 compounds, but only a subset of 567 compounds was suitable for use in the current study. This was due to several reasons. First, for some compounds only qualitative absorption categories were given (*i.e.*, "moderately absorbed", "incomplete absorption", etc.). Second, data were often provided as oral bioavailability (%*F*) without making a clear distinction between %*F* and the extent of absorption represented by %*HIA*. Since lower bioavailability may result from extensive first-pass metabolism, only the values in the range %*F* > 80% were considered a good approximation of %*HIA*. Finally, the data for some compounds could be affected by transport mechanisms different from passive diffusion, or other side processes.

For several compounds there was no clear evidence of carrier-mediated transport or hepatic first-pass metabolism, but suspicions regarding the possible contribution of these processes were raised during model development. Therefore, it was important to select a reference data set containing experimental data of higher quality in an attempt to minimize the influence of potential side effects. For this purpose, only the data provided by Zhao et al. were used. Some of their values were either changed or removed in later studies by the same authors[181,182]. %*HIA* for 169 out of 195 compounds in their initial compilation were considered "reasonably free" of unwanted effects and were used a reference data set in the current study.

Table 2.3 Human intestinal absorption and Caco-2 permeability data set sizes.

%HIA		P_e (Caco-2)	
Data set	N	Data set	N
Full training set	567	Training set	473
Reference data set[77]	169	Internal test set	209
log P_{eff} validation set[81]	25	External validation set	300
log K_a validation set[183]	22		

The obtained model was validated using two data sets containing other quantitative characteristics of human intestinal absorption: *in vivo* human jejunal permeability coefficients (P_{eff}) and absorption rate constants (K_a). The first external validation set consisted of P_{eff} values for 25 drugs taken from a review by Lennernäs[81], which may be considered unaffected by various side processes such as carrier-mediated absorption. The second validation set consisted of 22 K_a constants corresponding to passive diffusion that were determined by deconvolution analysis of drugs' concentration-time profiles in human plasma by Linnankoski et al.[183]. These two parameters are directly comparable since absorption rate constants are linearly related to jejunal permeability ($K_a \approx 2 \cdot P_{eff}/R$, where R is the radius of the intestine, so that log $K_a \approx$ log $P_{eff} +$ Const)[184]. Experimental data along with calculated permeability coefficients (log P_{eff}) and physicochemical properties of drugs constituting the validation sets, as well the reference data set from Zhao et al., are available in Ref[185].

Caco-2 permeability data were collected from many original experimental articles. Caco-2 data set differs from all other endpoints considered in this study in that aspect that a few data points per compound may be included in the database. Since it is an *in vitro* assay, experimental conditions can be easily varied, and separate data points represent permeabilities of the same compound measured at different pH values or at different stirring rates applied to the test solution. In some works stirring rate was not reported, and it was assumed that the experiments were conducted without stirring. In general, the results of such measurements were considered lower quality data, but most of them were retained in the data set apart from a few entries where given P_e values were significantly higher than the limit imposed by unstirred water layer at no stirring
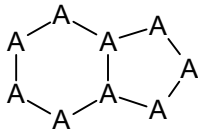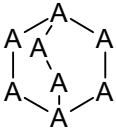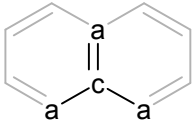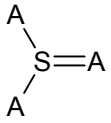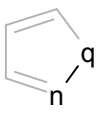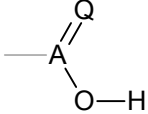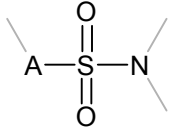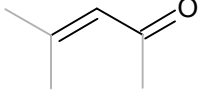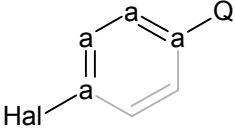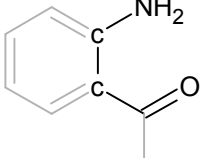
conditions (see Section 4.4). Overall, analysis of articles published before 2008 yielded a set of about 700 permeability coefficients for 512 drugs and drug-like compounds. After removal of questionable entries presumably affected by P-gp efflux (polarized transport with efflux ratio > 3) or PepT1 influx, the final data set contained 682 P_e values determined at pH ranging from 4.0 to 8.0 and stirring rates from 0 to 1100 rpm. When model development was finished, the data set was extended with further 300 entries mostly obtained from newer articles (2008-2010). The newly collected data were used for external validation of the model.

2.2. Descriptors

2.2.1. Fragmental descriptors

The most natural way for a chemist to describe a molecule is to list the structural features present in that molecule. In QSAR such description is accomplished by using fragmental descriptors. In the current study, fragmental approach was utilized for building the baseline model of plasma protein binding strength (Section 2.3.1). The molecules were fragmented using a predefined fragment set which is essentially an extension of the analogous set published by J. A. Platts and coworkers[186]. These fragments reflect a wide range of molecular properties such as size, topology (linear, branched, cyclic, etc.), aromaticity, presence of various functional groups, intramolecular interactions. A number of fragments indicate if the molecule belongs to a certain compound class (β-blockers, nucleobases, glycosides). The fragments were selected with an intention to cover as broad chemical space as possible. The overall list contains 379 different substructures, and several examples are shown in Table 2.4. Some of the fragments included in this list are often found in albumin ligands, while certain other substructures represent compounds that largely circulate unbound in human plasma. Statistically determined contributions of various substructures to plasma protein binding strength will be discussed in Section 4.1.

Table 2.4 Examples of structural fragments considered in the analysis. Notation: A – any atom; Q – any heteroatom from the list: N, O, S; Hal – any halogen atom; lower-case symbols denote atoms within aromatic rings.

Fragment	Type	Description
	Shape/Topology	Bicyclo[4.3.0] ring system
	Shape/Topology	Bicyclo[2.2.2] ring system
	Atom types	Permanently charged nitrogen
	Atom types	Fused aromatic carbon
	Atom types	sp^{2}d-hybridized sulfur
	Generalized substructures	Pyrazole, isoxazole or isothiazole type aromatic ring
	Functional groups	Any carboxylic, sulpho- or phosphonic acid
	Functional groups	Sulphonamide
	Functional groups	α,β-unsaturated carbonyl
	Interactions	p-Halophenol, haloaniline or halothiol
	Interactions	Intramolecular hydrogen bond – ortho- interaction between amino and carbonyl groups

2.2.2. Physicochemical properties

Permeabilities of cell membranes to drug molecules, as well as tissue to plasma partitioning processes, were modeled on the basis of mechanistic physicochemical approach. In other words, instead of deriving purely statistical models predicting the property of interest directly from chemical structure, experimental data were fitted to theoretical equations relating the analyzed property to key physicochemical parameters, such as lipophilicity, ionization, hydrogen bonding potential, and molecular size. The list of physicochemical descriptors considered in the current study along with their brief descriptions is presented in Table 2.5. In the next sections it will be shown that in most cases it is not feasible to use $\log D_{o/w}$ as a single descriptor accounting for both lipophilicity and ionization. Instead, one should prefer using separate descriptors for lipophilicity ($\log P_{o/w}$) and ionization state of the compound. The latter can be expressed by ion form fractions estimated from pK_a and pH of the environment[68]. *E.g.*, if the molecule contains both acidic and basic ionization centers:

$$f_0 = \frac{1}{1+10^{pH-pK_a}+10^{pK_a^B-pH}+10^{pH-pK_a}\cdot 10^{pK_a^B-pH}} \qquad (2.2)$$

$$f_- = \frac{10^{pH-pK_a}}{1+10^{pH-pK_a}+10^{pK_a^B-pH}+10^{pH-pK_a}\cdot 10^{pK_a^B-pH}} \qquad (2.3)$$

$$f_+ = \frac{10^{pK_a^B-pH}}{1+10^{pH-pK_a}+10^{pK_a^B-pH}+10^{pH-pK_a}\cdot 10^{pK_a^B-pH}} \qquad (2.4)$$

$$f_\pm = \frac{10^{pH-pK_a}\cdot 10^{pK_a^B-pH}}{1+10^{pH-pK_a}+10^{pK_a^B-pH}+10^{pH-pK_a}\cdot 10^{pK_a^B-pH}} \qquad (2.5)$$

Here f_i terms represent the fractions of ith ionic species, pK_a^B denotes pK_a of basic group. If the compound has no acid or base pK_a, the respective terms in the above equations become equal to zero.

The considered hydrogen bonding descriptors involved several continuous properties – *TPSA*, Abraham's solvation parameters (*A* and *B*), as well as simple counts of hydrogen bond donor and acceptor sites in the molecule. A composite descriptor ($N_{HD} + 1/2\ N_{HA}$) reflecting both hydrogen bond donating and accepting potential was also used in the analysis.

Table 2.5 The list of simple physicochemical descriptors used in the analysis.

Group	Parameter	Description	References
Lipophilicity	$\log P_{o/w}$	Octanol/water partitioning coefficient of neutral species	[68]
Lipophilicity & Ionization	$\log D_{o/w}$	Octanol/water partitioning coefficient of all ionic species of a compound at given pH	[68]
Ionization	pK_a	Ionization constant. Strongest acid and base pK_a of a compound were used	[68]
Hydrogen bonding	N_{HD}	Count of OH + NH atoms	–
	N_{HA}	Count of all O and N atoms	–
	TPSA	Surface area occupied by polar atoms	[187]
	A	Overall solute hydrogen bond acidity	[188]
	B	Overall solute hydrogen bond basicity	[188]
Molecular size	*MW*	Molecular weight	–
	V_x	McGowan characteristic volume	[189]

2.3. Statistical methods

2.3.1. GALAS modeling methodology

The model for predicting the extent of plasma protein binding was developed using recently introduced GALAS (Global, Adjusted Locally According to Similarity) modeling methodology[4,5]. Each GALAS model consists of two parts:

- A global (baseline) statistical model.

- A similarity based routine that introduces corrections to baseline predictions according to experimental data for the most similar compounds from the training set (local model).

In most cases the described approach allows achieving better prediction accuracy compared to conventional QSAR models and also enables the assessment of the model applicability domain. Another important feature of GALAS models is their trainability, *i.e.*, the possibility to expand the applicability domain and adapt the model to novel compound classes without the need for full statistical reparameterization.

Global model (PLS). In GALAS methodology baseline models for continuous quantitative properties are derived using PLS statistical method (Projection to Latent Structures by means of Partial Least Squares)[190,191] combined with multiple bootstrapping procedure[192] that defines the so called dynamic similarity key. The concept of dynamic similarity is one of the important features of GALAS methodology – it means that similar compounds are extracted from the data set not by purely chemical similarity criteria, but instead, determination of what is similar is made in the context of the particular property.

The optimal number of components is determined using *Leave-One-Out* cross-validation minimizing *PRESS* (*Predictive Residual Sum of Squares*) statistic (or maximizing the cross-validation Q^2)[190,193]:

$$PRESS = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{2.6}$$

$$Q^2 = 1 - \frac{PRESS}{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2} \tag{2.7}$$

y_i are experimental values, $\hat{y}_i$ – predicted values, and $\bar{y}$ – response mean.

An important aspect to consider before using linear methods such as PLS is to ensure that the analyzed data uphold the LFER (Linear Free Energy Relationship) principle. Plasma protein binding data are reported as percentage fractions varying within a confined scale (0 to 100%). *%PPB* exhibits sigmoid relationship with energy-representing parameters and should be linearized prior to modeling. The obvious solution would be to convert *%PPB* values to apparent plasma binding constants $\log K_{app}$ in a similar way like the affinity constants for any particular protein could be calculated:

$$\log K_{app} = \log \left(\frac{\max \left(\%PPB, 1\% \right)}{100\% - \min \left(\%PPB, 99.99\% \right)} \right) \tag{2.8}$$

Note that lower and upper limits of *%PPB* were set to 1% and 99.99%, respectively. This was done to avoid infinities when the data are reported as "not bound" or "completely bound". Different cut-offs were selected for strong and weak binders due to higher sensitivity of measurements on the upper end of the

scale. This is understandable as differentiating between 0 and 1% binding does not have any practical importance while an increase in %*PPB* from 99% to 99.9% leads to a tenfold decrease in plasma free fraction.

Local model. Baseline model may reflect the general trends in the analyzed property reasonably well, but it cannot account for the specific effects observed locally within a particular chemical class. The local part of the GALAS modeling methodology allows correcting the systematic deviations introduced by the global model by analyzing its performance in the local chemical neighborhood of the analyzed compound.

Obviously, in the case of plasma protein binding, the local correction (denoted Δ) should be applied to apparent binding constants $\log K_{app}$ as shown in Eq. (2.9), whereas the final predicted value can be converted back to the percentage scale:

$$\log K_{app,pred.} = \log K_{app,baseline} + \Delta \tag{2.9}$$

$$\%PPB_{pred.} = \frac{\log K_{app,pred.}}{1 + \log K_{app,pred.}} \cdot 100\% \tag{2.10}$$

Each GALAS model prediction is supplemented by calculated Reliability Index (*RI*) – a value ranging from 0 to 1 that serves as a quantitative measure of the prediction reliability. *RI* is a product of the compound's Similarity Index (*SI*) to the training set compounds and Model-Data Consistency Index (*MDCI*) that reflects the degree of variation among Δ values of individual compounds.

$$RI = SI \cdot MDCI \tag{2.11}$$

These two components of the *RI* value correspond to the main principles of reliability estimation:

- A reliable prediction can only be made if the training set used to derive the model contains compounds similar enough to the analyzed molecule (the degree of similarity is given by *SI*).

- A prediction should also be considered unreliable if experimental data for similar compounds are inconsistent with the corresponding baseline predictions (*i.e.*, obtain large Δ values of opposite signs, which is reflected by low *MDCI*).

If at least one of the outlined criteria is not met, the corresponding index, and ultimately *RI*, approaches zero. After analyzing a number of GALAS models, it was empirically determined that *RI* = 0.3 can be viewed as a cut-off for figuring out if a compound belongs to the model applicability domain[4].

2.3.2. Non-linear regression

A common goal of QSAR analysis is finding a multivariate predictor function (2.12) that relates the response vector *y* with descriptor matrix *x*:

$$y = f(x, \beta) + \varepsilon \tag{2.12}$$

Here *β* is a vector of model parameters, and *ε* is the error term that corresponds to unknown sources of variation in response values[194,195]. The majority of statistical methods including PLS assume that the function *f* represents linear dependence between response and independent variables, though it is not always the case, especially in biological systems. The method described in this section applies when the considered relationship is non-linear.

The basis of Non-linear Least Squares (NLS) regression procedure is minimization of the residual sum of squares (*RSS*) with respect to model parameters *β*:

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - f(x_i, \beta))^2 \tag{2.13}$$

Minimizing *RSS* is equivalent to maximizing the likelihood parameter (*L*) of the model, which is defined as follows:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{RSS(\beta)}{2\sigma^2}\right) \tag{2.14}$$

The best fit is obtained when the parameter estimates $\beta = \hat{\beta}$ are used and the maximum likelihood is given by Eq. (2.15) where $\hat{\sigma}^2 = \frac{n-p}{n} s^2$ is used as an estimate of σ^2 (Ref.[194]).

$$L(\hat{\beta}, \hat{\sigma}^2) = \frac{1}{\left(2\pi \cdot \frac{RSS(\hat{\beta})}{n}\right)^{n/2}} \exp\left(-\frac{n}{2}\right) \tag{2.15}$$

The values of the parameter estimates producing the best fit are found by differentiating the *RSS* and setting partial derivatives to zero[195]:

$$\frac{\partial RSS(\beta)}{\partial \beta} = -2\sum_i (y_i - f(x_i, \beta)) \frac{\partial f(x_i, \beta)}{\partial \beta}$$ (2.16)

Due to the non-linearity of these equations, they are usually solved by iterative numerical optimization methods, such as the commonly used Gauss-Newton algorithm[194,195].

2.3.3. Linear discriminant analysis

Linear Discriminant Analysis (LDA) introduced by R. A.. Fisher has become one of the most popular techniques for dealing with two-class classification problems along with logistic regression[196,197]. In comparison to the latter, LDA is more sensitive to predictor normality violations, but it is a more robust method that yields more precise predictions when the level of noise is sufficiently low and response classes are well differentiated by the predictor values[197]. In the current study, LDA method was used for the assignment of BBB+/− categories on the basis of qualitative CNS access data.

The goal of LDA is to find an optimal discriminant function Δ defined as a linear combination of predictor parameters x_i that would allow for the best possible separation of response classes[197]:

$$\Delta = \beta_o + \sum_i \beta_i x_i$$ (2.17)

Discrimination is based on the idea that the optimal Δ function should maximize the ratio of between-class sum of squares (*SSB*) to within-class sum of squares (*SSW*)[196,197]. The statistical significance of a descriptor x_i used for classification is given by the respective *F*-value:

$$F_i = \frac{SSB_i}{SSW_i} = \frac{SST_i - SSW_i}{SSW_i}$$ (2.18)

$$SST_i = \sum_j \sum_k (x_{ijk} - \bar{x}_i)^2$$ (2.19)

$$SSW_i = \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2$$ (2.20)

Here *SST* is the total sum of squares, $\bar{x}_{ij}$ – the mean of parameter x_i within the jth group, and $\bar{x}_i$ – the overall mean of x_i in the data set.

If the population consists of two classes and predictor variables distribute normally with the same covariances in both classes, linear discriminant rule is equivalent to the maximum likelihood rule that minimizes the expected cost of misclassification (*ECM*)[196]:

$$ECM = p_{21}\pi_1 + p_{12}\pi_2 \qquad (2.21)$$

Here π_1 and π_2 are the prior probabilities of a data point belonging to a particular class, and p_{ij} denotes the probability of erroneous assignment of a data point to class i when it actually belongs to class j. Eq. (2.21) assumes the equal cost for both possible cases of misclassification.

In addition to the *F*-value defined in Eq. (2.18), the discriminative power of the classifier function is sometimes evaluated using a related term called Wilk's lambda or *U*-statistic:

$$U = \frac{SSW}{SST} \qquad (2.22)$$

Wilk's lambda ranges from 0 to 1 where $U = 1$ shows that the class means do not differ at all and the classifier performs no better than chance. The smaller the *U*-value – the better is the achieved separation of classes[198].

2.4. Software

All database manipulations, generation of physicochemical and fragmental descriptors, statistical analysis of plasma protein binding data using PLS method, and development of GALAS models were performed using Algorithm Builder program[199]. Nonlinear physicochemical models of other distribution-related properties, BBB and Caco-2 permeability, and human intestinal absorption were derived using the implementation of NLS regression procedure provided in the R program[200]. Linear discriminant analysis of qualitative CNS access data was also performed in the R program using "MASS" package.

3. THEORY

3.1. General considerations

According to thermodynamic partitioning theory of electrolytes, a compound's partitioning coefficient between two phases should depend on its ionization state. If we assume that all types of ionic species partition independently of each other, then the overall partitioning coefficient should be equal to the sum of partitioning coefficients of all ionic species. A good example illustrating this concept is the commonly used expression for calculating pH-dependent octanol-water distribution coefficient $\log D_{o/w}$ from the respective partitioning coefficient of neutral species ($\log P_{o/w}$)[68]:

$$\log D_{o/w} = \log\left(\sum_i f_i \cdot P_i\right) = \log\left(\sum_i f_i \cdot 10^{\log P_o + \Delta_i}\right) \tag{3.1}$$

Here i denotes the particular type of ionic species (neutral, mono-cation, mono-anion, zwitterion), f_i – ion fraction of the ith species that depends on the solution pH and solute pK_a.

A common approach in QSAR is to use octanol-water $\log D$ as a composite descriptor to evaluate both ionization and lipophilicity dependence of the analyzed property. However, the particular Δ_i values describing the ionization influence in a relevant biological system should not necessarily be equivalent to those observed in octanol-water system. The main idea employed in the current study states that the same type of ionization–partition dependence as shown in Eq. (3.1) should be applied separately to each analyzed system, so that specific Δ_i values typical for that process (its quantitative representation is designated $\log K_p$) could be obtained[201]:

$$\log K_p = \log\left(\sum_i f_i \cdot K_p^i\right) = \log\left(\sum_i f_i \cdot 10^{\log K_p^o + \Delta_i}\right) \tag{3.2}$$

In Eq. (3.2) characteristic constants $\log K_p^i$ are expressed as $\log K_p^o$ of neutral species shifted by an appropriate Δ_i value specific for each type of ionic species. This is consistent with the common understanding that all species partition by the same mechanism with the only difference introduced by electros-

tatic contributions. $\log K_p^o$ can be described in terms of lipophilicity (octanol/water $\log P$) and optional hydrogen bonding (HB) descriptors:

$$\log K_p^o = c_o + c_1 \log P_{o/w} + \alpha \log\left(1 + 10^{\log P_{o/w} - \beta}\right) + [c_2 HB] \tag{3.3}$$

For many biological properties, the relationship with lipophilicity is not linear throughout the entire $\log P$ scale. Membrane permeation rate increases with lipophilicity only up to a certain point. Highly hydrophobic molecules permeate slower as they may be "trapped" in the membrane. This type of $\log P$ dependence is best described by bilinear equation first suggested by H. Kubinyi[192]. As shown in Eq. (3.2), bilinear transformation introduces two additional terms: α, that determines the rate of decrease in permeability after passing $\log P_{o/w}$ optimum, and β, that is close to the optimal $\log P_{o/w}$ value and controls the inflection point in the permeability-lipophilicity relationship.

3.2. Membrane permeability

Consider the modified Fick's law:

$$J = P_m \Delta C_m = \left(\frac{D_k \cdot K_d}{h}\right) \Delta C_m \tag{3.4}$$

Here J is transmembrane flux of drug, P_m – its membrane permeability, D_k – kinetic diffusion coefficient, K_d – thermodynamic partitioning ratio between lipid phase and water, ΔC_m – the concentration difference across the membrane, h – thickness of the membrane.

Since h is a property of the membrane that can be assumed constant for any considered drug molecules, Eq. (3.4) can be rewritten as follows:

$$P_m = \left(\frac{1}{h}\right) D_k K_d = C_{mem} D_k K_d \tag{3.5}$$

The objective of a QSAR study is to derive predictive models for non-constant terms D_k and K_d.

According to the Stokes-Einstein equation, D_k is inversely proportional to the compound's radius that can be approximated by the cubic root of McGowan characteristic volume ($V_x^{-1/3}$)[189]:

$$D_k \approx C_{diff} V_x^{-1/3} \tag{3.6}$$

Analysis of blood-brain barrier permeability data revealed that D_k only accounts for about 10% of variation in log PS constants, while the major part of variation is explained by the differences in the compounds' K_d[139]. Therefore, partitioning between the membrane and water phase is a key determinant of permeability of drugs across cell membranes. Since K_d represents a non-specific partitioning process, it can be modeled according to Eq. (3.2).

The considerations discussed above are common to all absorptive systems. The following paragraphs will focus on specific differences between these systems.

Intestinal permeability. Prior to reaching mucosal epithelium drug molecules must first pass the unstirred water layer (UWL). After that, the molecules can be absorbed either by transcellular route across phospholipid bilayers or by paracellular route through intercellular tight junctions. The overall resistance to diffusion (inverse of permeability) is calculated as a sum of all particular resistances in the system[68]:

$$\frac{1}{P_e} = \frac{1}{P_{UWL}} + \frac{1}{P_{para} + P_{trans}} \tag{3.7}$$

Passive transcellular permeability (P_{trans}) is the main contributor to the absorption of the majority of drug-like compounds. This mechanism plays an important role in all analyzed systems, and the considerations outlined in the beginning of the current section apply to this particular transport route. Transcellular permeability was modeled following the approach represented by Eqs. (3.2-3.6) where the term P_m was replaced by P_{trans}.

Paracellular permeability (P_{para}) was described by a theoretical model of molecular size-restricted diffusion through aqueous pores in tight junctions[89]:

$$P_{para}^{o} = \frac{\varepsilon D F\left(\dfrac{r}{R}\right)}{\delta} \tag{3.8}$$

$$P_{para}^{i} = P_{para}^{o}\left(\frac{\kappa z_i}{1 - e^{-\kappa z_i}}\right) \tag{3.9}$$

Here ε is the porosity of the epithelium, D is the aqueous diffusivity of the molecule expressed by Eq. (3.6), δ is the tortuosity factor of the pore, κz is the electrochemical energy function, and $F(r/R)$ is the Renkin molecular sieving function[202]:

$$F\left(\frac{r}{R}\right) = \left(1-\frac{r}{R}\right)^2\left(1-2.104\frac{r}{R}+2.09\left(\frac{r}{R}\right)^3-0.95\left(\frac{r}{R}\right)^5\right) \tag{3.10}$$

Replacing all constant terms in the above equations by scaling factor C_{para} yields the overall model for paracellular transport:

$$P_{para} = C_{para} \cdot V_x^{-1/3} \cdot F\left(\frac{r}{R}\right)\left[\sum_{i=-2}^{2} f_i + \sum_{i=-2}^{2(i\neq0)} f_i \cdot \left(\frac{\kappa z_i}{1-e^{-\kappa z_i}}\right)\right] \tag{3.11}$$

Several types of ionic species with charges ranging from -2 to +2 had been considered. For human jejunum the values $R = 5.6$ Å and $\kappa = 3$ (corresponding to transepithelial potential drop of 80 mV) determined by Sugano et al.[202,203] were used, while the respective values for Caco-2 monolayers were obtained in the current study.

Permeability of the unstirred water layer (P_{UWL}) is expressed differently in human jejunum and Caco-2 systems. It has been shown experimentally that effective jejunal permeability of glucose in humans is equal to 10^{-3} cm/s[204]. This value is assumed to be fully limited by diffusion through unstirred water layer as glucose is rapidly absorbed by carrier-mediated mechanisms. Thus, glucose may be considered an UWL permeability marker, and it is possible to express P_{UWL} of other compounds by Eq. (3.12) where 3.6 Å corresponds to the radius of glucose molecule. The radius of test molecule's spherical equivalent (r) is calculated from its McGowan volume (V_x) according to Eq. (3.13)[202]:

$$P_{UWL} = 10^{-3} \times \left(\frac{3.6}{r}\right) \tag{3.12}$$

$$r = \sqrt[3]{\frac{3V_x}{4\pi}166} \tag{3.13}$$

It should be noted, that in human jejunum UWL is only significant when permeability coefficients (P_{eff}) are considered since P_{UWL} corresponds to %HIA $> 99.9\%$[158] and such values are experimentally indistinguishable from 100%.

In Caco-2 monolayers the situation is quite different. It is an *in vitro* system and here UWL thickness depends on experimental conditions (presence and rate of stirring). Often the interdependence between P_{UWL} and stirring rate is described by the following empirical equation:

$$P_{UWL} = Kv^{\alpha} \tag{3.14}$$

Here K is a fitted constant reflecting diffusivity, viscosity, and geometric factors; v is the stirring rate; and the exponent α is proposed to reflect the type of hydrodynamics[87,88]. The major flaw of Eq. (3.14) is that it implies $P_{UWL} = 0$ in the absence of stirring, which is definitely not true. Therefore, a modified version of this equation incorporating an additional constant (K_o) equal to P_{UWL} at $v = 0$ was used in the current study:

$$P_{UWL} = K_o + Kv^{\alpha} \tag{3.15}$$

Parameters K_o, K, and α were fitted on the basis on Caco-2 permeability data for four model compounds at varying H and stirring rates provided in Ref.[88].

Permeability *vs.* fraction absorbed. Traditionally the relationship between fraction absorbed and effective permeability has been described by first-order kinetic equation. Yu and Amidon suggested a compartmental absorption and transit (CAT) model that allows for a better fit to experimental *%HIA* data[184]. Their model makes an assumption that the drug passing the intestine flows through a series of seven segments defined by separate compartments, all of which have the same residence time. Linear transfer kinetics between compartments is assumed, whereas colon is viewed only as a reservoir. *%HIA* can then be expressed as follows:

$$\%HIA = 100\% \times \left(1 - \left(1 + C_A P_{eff}\right)^{-7}\right) \tag{3.16}$$

C_A is a scaling factor between fraction absorbed and effective permeability. Eq. (3.16) was used for interconversion between these parameters in the current study.

Blood-brain barrier permeability. Due to the hydrodynamic conditions in brain capillary endothelial cells, UWL thickness at the blood-brain barrier is close to zero[68], and P_{UWL} term in Eq. (3.7) can be disregarded. However, at

BBB another limiting effect on permeability can be observed. In this case cerebral blood flow-limited distribution into the brain occurs[175]. The influence of the new term – flow-limited permeability (P_F) can be expressed in a similar way, as a component of overall resistivity to diffusion:

$$PS = \left(\frac{1}{P_F} + \frac{1}{P_m S}\right)^{-1} = \left(10 + \frac{1}{P_m S}\right)^{-1} \tag{3.17}$$

P_F has profound influence for highly permeable compounds like tricyclic antidepressants that diffuse across membrane much faster than they are delivered to the diffusion place. The influx rate constant of such compounds becomes almost indistinguishable from cerebral blood flow ($K_{in} \approx P_F$). If $P_m \cdot S$ exceeds experimental flow more than about twice, then reliable determination of $P_m \cdot S$ from K_{in} is impossible[55]. Since the cerebral blood flow in a typical *in situ* rat brain perfusion experiment is *ca.* 0.05 mL/s/g[61,136,138], the maximum measurable $PS \approx 2 \cdot P_F \approx 0.05 \times 2 = 0.1$ (mL/s/g). Hence, the value of ($1/P_F$) was fixed at 10 (s·g/mL).

Combining Eqs. (3.2-3.6), and (3.17) into a single expression yields the final equation for calculating log PS:

$$\log PS = -\log\left(10 + \frac{CV_x^{1/3}}{\sum_i f_i 10^{\log K_d^{\circ} + \Delta_i}}\right) \tag{3.18}$$

where $C = C_{diff} \cdot C_{mem}$ is a product of two above mentioned normalizing factors.
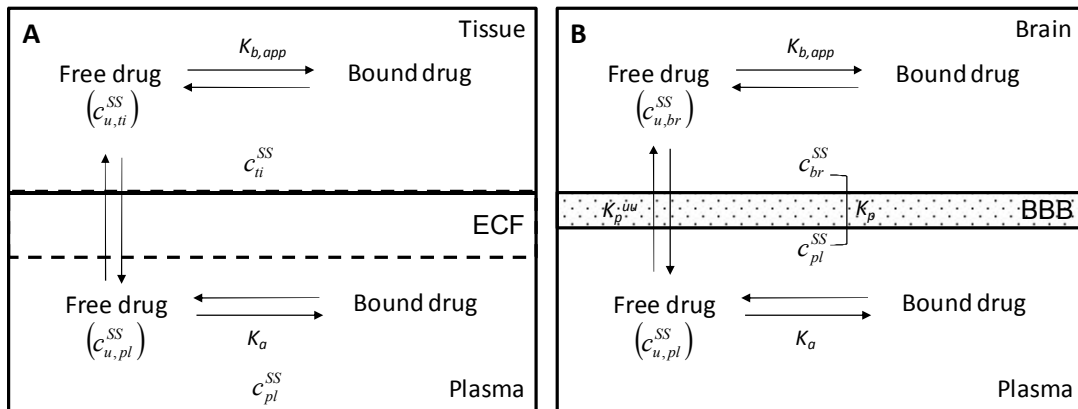

Figure 3.1 An overview of the processes involved in drug distribution in the body: (A) generalized scheme of distribution between plasma and tissues; (B) specific case of brain/plasma partitioning.

3.3. Tissue-plasma partitioning

Figure 3.1, (A) presents a simplified outline of the processes involved in the distribution of drug molecules between plasma and tissues, while part (B) of the same figure illustrates solute partitioning between plasma and brain – an example of drug distribution into a particular tissue. In plasma drug molecules circulate in both bound to proteins and free (unbound) form, and it is the unbound form that equilibrates between plasma and tissues[205,206]. Once in the brain or any other tissue, the molecules redistribute intra- and extracellularly. As one can see from the figure, brain is unique amongst other tissues since it is separated from circulation by one of the most restrictive physiological barriers – the blood-brain barrier.

In the current study, brain tissue was treated as a single compartment without making a distinction between drug partitioning into cells and distribution within brain interstitial fluid. Although this may seem an oversimplification, it is unlikely to introduce significant prediction errors. First, once the molecules have overcome the tightest barrier (BBB), permeation into the cells inside brain tissue should not be a limiting factor. Second, binding of drug molecules to both intracellular lipids[207] and extracellular tissue components such as lipoproteins[208] is in general governed by the same nonspecific processes which can be accounted by a single equation.

At steady state, the unbound drug concentrations in both plasma ($c_{u,pl}^{SS}$) and tissue ($c_{u,ti}^{SS}$) compartments are equal[206], whereas the ratio of total drug concentrations (c_{pl}^{SS} and c_{ti}^{SS}) represents the experimentally observed partitioning coefficient (K_p). Decimal logarithm of the respective coefficient for brain/plasma system is commonly known as log BB:

$$K_p = \frac{c_{ti}^{SS}}{c_{pl}^{SS}} \qquad (3.19)$$

Other parameters that reflect tissue distribution in terms of unbound drug concentrations have been introduced[205,206]. Because the extent of plasma protein binding is relatively easy to measure, tissue/plasma distribution is some-

times expressed as a $K_p{}^u$ coefficient, that is, K_p corrected for unbound fraction in plasma:

$$K_p^u = \frac{c_{ti}^{SS}}{c_{u,pl}^{SS}} = \frac{K_p}{f_{u,pl}} \tag{3.20}$$

When only free drug concentrations are considered in both compartments, the resulting concentration ratio is the "unbound partition coefficient" denoted $K_p{}^{uu}$:

$$K_p^{uu} = \frac{c_{u,ti}^{SS}}{c_{u,pl}^{SS}} = K_p \frac{f_{u,ti}}{f_{u,pl}} \tag{3.21}$$

The latter parameter is independent of differences in binding to plasma and tissue constituents. Thus, $K_p{}^{uu}$ is not a measure of compound's relative affinity to brain tissue but rather an indicator of the mechanism of its transport across BBB[205,206]. At steady-state the ratio of free drug concentrations in brain interstitial fluid and blood plasma should be close to unity because the rates of passive diffusion-driven influx and efflux would be similar. If this ratio is significantly different from unity, the steady state has most likely not been achieved due to carrier-mediated transport. $K_p{}^{uu} > 1$ may be attributed to facilitated diffusion when influx clearance is much greater than the opposite process, whereas $K_p{}^{uu} < 1$ is typical for compounds that undergo active efflux mediated by P-gp or other efflux pumps. $K_p{}^{uu}$ values significantly lower than unity may also be observed for compounds that do not reach steady state because their transmembrane permeability is too low.

If only passive transport is considered and the assumption that at steady state $c_{u,pl}^{SS} = c_{u,ti}^{SS}$ is correct, Eq. (3.21) can be rewritten as follows:

$$K_p \frac{f_{u,ti}}{f_{u,pl}} = 1 \Rightarrow K_p = \frac{f_{u,pl}}{f_{u,ti}} \tag{3.22}$$

Applying the same considerations to $K_p{}^u$ yields Eqs. (3.23) and (3.24):

$$K_p^u = \frac{1}{f_{u,ti}} \tag{3.23}$$

$$\log K_p^u = -\log f_{u,ti} = pf_{u,ti} \tag{3.24}$$

If we compare Eqs. (3.22) and (3.23), the fundamental difference between K_p and K_p^u becomes evident. K_p corresponds to the ratio of free fractions in tissue and plasma (*i.e.*, represents two different processes), while K_p^u provides an unambiguous measure of tissue binding[209,210]. The latter fact is emphasized in Eq. (3.24) where $\log K_p^u$ is replaced with a new term $pf_{u,ti}$ that clearly refers to the tissue unbound fraction. Obviously, the influences of the two processes should be evaluated separately, and $pf_{u,ti}$ but not $\log K_p$ is a preferable parameter for modeling.

Tissue unbound fractions and data quality. An important observation can be made if the nature of $pf_{u,ti}$ parameter is considered. Because $f_{u,ti}$ is a free fraction ranging from 0 to 1, $pf_{u,ti}$ cannot be negative. Yet, adjusting raw experimental data for the extent of plasma protein binding in order to yield the fraction unbound in tissue may result in $pf_{u,ti} < 0$ (see Eqs. (3.28) and (3.30) for details of converting $\log BB$ and V_{SS}, respectively). In these cases it may be implied that the assumption of passive transport is not valid and the reported values were not obtained at steady-state conditions. This conclusion serves as a powerful criterion to identify experimental data affected by carrier-mediated efflux.

Affinity to tissues. Because $f_{u,ti}$ does not uphold linear free energy relationship principle, it cannot be directly correlated with physicochemical parameters such as $\log P_{o/w}$. Instead, by analogy with binding to albumin and other proteins in plasma[16], the strength of the interactions between drug molecules and tissue constituents may be characterized by an apparent binding constant $K_{b,app}$:

$$K_{b,app} = \frac{1 - f_{u,ti}}{f_{u,ti}} \tag{3.25}$$

Rearrangement of the above expression yields the following equation for calculating $pf_{u,ti}$:

$$pf_{u,ti} = \log(1 + K_{b,app}) \tag{3.26}$$

It can be assumed that the apparent binding constant $K_{b,app}$ describes non-specific binding of drug molecule to lipid components of the tissue and in some cases to proteins (*e.g.*, brain albumin). Therefore, $\log K_{b,app}$ can be expressed in terms of simple physicochemical properties such as $\log P_{o/w}$ and ionization according to Eq. (3.2). The equation for calculating $\mathrm{p}f_{u,ti}$ can then be rewritten as shown below:

$$\mathrm{p}f_{u,ti} = \log\left(1 + \sum_i f_i 10^{\log K_{b,app}^o + \Delta_i}\right) \tag{3.27}$$

Brain/Plasma partitioning. Rearranging Eq. (3.22) and replacing $\mathrm{p}f_{u,ti}$ with brain-specific term $\mathrm{p}f_{u,br}$ yields the following expression relating brain/plasma partitioning ratio $\log BB$ to the extent of binding in plasma and brain tissue:

$$\log BB = \log K_p = \log f_{u,pl} + \mathrm{p}f_{u,br} \tag{3.28}$$

This equation was used to recalculate original $\log BB$ values collected from literature to $\mathrm{p}f_{u,br}$, and these in turn were subjected to non-linear fitting according to Eq. (3.27). Most erroneous values ($\mathrm{p}f_{u,br} < 0$) were omitted from the data set as violating the steady-state assumption. However, some entries with $\mathrm{p}f_{u,ti}$ marginally less than zero were retained if the error could possibly be attributed to the uncertainty in plasma protein binding data.

Volume of distribution. Steady-state volume of distribution (V_{SS}) can be related to free fractions of drug in plasma ($f_{u,pl}$) and tissue ($f_{u,ti}$), and organism-specific physiological parameters by Øie-Tozer equation [211]:

$$V_{SS} = V_P\left(1 + R_{E/I}\right) + f_{u,pl}V_P\left(\frac{V_E}{V_P} - R_{E/I}\right) + V_R\frac{f_{u,pl}}{f_{u,ti}} \tag{3.29}$$

Here $R_{E/I}$ is extra-/intravascular ratio of albumin, V_P – plasma volume, V_E – extracellular fluid volume, and V_R – volume of the remainder (tissue) fluid.

In the current study, a modified version of the Øie-Tozer equation was used. If human-specific values of the respective parameters[211] are entered into Eq. (3.29), similar values are obtained for albumin distribution volume $\left(V_A = V_P\left(1 + R_{E/I}\right)\right)$ and volume of extracellular fluid devoid of extracellular al-

bumin $\left(V_E - V_P R_{E/I}\right)$. Both terms can then be replaced by a single number: $V_A \approx 0.1$ L/kg. Additionally, the original Øie-Tozer equation does not reflect the fact that very hydrophilic molecules may be restricted to plasma and extracellular fluid due to their inability to cross cell membranes. This was accounted by introducing an additional permeability indicator variable I_P. Molecules having $\log D_{o/w} < -5$ at pH = 7.4 can be assumed impermeable and assigned the value $I_P = 0$. These compounds can have maximal $V_{SS} = 0.2$ L/kg. Less hydrophilic compounds ($I_P = 1$) may be distributed within total body water ($V_{SS} = 0.6$ L/kg) even if they do not significantly interact with tissue components. The proposed modifications resulted in the following empirical equation for estimating the volume of distribution at steady-state:

$$V_{SS} = V_A\left(1 + f_{u,pl}\right) + I_P V_R \frac{f_{u,pl}}{f_{u,ti}} \tag{3.30}$$

In Eq. (3.30), the right-hand side term $f_{u,pl}/f_{u,ti}$ is equivalent to the tissue/plasma partitioning coefficient K_p, which is modeled as described above.

Multi-step fitting. All endpoints discussed in this section represent complex multi-mechanism properties, and each was described by a system of non-linear equations. In order to avoid statistical artifacts and ensure that parameters related to different processes are obtained independently, it was necessary to perform data fitting in a multi-step approach. The first step in the analysis of permeability across human jejunum and Caco-2 cell monolayers involved optimization of P_{para} parameters from Eq. (3.11) based on experimental data for small hydrophilic compounds ($MW < 250$, $\log P_{o/w} < 0$). This step was not needed in $\log PS$ or tissue binding models. Parameters characterizing transcellular transport or lipid binding affinity of non-electrolytes were determined in the next step, and finally data for electrolytes were utilized to establish the ionization dependence of the considered endpoint (Δ_i coefficients).

4. RESULTS & DISCUSSION

4.1. Plasma protein binding

Baseline model. The best baseline model of log K_{app} according to *PRESS* statistic was obtained using PLS with 7 components. Models of this kind can be interpreted in terms of descriptor coefficients. Although fragmental descriptors have a clear meaning, the interpretation is not always unambiguous due to fragment overlapping. Still, the general patterns of structural features frequently found in highly protein bound molecules are evident. Several representative fragments and their coefficients in the PLS model are listed in Table 4.1 along with short comments. The table includes characteristic substructures found in the ligands of all major plasma proteins. *E.g.*, arylacetic acids (ibuprofen, flurbiprofen) are typical human serum albumin Site II ligands, whereas many non-carboxylic acids that are at least partially ionized at physiological conditions also interact with albumin. 2-aminopyrimidine fragment is a part of the common scaffold of α-adrenergic blockers (doxazosin, prazosin, terazosin, etc.) that predominantly bind to AAG. Terpene fragment is frequently encountered in large lipophilic molecules that are transported in lipoprotein particles. Dihydropyridine (dipine) compound class does not demonstrate a clear preference towards a particular protein – dipines readily interact with all major carriers, resulting in a very low free fraction in plasma.

Final model. After similarity correction, the resulting GALAS model was able to predict %*PPB* of test set compounds with $R^2 = 0.70$ (Figure 4.1, (A)). Note that predictions were filtered according to Reliability Index (*RI*) values – predictions with $RI < 0.3$ were considered unreliable and excluded from further analysis. As shown in Figure 4.1, (B), even better correlation between experimental and predicted %*PPB* was observed for compounds obtaining moderate to high *RI* values ($RI > 0.6$).

Despite good overall accuracy, significant deviations between the model estimates and experimentally determined %*PPB* values are observed for some compounds. The majority of them obtain borderline reliability, and their omis-

sion leads to a significant improvement in predictivity for the remaining molecules. If only predictions of at least moderate reliability are considered, in > 95% cases calculated %*PPB* stays within 30% from the experimental values. These results support the previously reached conclusion that Reliability Index values produced by GALAS models serve as a suitable measure for identifying accurate predictions[4,5]. *RMSE* of predictions expressed in the logarithmic scale is comparable to the values reported in previous studies[120,121]. Another important aspect is the proportion of test set compounds that obtain *RI* within acceptable range. In the current work, Reliability Index was at least borderline ($RI \geq$ 0.3) for 265 of 291 test set molecules (91.1%), while $RI \geq 0.6$ was obtained for 107 compounds (36.8%). In practice this means that the model could be expected to produce highly reliable predictions for about one third of a previously unseen data set.

It has been already mentioned that fraction unbound in plasma is a key property that influences all subsequent distribution processes. Precision of protein binding data for use as a descriptor in other QSAR models should be as high as possible. For that reason, models of plasma to tissue distribution described further mostly rely on experimentally measured %*PPB* values, while the presented GALAS model has been used to "fill the gap" where experimental data were not available.
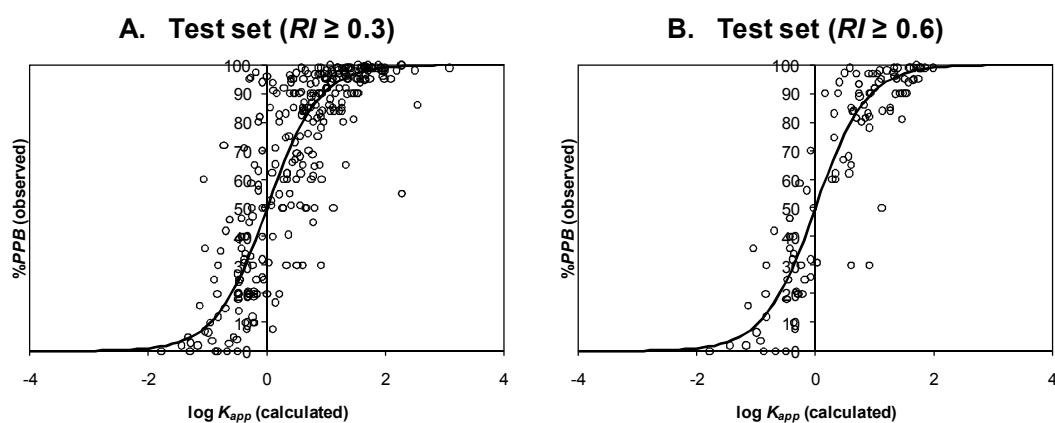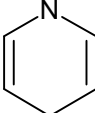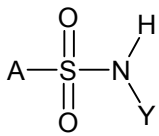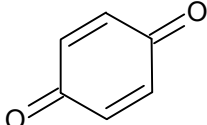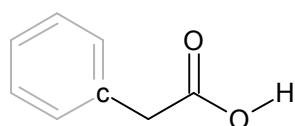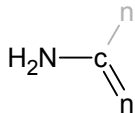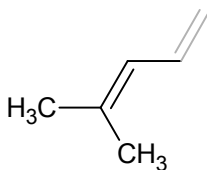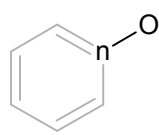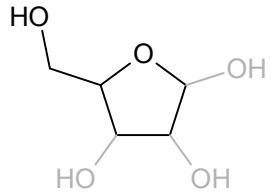
Figure 4.1 Performance of the GALAS model for predicting human plasma protein binding on test set compounds at different reliability thresholds: (A) $RI \geq 0.3$: $N =$ 265; $R^2 = 0.70$; *RMSE* = 0.59 (in log K_{app} units); (B) $RI \geq 0.6$; $N = 107$; $R^2 = 0.82$; *RMSE* = 0.51.

95

Table 4.1 Statistically determined contributions of selected structural fragments to overall extent of protein binding in human plasma. Notation: A – any atom; lowercase symbol – atom within aromatic ring.

Fragment	Description	Coef.	Comment
Positive contributions			
	Dihydropyridine	+0.62	Drugs of this class are known as typical albumin ligands, their %*PPB* values are in the range 95-99%[212]
	Acidic sulphona-mide	+0.61	Y is an electron withdrawing substituent (-CN, -NO$_2$, -SO$_2$R) or an unsaturated bond
	Quinone	+0.51	Presence of planar quinone ring is favorable for albumin binding
	Arylacetic acid	+0.33	This substructure closely resembles the typical pharmacophore of albumin ligands (acidic group located close to a hydrophobic moiety)
	2-Aminopyrimidine	+0.14	Encountered in the common scaffold of zosin class α-adrenoreceptor blockers
	Terpene fragment	+0.13	Neutral lipophilic terpenes, such as etretinate bind with high affinity to plasma lipoproteins
Negative contributions			
	Aromatic N-oxide	-0.86	Presence of polar atoms within the aromatic ring is highly unfavorable for ring localization within hydrophobic binding pocket of the protein
	Pentose	-0.31	Most carbohydrates are very poorly bound in human plasma
	Permanently charged cation	-0.29	Although AAG exhibits a preference towards basic ligands, permanently charged molecules are poorly bound

4.2. Blood-brain barrier permeability

Preliminary analysis. In order to determine the pattern of the BBB permeability-lipophilicity relationship, experimental log PS values were plotted against log $P_{o/w}$ for separate compound classes representing different ionization states at physiological pH. As illustrated in Figure 4.2, (A-C), a wide plateau at log $P_{o/w} \approx 2$–4 is observed for neutral compounds, a narrower one for bases, whereas acidic molecules do not reach the plateau due to the detrimental effect that ionization has on permeability. These results are fully consistent with the findings recently reported by Summerfield et al.[175] – in that study a similar inverted "U"-shaped relationship between *in situ* log PS and calculated log $P_{o/w}$ was obtained with a plateau at the optimum range of log $P_{o/w}$ values. Common explanation of such phenomenon is that hydrophobic drugs are entrapped within lipid bilayer, thus not reaching brain interstitium.

Final results and external validation. Figure 4.2, (D-E) shows the obtained scatter plots for training and internal validation sets. Note that in these plots the lower limit of BBB permeability was set at -5 since log $PS < -5$ are beyond sensitivity of *in situ* perfusion measurements. All predicted values lower than this limit were assigned with log $PS = -5$. In both plots very good correlations are observed with R^2 above 0.8 and $RMSE < 0.5$ log units. The statistical parameters are much better than in any previous methods, and the considered diversity of compounds is much wider. An important factor is that both training and validation sets produced nearly identical results indicating that the model is not likely to be over-fitted. The model was additionally validated using a data set of 50 log PS values from Ref.[175], that was compiled when model development was finished. Several entries were removed from this set since the respective compounds were present in our training set. Model performance for the remaining 43 molecules is depicted in Figure 4.2, (F). As one can see from this figure, previously unseen compounds are predicted with slightly lower R^2, but this may be related to the fact that the log PS scale here is constrained in comparison with the data used for modeling. Summerfield set only contains drugs marketed for CNS indications, and naturally, log PS of the absolute ma-

jority of these compounds varies between -1 and -3. On the other hand, *RMSE* of predictions is on the same level as for the internal test set, showing that the model is suitable for estimating brain penetration potential of new compounds in drug discovery projects.
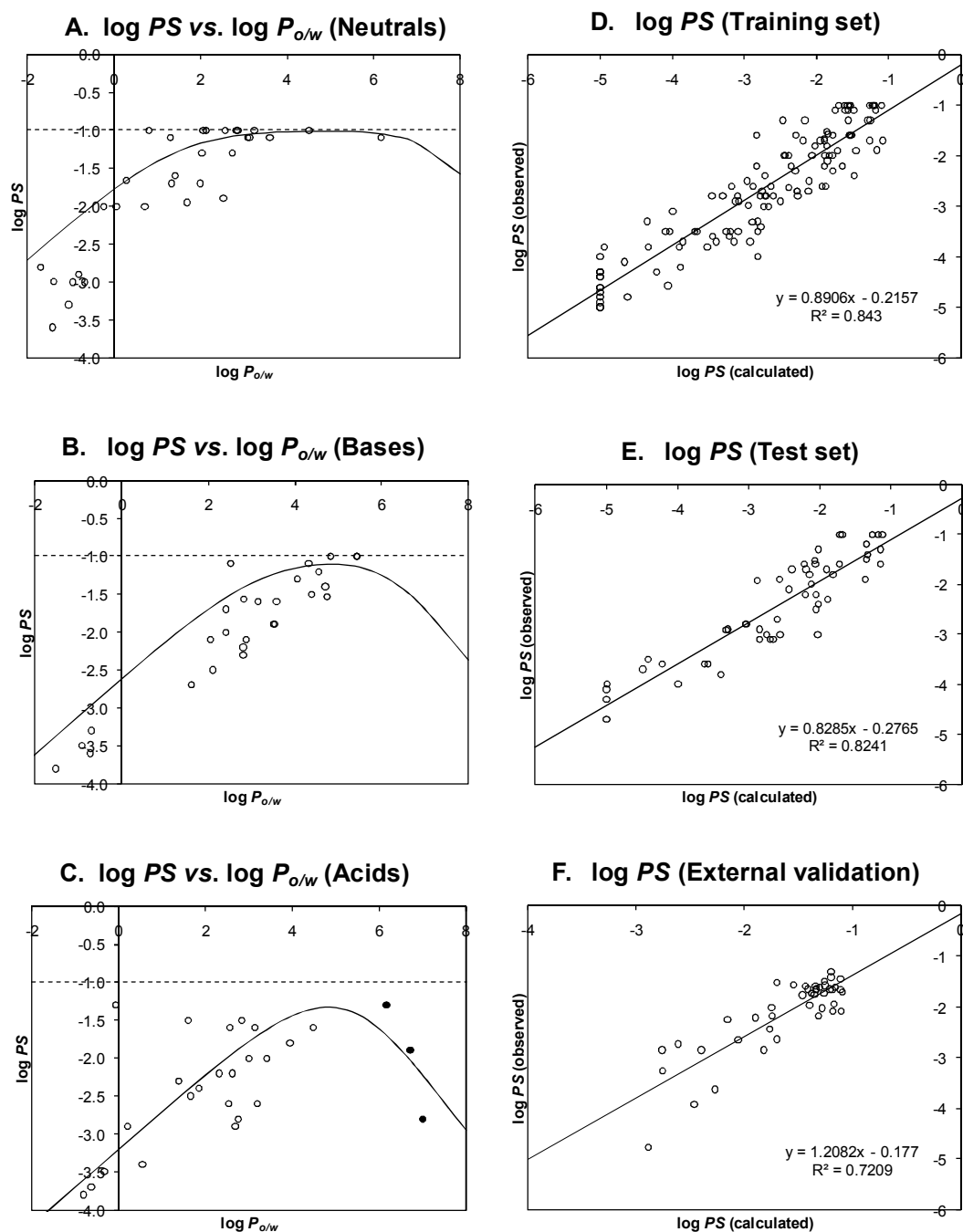
Figure 4.2 log *PS* modeling results: (A-C) Theoretical dependences of log *PS* on log $P_{o/w}$. Dashed lines at log *PS* = -1 indicate flow-limited permeability; filled data points represent fatty acid molecules, that can be retained in the membrane. (D-F) Observed *vs*. predicted log *PS* for training set (*N* = 125; *RMSE* = 0.48), test set (*N* = 53; *RMSE* = 0.49), and external validation set compounds (*N* = 43; *RMSE* = 0.52).

4.3. Human intestinal absorption

Preliminary Analysis. Since the current study focuses on compounds that are predominantly absorbed by passive diffusion and have no known issues such as dose-limited, dose-dependent, or formulation-dependent absorption, their intestinal permeability should follow simple physicochemical rules. The obvious first step in the analysis would be to evaluate the general trend in the relationship between %*HIA* and lipophilicity. For this purpose, quantitative %*HIA* values were plotted against calculated log $P_{o/w}$, as shown in Figure 4.3. In order to eliminate complications due to paracellular transport, only compounds having *MW* > 250 were selected for this plot. As expected for passive transcellular absorption, there is clear evidence for an approximate sigmoid relationship between %*HIA* and log $P_{o/w}$ – lipophilic compounds are in general well absorbed, and hydrophilic compounds – poorly absorbed. Comparison of data points corresponding to different classes of electrolytes demonstrates a significant dependence on charge type. A shift of the curve to higher lipophilicities is more marked for acids than it is for bases. To avoid cluttering, data points corresponding to zwitterionic compounds have not been included in the plot. The shift of sigmoid curve for zwitterions is slightly lower than that for cations. In all cases the effect is much smaller than expected from partitioning into 1-octanol at pH = 7.4. In other words, the intestinal permeability of ionizable compounds is greater than could be calculated using log $D_{o/w}$ as a descriptor. The challenge for a quantitative model was to devise appropriate corrections to log $P_{o/w}$ that would correctly predict passive permeability of electrolytes across intestinal epithelium.

Final results. After initial modeling steps that involved estimating relative contributions of paracellular and transcellular transport and the influence of hydrogen bonding on transcellular permeability, the final model was derived making a connection between permeability and fraction absorbed according to Eq. (3.16). %*HIA* was predicted with $R^2 = 0.93$ for training set compounds. The obtained sigmoid relationships between predicted log P_{eff} and observed %*HIA* values for the entire training set and reference data set from Zhao et al.
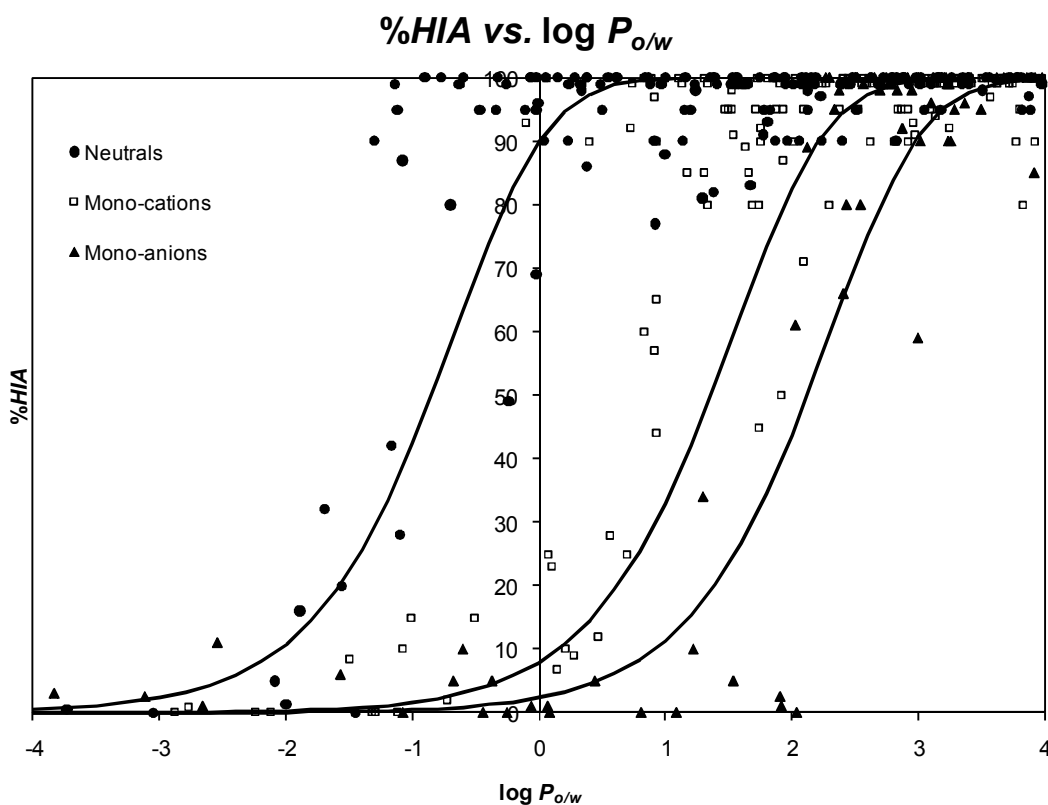
Figure 4.3 Relationship between *%HIA* and log $P_{o/w}$ for compounds with *MW* > 250. Here bases were defined as compounds having pK_a > 7, acids – pK_a < 5.

are shown in Figure 4.4. A good fit is observed in both plots, and the sigmoid curves reflect the experimental observation that jejunal $P_{eff} \approx 1.5$–$2.0 \cdot 10^{-4}$ cm/s (log $P_{eff} > \approx -4$) corresponds to complete absorption[81,213].

Figure 4.5 demonstrates the correlations between calculated log P_{eff} and experimentally obtained jejunal permeability coefficients or absorption rate constants in the external validation sets. In both plots very good correlations are observed with R^2 of log P_{eff} prediction reaching 0.72–0.84, and *RMSE* being as low as 0.35–0.45 log units. The obtained statistical parameters are very good, while the data set used for modeling covers a wide and chemically diverse range of drugs and drug-like compounds. The approach employed in this part of the study is somewhat unconventional – it skips the usual internal validation on a part of the data set not used in modeling. The reason was the lack of high quality experimental *%HIA* data in the intermediate range (30 to 70%). Under these circumstances, splitting the data set into training and test sets was not feasible – both subsets would contain too little data points in the moderate

absorption region, and it would be impossible to evaluate whether the obtained model correctly conveys the deflection of the sigmoid. This issue does not manifest in the external validation sets where the entire range of permeability coefficients (or absorption rate constants) is evenly represented. Also, validating the model with various types of experimental data different from those used for model development provides a better insight on the intrinsic correctness of the algorithm rather than just evaluating the goodness of fit between predicted and experimental %*HIA* data. This is especially true when taking into account that most %*HIA* values are concentrated on the opposite ends of the scale.
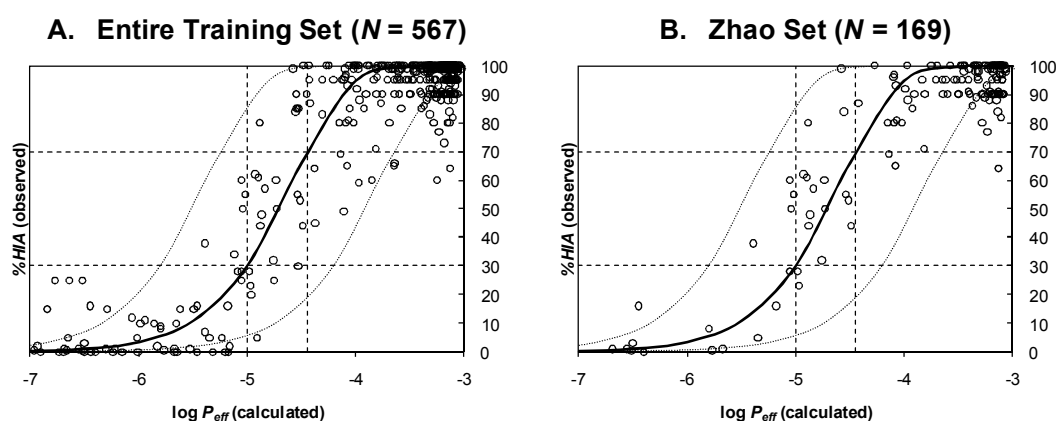
Figure 4.4 Relationship between experimental %*HIA* and calculated log P_{eff}: (A) the entire training set used for modeling; (B) reference data set from Zhao et al. Solid curves represent the simulation sigmoids, dotted curves – 95% confidence intervals of predictions assuming average log P_{eff} estimation error of 0.4 log units.
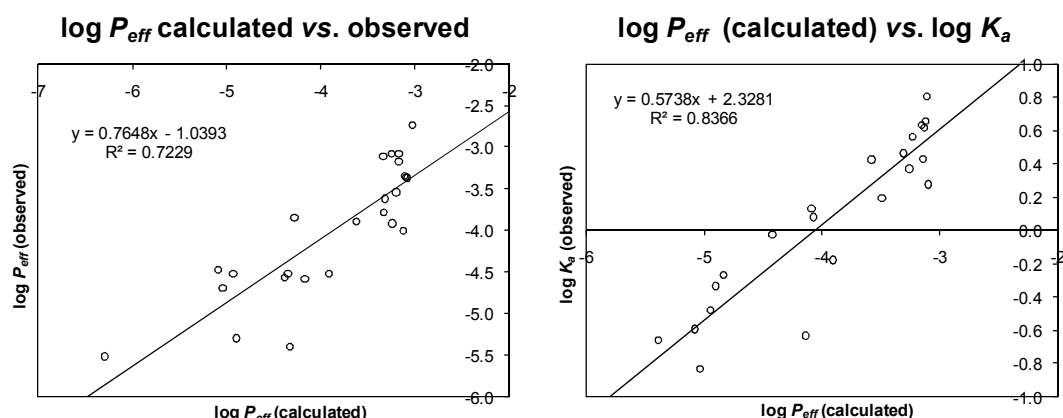
Figure 4.5 External validation results. Correlation between calculated log P_{eff} and: (A) observed log P_{eff} from Lennernäs[81] ($N = 25$, *RMSE* = 0.45); (B) observed absorption rate constants (log K_a) from Linnankoski et al.[183] ($N = 22$, *RMSE* = 0.35).

4.4. Caco-2 permeability

UWL permeability. In Caco-2 monolayers the limiting effect of UWL permeability is much more pronounced than in small intestine and is strongly dependent on the solution stirring. Analysis of P_{UWL} was based on a recent publication by Korjamo et al.[88] where Caco-2 permeabilities of three basic drugs (metoprolol, propranolol, and verapamil) and one acidic drug (ibuprofen) were measured at four different pH values (from 5.8 to 7.4) and varying stirring rates (from 250 to 420 rpm). From these data, the authors estimated the apparent contributions of UWL to the diffusion resistance of the monolayer at various experimental conditions. In the current work, the resulting values were used to derive relevant coefficients in Eq. (3.15). Since Ref.[88] does not provide any values for experiments without stirring, the data from an earlier work by Artursson & Karlsson[214] were utilized to derive a reasonable estimate of parameter K_o. The stirring rates analyzed therein ranged from 0 to 1100 rpm, yet the data depicted in Figure 4.6, (A) do not allow making a definitive conclusion about the pattern of relationship between v and P_{UWL}. The observed dependence is clearly linear up to 600-800 rpm, but it departs from linearity at higher stirring rates. In spite of this, it is quite obvious that in the absence of stirring UWL-controlled Caco-2 permeation rate of lipophilic drugs does not fall lower than $50 \cdot 10^{-6}$ cm/s, which was taken as an estimate of K_o.
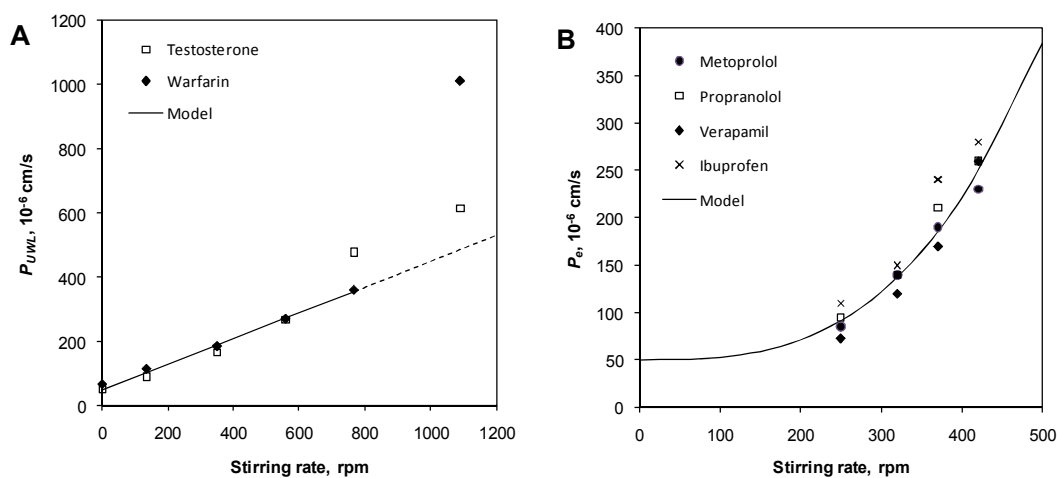
Figure 4.6 Dependence of P_{UWL} in Caco-2 monolayers on stirring rate. (A) data from Artursson & Karlsson[214]; (B) data from Korjamo et al.[88]

Table 4.2 Parameters characterizing paracellular permeability in human jejunum and Caco-2 monolayers.

Parameter	Explanation	Value	
		Jejunum	Caco-2
R	Average pore radius, Å	5.6	4.6
κ	Electrochemical energy parameter	3.0	2.3
$\Delta\psi$	Transepithelial potential drop, mV	80	60
Derived ionization parameters			
Δ_{para}^{+2}	Electrostatic charge effects on	+0.8	+0.7
Δ_{para}^{+1}	paracellular permeability of ionic	+0.5	+0.4
Δ_{para}^{-1}	species calculated according to	–0.8	–0.6
Δ_{para}^{-2}	Eq. (3.11)	–1.8	–1.3

P_{UWL} data reported by Korjamo et al. and the curve fitted according to Eq. (3.15) are shown in Figure 4.6, (B). As one can see from this figure, UWL thickness greatly reduces if the solution is stirred. Application of stirring at 250 rpm would result in two-fold, and at 500 rpm rate – in almost eight-fold increase of UWL-controlled permeability limit. It should be noted that when compounds are classified into quantitative categories on the basis of their absorptive characteristics, "good permeability" is typically defined as P_e(Caco-2) $> 10 \cdot 10^{-6}$ cm/s. Therefore, despite its importance for freely diffusible molecules, permeability limit imposed by UWL is not significant for moderately or poorly permeable compounds.

Paracellular permeability. Paracellular transport parameters obtained by Sugano et al.[202,203] were used in the *%HIA* model described in Section 4.3, and only the scaling factor C_{para} in Eq. (3.11) needed to be determined. The respective parameter values for Caco-2 system were not available, and they were fitted in the current study. The key characteristics of paracellular transport that determine the contribution of this route to overall permeation rate in two absorptive systems are compared in Table 4.2.

The major difference between Caco-2 cell line and human small intestine is reduced leakiness of paracellular pathway in the former. According to our model, which is based on Renkin molecular sieving function, the aqueous pores in the intercellular tight junctions are on average about 1 Å smaller in

Caco-2 monolayers compared to jejunum. This result is in good agreement with recent publications stating that *in vivo* human intestine is indeed more permeable to small hydrophilic molecules[82,83]. For that reason, moderately sized and relatively hydrophilic compounds may be absorbed by different pathways in different cell systems. A good example of such compound is atenolol. This drug is not completely absorbed in human intestine (*%HIA* = 44%)[77], and its absorption predominantly takes place by paracellular route (estimated P_{para} contribution: 71%). However, predicted contribution of paracellular transport in Caco-2 cells at pH = 7.4 and no stirring is only 6%. This is fully consistent with experimental data stating that in Caco-2 system transcellular route dominated, while paracellular pathway accounted for < 20% of atenolol absorption[215].

Different authors report widely varying values of transepithelial potential drop for Caco-2 monolayers. As outlined in Ref.[91], the published values range from 15 to 82 mV. In the current study, the best fit to experimental data was obtained using $\Delta\psi$ = 60 mV corresponding to κ = 2.3. When translated into Δ_{para} values for different ionic forms, this means that the difference between permeabilities of neutral and charges species is not as pronounced as in jejunum, but bications still permeate almost 5 times faster than non-electrolytes.

Transcellular permeability and final model. The implementation of transcellular permeability model did not differ from the approach utilized for predicting *in vivo* intestinal absorption and BBB permeability. The obtained regression coefficients were very similar to the respective values in aforementioned systems. A comparison and a more detailed discussion of the model coefficients will be presented in Section 4.7.1.

The overall model was derived according to Eq. (3.7) incorporating the contributions of both paracellular and transcellular routes as well as UWL resistance. The obtained correlations between predicted and observed log P_e values for training and test sets are depicted in Figure 4.7, while the results of external validation are presented in Figure 4.8. Similarly to blood-brain barrier permeability, log P_e scale is constrained from the bottom with lower limit

about -1.7. This value corresponds to $P_e = 0.02 \cdot 10^{-6}$ cm/s and reflects sensitivity of typical permeability measurements. The upper limit of Caco-2 permeation rate is imposed by UWL, but there is no clear boundary for the entire data set since the collected data points represent a variety of stirring conditions, and measured P_e values as high as $500 \cdot 10^{-6}$ cm/s have been encountered in experiments with vigorous stirring.

The provided plots demonstrate good predictive power of the model, and even in the external validation set, *RMSE* of predictions only slightly exceeds 0.5. Slightly larger average prediction error for the external set can be partly explained by somewhat lower data quality. In some articles stirring rate was not reported, and it was assumed that the experiments had been conducted without stirring. Sometimes, P_e values reported in such studies were higher than would be theoretically possible at 0 rpm. Such discrepancy is evident in Figure 4.8 where a number of data points are concentrated in the region of calculated $\log P_e = 1.5\text{-}1.6$ (P_{UWL}-dependent maximum), although experimental values for these compounds span a much wider $\log P_e$ range. This issue is only observed for freely diffusible molecules and does not manifest when the overall permeation rate is controlled by P_{para} or P_{trans}. Also, even in the validation set, no occurrences of misclassification are observed where poorly absorbed compound is predicted to have good permeability or vice versa.

log P_e (Training set) | **log P_e (Test set)**

y = 0.8681x + 0.0701
R² = 0.8086

y = 0.8394x + 0.0862
R² = 0.7843
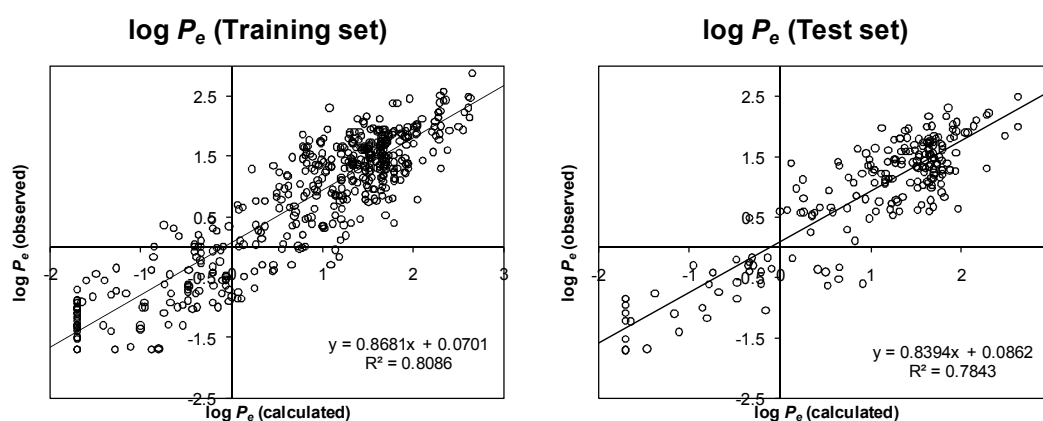
Figure 4.7 Observed *vs.* predicted $\log P_e$ in Caco-2 cell monolayers for training ($N = 473$; *RMSE* = 0.50) and test set compounds ($N = 209$; *RMSE* = 0.47).

log P_e (External validation set)

y = 0.8065x + 0.1102
R² = 0.753

log P_e (observed)

log P_e (calculated)
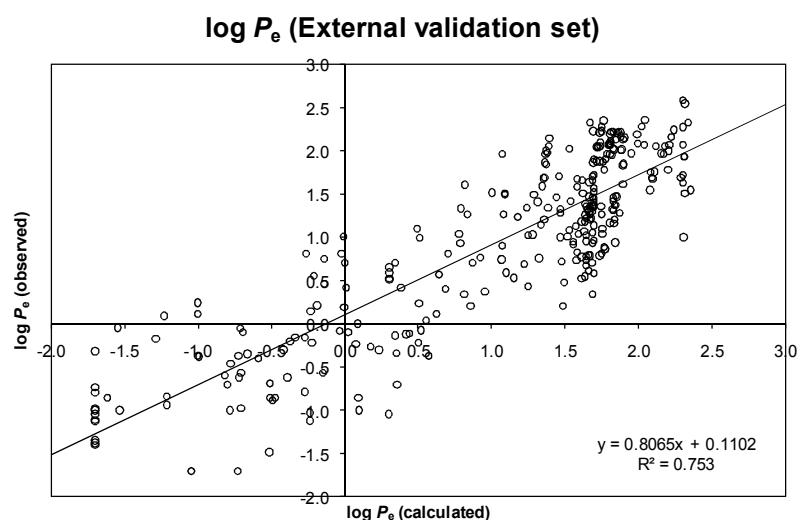
Figure 4.8 External validation results of the presented Caco-2 permeability model ($N = 300$; $RMSE = 0.55$).

4.5. Blood-brain distribution

Preliminary analysis. Prior to data verification or any subsequent fitting procedures, preliminary analysis of the relationship between brain/plasma partitioning and lipophilicity was performed. All available quantitative log BB values collected from the literature were plotted against log $P_{o/w}$, as shown in Figure 4.9, (A). No correlation was observed between log BB and log $P_{o/w}$ with the entire range of variation in blood-brain partitioning evenly represented among compounds with highly varying lipophilicities. Apparently, no tendency in the changes in the extent of brain delivery with varying log $P_{o/w}$ could be identified. The situation noticeably improved when experimental log BB was corrected by the extent of protein binding in plasma, yielding p$f_{u,br}$ values. As illustrated in Figure 4.9, (B), p$f_{u,br}$ reasonably well correlated with log $P_{o/w}$. Almost half of the variation in this parameter among the analyzed compounds could be explained by the differences in their lipophilicities ($R^2 = 0.46$), even despite the fact that the values affected by carrier-mediated transport or other entries of questionable quality were not yet removed from the data set depicted in this figure. These findings clearly demonstrate that log BB is not a measure of lipophilicity-driven BBB permeability, as confusingly stated by some authors[144-147,216], but a more complex parameter. Moreover, QSAR models that directly relate log BB to log $P_{o/w}$ or other lipophilicity measure may lead to de-

106

A. log *BB* vs. log *P*$_{o/w}$

R² = 0.0014

B. p*f*$_{u,br}$ vs. log *P*$_{o/w}$

R² = 0.4628
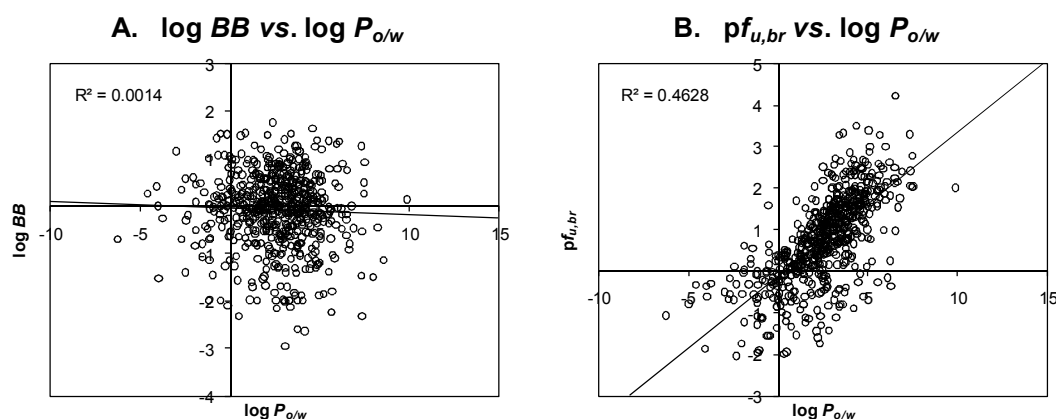
Figure 4.9 (A) Lack of correlation between log $P_{o/w}$ and log *BB* in a diverse data set. (B) Apparent positive correlation between log $P_{o/w}$ and p$f_{u,br}$ for the same compounds.

ceptive conclusions as in a diverse data set containing about 500 compounds these properties do not correlate. Any statistically significant positive coefficients of log $P_{o/w}$ obtained in such models may arise either due to insufficient diversity of the training set compounds or as a consequence of other factors such as interdependences between the used descriptors. In contrast, variations in p$f_{u,br}$ are well reflected by log $P_{o/w}$, supporting the decision to select this particular parameter for modeling.

Statistical characteristics of the final model. Figures 4.10-4.11 illustrate the performance of the obtained model on training set (A) and internal test set (B). A very high correlation is observed between predicted and experimental p$f_{u,br}$ displayed in Figure 4.10 with almost 75% of variance in tissue binding strength explained by the model. The *RMSE* of prediction is close to 0.4 in both training and test sets. log *BB* is predicted with lower R^2 while maintaining the same *RMSE* (Figure 4.11). The mean square error for the prediction of both parameters is identical by definition of the model. Indeed, p$f_{u,br}$ and log *BB* are inter-related by Eq.(3.28), which is used both for deriving p$f_{u,br}$ from original log *BB* data and for back-calculating log *BB* from predicted p$f_{u,br}$, so that the residual for the prediction of these two properties always remains the same.

The observed results highlight the major difference between the two analyzed properties. As discussed above, p$f_{u,br}$ is an unambiguous characteristic of lipophilicity-driven binding to brain tissue constituents, which is well de-

scribed by log $P_{o/w}$ and ionization. On the contrary, log *BB* is a "mixed" parameter that represents a cumulative effect of two different processes shifting the equilibrium in opposite directions. Hydrophobic interactions have a significant influence on drug binding in both plasma and brain tissue, so that the contributions of lipophilicity to these processes compensate each other. As a consequence, brain/plasma ratios of most passive permeants concentrate in the middle of the scale, which is clearly seen in Figure 4.11. In this situation, very high R^2 for predicting log *BB* can only be obtained after artificially expanding the scale by inclusion of extreme values corresponding to facilitated diffusion or efflux. Even though analysis of passively transported compounds with log *BB* values concentrated around zero results in lower R^2, the quality of predictions produced by the current model is sufficiently high judging by the *RMSE*. Given the experimental log *BB* measurement error of about 0.3 log units[142,150] and taking into account the uncertainty in available $f_{u,pl}$ values, it can be concluded that the accuracy of the model is close to that of underlying data. At the same time, the presented results serve as a further indication that brain/plasma partitioning ratio becomes a much more useful estimate of brain delivery potential if it is interpreted in conjunction with $f_{u,br}$ and $f_{u,pl}$.

The results of the external validation using experimental brain tissue binding data are presented in Figure 4.12. The proposed model performs well on both data sets with *RMSE* of predicting p$f_{u,br}$ in mice similar to the value obtained for internal test set, and *RMSE* for rat p$f_{u,br}$ being as low as 0.35 log units. It is important to mention that external validation based on a different assay compared to the training set data is of utmost importance for the model described in this study. The necessity of such kind of validation arises from the fact that the model was parameterized not on the basis of the original blood-brain distribution data but rather on p$f_{u,br}$ values recalculated from log *BB*. In this situation, the ability to predict directly measured unbound fractions in brain supports the feasibility of theoretical approach employed to derive the model and illustrates its intrinsic correctness. This could not be accomplished by a simple evaluation of model performance on a set of new log *BB* constants.

p$f_{u,br}$ (Training set)

y = 0.9553x + 0.0068
R² = 0.7491

p$f_{u,br}$ (Test set)

y = 0.9787x + 0.0203
R² = 0.7377
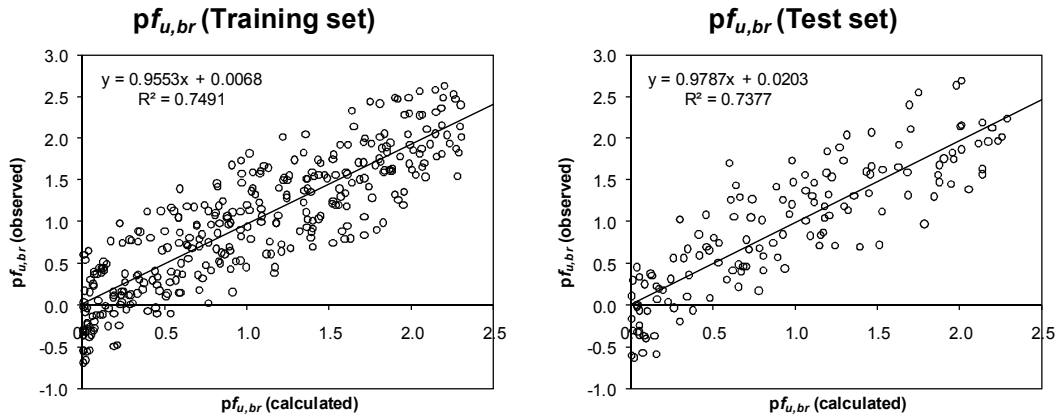
Figure 4.10 Observed *vs.* predicted p$f_{u,br}$ for training set ($N = 329$; *RMSE* = 0.38) and test set ($N = 141$; *RMSE* = 0.39).

log *BB* (Training set)

y = 0.8141x - 0.0154
R² = 0.5176

log *BB* (Test set)

y = 0.7962x + 0.0161
R² = 0.5398
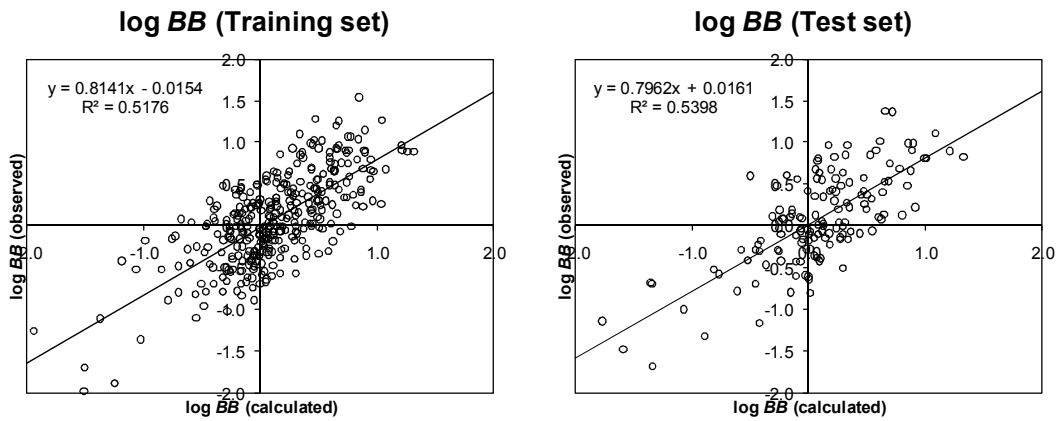
Figure 4.11 Observed *vs.* predicted log *BB* for training set ($N = 329$; *RMSE* = 0.38) and test set ($N = 141$; *RMSE* = 0.39).

p$f_{u,br}$ validation (Kalvass set)

y = 1.0442x + 0.048
R² = 0.7204

p$f_{u,br}$ validation (Summerfield set)

y = 0.9465x + 0.0153
R² = 0.7117
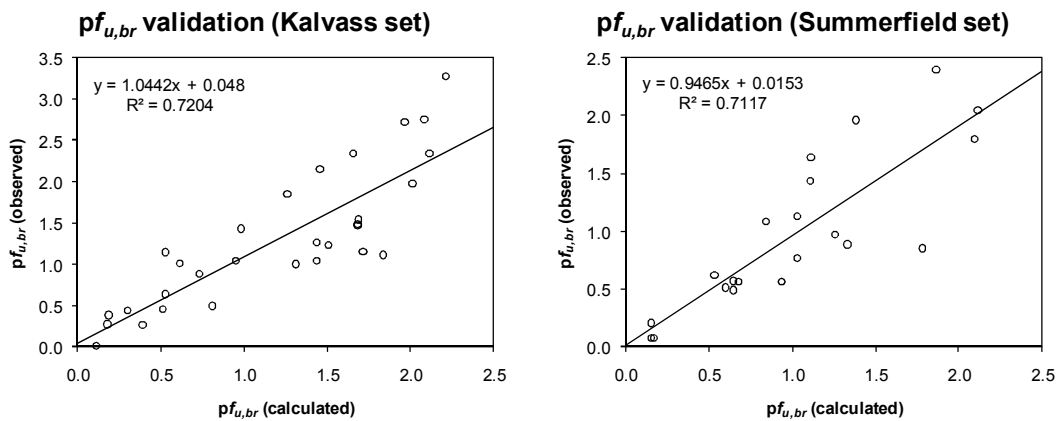
Figure 4.12 Observed *vs.* predicted p$f_{u,br}$ for external validation sets (Kalvass set: $N = 30$; *RMSE* = 0.43; Summerfield set: $N = 21$; *RMSE* = 0.35).

4.6. Volume of distribution

The predictive performance of V_{SS} model developed in the current study is illustrated in Figs. 4.13-4.15. As stated in Section 2.1.1, both training and internal test sets contained only those compounds for which experimental plasma protein binding data were known. Due to this reason, once the compounds actively transported in/out of the tissues and other specific cases were excluded, there were no "problematic" entries with $pf_{u,ti} < 0$ remaining in the internal data set. Overall, the correlation between calculated and observed tissue binding strength was very good with test set parameters even slightly better than training set parameters. The scale of $pf_{u,ti}$ values was much wider than that of $pf_{u,br}$ (≈ 4 *vs.* 2.5 log units), and oppositely to blood-brain distribution, no apparent saturation in tissue binding was observed with increasing log $P_{o/w}$.

Reporting R^2 or *RMSE* of pV_{SS} prediction would not be very informative since V_{SS} values of drugs are usually compared directly, without log-transformation. Instead, accuracy of V_{SS} predictions is often evaluated using *AFE* (Average Fold Error) statistic, calculated as a geometric mean of the prediction residuals:

$$AFE = 10^{\frac{1}{N}\sum \left| pV_{SS,calc.} - pV_{SS,obs.} \right|} \tag{4.1}$$

The presented model predicts V_{SS} with only about 2-fold average error in both training and test sets. Also, predictions for more than 90% of the test set compounds fall within 3-fold error margin indicated by dashed lines in Figure 4.14. Comparable statistical parameters were achieved in both external validation sets, and *RMSE* of predicting drug affinity to tissues did not exceed 0.5 log units in all cases (Figure 4.15). Yet, after transformation of V_{SS} to $pf_{u,ti}$ according to Eq. (3.30), a number of compounds obtained $pf_{u,ti} < 0$ indicating an error in the data. In external set 1 this issue could arise due to the uncertainty in predicted $f_{u,pl}$ values. Nevertheless, the obtained statistical characteristics were reasonably good. These results show that the predictions produced by human plasma protein binding model described in Section 4.1 are suitable for use as rough estimates of $f_{u,pl}$ in models of subsequent drug distribution processes.
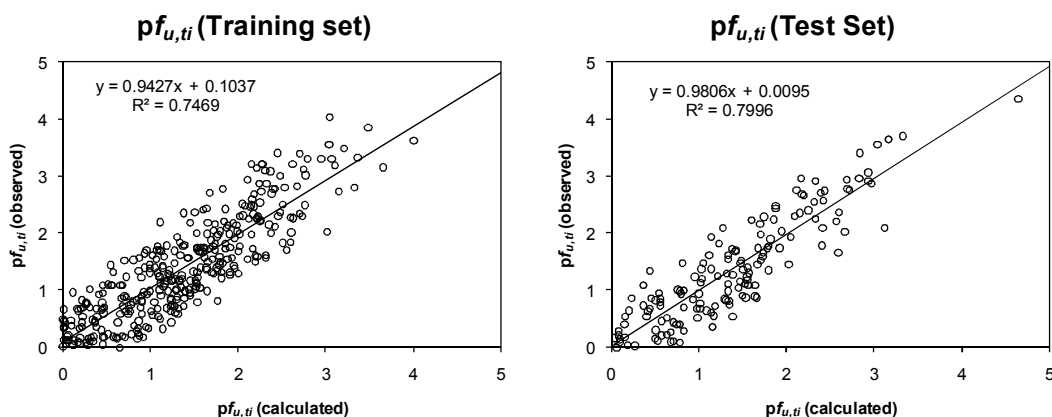
Figure 4.13 Observed *vs.* predicted $pf_{u,ti}$ for training set ($N = 346$; *RMSE* = 0.44) and test set compounds ($N = 150$; *RMSE* = 0.41).
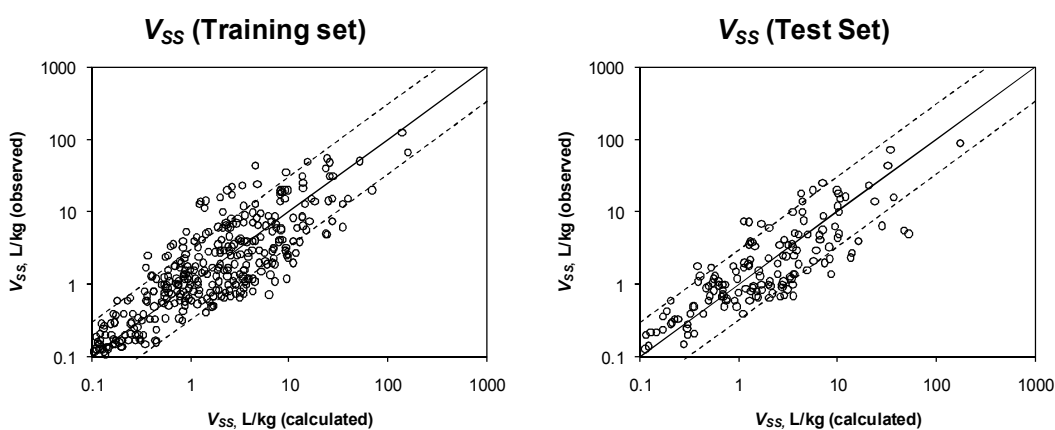
Figure 4.14 Evaluation of model performance for predicting V_{SS} values (training set: $N = 346$; *AFE* = 2.03; test set: $N = 150$; *AFE* = 1.96). Dashed lines correspond to 3-fold error of prediction.
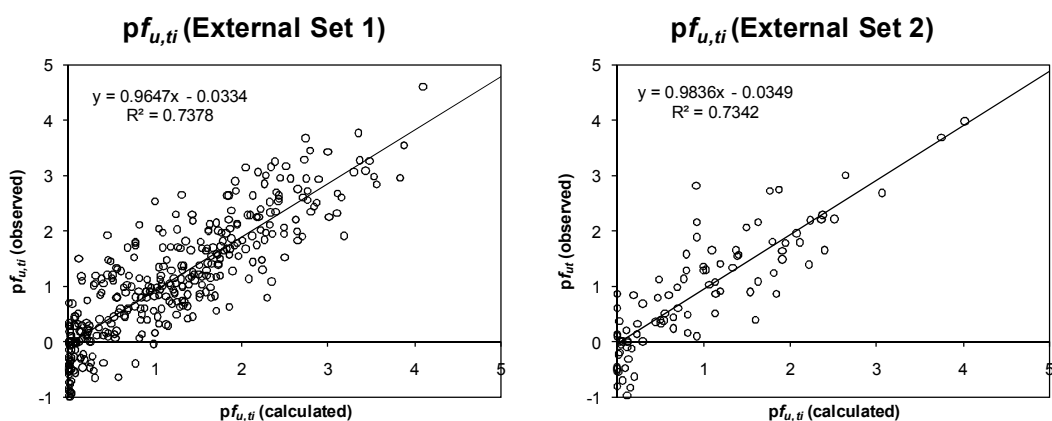
Figure 4.15 Observed *vs.* predicted $pf_{u,ti}$ for the external validation sets (external set 1: $N = 352$; *RMSE* = 0.50; external set 2: $N = 90$; *RMSE* = 0.47).

Entries with $\mathrm{p}f_{u,ti} < 0$ were also present in external set 2, even though for most compounds in this set experimental $f_{u,pl}$ values were available. In a recent publication[211], Waters & Lombardo discussed the presence of such erroneous data in the same data set and outlined that in many cases unexpectedly low V_{SS} could be explained by involvement of active transport processes in the distribution of the respective drugs. These findings show that predictions taking into account the intrinsic relationship between V_{SS} and plasma protein binding help to identify potentially problematic experimental data.

4.7. Generalized models of permeability and distribution

In Sections 4.2-4.6 the predictive performance of the models was discussed without any reference to the fitted descriptor coefficients. The current section will focus on this particular aspect of the models as well as the influence of physicochemical characteristics of drugs on their ability to diffuse through cellular membranes and affinity to tissues.

4.7.1. Membrane permeability

The optimized parameter values obtained in absorptive systems (BBB, small intestine, and Caco-2) are summarized in Table 4.3. Parameters listed in "Scaling" category are needed to bring the data from different systems and expressed by different measurement units onto a single scale. These values are not directly comparable and are provided only for the sake of completeness.

Lipophilicity. Clearly, the most important parameters are related to lipophilicity, which characterizes non-electrostatic properties of cellular membranes. Slope $c_1 = +0.5..+0.7$ indicates that in all analyzed systems transmembrane permeability should increase with increasing log $P_{o/w}$ similarly to liposome/water partitioning[68]. Yet, if brain log PS or intestinal log P_e was modeled by liposome/water partitioning, then the difference between c_1 coefficients (*e.g.*, $\Delta c_1 = c_1^{LIP} - c_1^{BBB}$) would yield systematic deviations equal to $\Delta c_1 \cdot \log P_{o/w}$.

Parameters α and β determine the shape of permeability-lipophilicity curve. The obtained values indicate that log PS reaches an optimum and then

sharply decreases with further increasing log $P_{o/w}$, as illustrated in Figure 4.2. Very similar pattern is observed for Caco-2 permeability, whereas bilinear relationship of this kind could not be established for human intestinal absorption. The reason is the sigmoidal character of the absorption dependence on lipophilicity (Figure 4.3). Fraction absorbed quickly reaches 99.9%, which is experimentally indistinguishable from 100%, and such tendency of drugs to be fully absorbed holds throughout a wide range of log $P_{o/w}$ values. Therefore, the deflection point in permeability *vs.* lipophilicity curve remains beyond observation limit given the available experimental data.

Hydrogen bonding. Slope $c_2 = -0.2\ldots-0.4$ indicates unfavorable contribution of hydrogen bonding to transmembrane permeability in all analyzed systems. The value observed at BBB is twice smaller than in the intestinal epithelium by its absolute magnitude, though this may be attributed to different hydrogen bonding descriptors used in the respective models. In Caco-2 and *%HIA* models the lesser solvating ability of hydrocarbon core of membranes compared to 1-octanol was accounted by inclusion of N_{HD} term. In log PS model the best results were achieved when a composite parameter (N_{HD} + $N_{HA}/2$) was used as a H-bonding descriptor. The definition of this parameter is consistent with Lipinski's "rule of five", stating that drug-like molecular structures should contain no more than 5 H-donors and 10 H-acceptors, *i.e.*, the tolerable number of H-bond acceptors is twice as large as that of donors[217]. In all cases simple hydrogen bond count worked slightly better than Ertl's TPSA or Abraham's *A* and *B*, and was also much easier to calculate. It is not completely clear why the influence of N_{HA} term was only evident at BBB, but this could be related to the difference between membrane-water interfaces in the analyzed systems, namely, the absence of detectable UWL in brain endothelium.

Ionization. Analysis of various electrolyte classes yielded the electrostatic charge parameters of transcellular diffusion ($\Delta_i^{trans} = \Delta_i/c_1$) listed in Table 4.3. The values for the biological systems under consideration including liposome/water partitioning (LIP) were normalized by the corresponding log $P_{o/w}$ slopes (c_1 coefficients) to bring the data from different systems onto a single

scale. When compared among different partitioning systems, these numbers indicate relative electrostatic properties of these systems. As one can see from Table 4.3, in all analyzed absorptive systems the fitted Δ_i^{trans} values were significantly smaller than in octanol, being qualitatively (though not quantitatively) more similar to the liposomes. Liposomes are free of transporters, meaning that relatively high affinity for ionic species is a general feature of phospholipid membranes. Mechanistically, this can be explained by partitioning of charged species through ion-pairing with phosphate groups or counter-ions around these groups. Another possible explanation would be the apparent shift in pK_a of ionizable groups in the vicinity of the membrane. The significance of high affinity of cationic species for negatively charged membranes has been discussed by a number of authors[218-220].

Table 4.3 Optimized parameter values in transcellular permeability models compared to simple *in vitro* partitioning (OCT – octanol, LIP – liposomes, BBB – blood-brain barrier, HIA – human jejunum, Caco-2 – Caco-2 cell monolayers).

Parameters		Explanations	OCT	LIP	BBB	HIA	Caco-2
Scaling	C_{para}	Scaling factor for P_{para}	–	–	–	$7.0 \cdot 10^{-4}$	450
	C_A	Scaling factor between %HIA and P_{eff}	–	–	–	5300	–
	c_o	Intercept for non-electrolytes	0.0		-1.7	-2.5	+2.2
Lipophilicity	c_1	Slope of log $P_{o/w}$	+1.0	+0.8	+0.5	+0.7	+0.6
	α	Fall with log $P_{o/w}$ after optimum (Slope = $c_1 + \alpha$)	–	–	-1.25	–	-1.6
	β	Deflection point in Eq. (3.3)	–	–	5	–	4
H-Bonding	c_2	Slope of H-bonding term	–	–	-0.2	-0.4	-0.4
Ionization	$\Delta_\pm/c_1$	Electrostatic charge effect on P_{trans} of ionic species	-2.5	–	-1.4[a]	-2.0	-1.4
	Δ_+/c_1		-3.1	-1.25	-1.8	-2.3	-1.8
	Δ_-/c_1		-4.1	-2.5	-3.0	-3.3	-2.8
	Δ_{-2}/c_1		–	–	–	–	-4.2

[a] Due to the lack of high quality experimental log PS data for zwitterions, the respective value was extrapolated from Caco-2 system.

Generalized model. Another striking observation is that the normalized Δ_i values for monoprotic electrolytes are very similar among different types of biological membranes analyzed in the current study (brain endothelium, intestinal, and Caco-2 epithelium). The largest $\Delta\Delta_i$ (*i.e.*, $\Delta_+^{HIA} - \Delta_+^{Caco-2}$) is only 0.5 log units, which is well within the uncertainty limit of underlying experimental data. Therefore, it can be assumed that relative permeabilities of biological membrane barriers to ionic species are almost identical, and they are not significantly influenced by the differences in membrane composition.

The set of Δ_i coefficients for 1-octanol/water system is sometimes referred as *diff* ≈ 3-4 rule, *i.e.*, the partitioning ratio of monocations is about 3, and that of monoanions – about 4 log units lower than the value for non-electrolytes[68]. When *in vitro* liposome/water partitioning is considered, the rule slips to *diff* ≈ 1-2. On the basis of Δ_i values obtained in BBB, HIA, and Caco-2 systems, we can state that the electrostatic charge effect observed at cellular diffusion barriers lies in between these two borderline cases and can be approximated by an empirical *diff* ≈ 2-3 rule. Further research is needed to establish the limits of applicability for this hypothesis. The obvious direction for future studies would be to analyze other relevant absorptive systems such as drug permeation across human skin[221] or placental barrier[222].

Table 4.4 Optimized parameter values in tissue binding models (BBB – fraction unbound in brain obtained from brain/blood partitioning ratio, V_{SS} – fraction unbound in tissue obtained from steady state volume of distribution).

Parameters		Explanations	BBB	V_{SS}
Scaling	c_0	Intercept for non-electrolytes	-0.7	-1.7
Lipophilicity	c_1	Slope of log $P_{o/w}$	+0.6	+0.5
	α	Fall with log $P_{o/w}$ after optimum (Slope = $c_1 + \alpha$)	-0.6	–
	β	Deflection point in Eq. (3.30)	5	–
Ionization	$\Delta_\pm/c_1$	Electrostatic charge effect on tissue	-0.2	0.0
	Δ_+/c_1	binding strength of electrolytes	-0.2	0.0
	Δ_-/c_1		-1.0	-1.8
	Δ_{-2}/c_1		–	-2.8

4.7.2. Tissue distribution

Table 4.4 provides a summary of the optimized parameter values obtained in plasma to tissue distribution models. Again, relatively large positive coefficients were observed for log $P_{o/w}$ (c_1 = +0.5..+0.6). This is consistent with the fact that nonspecific hydrophobic interactions with intra- or extracellular lipids represent the main process governing accumulation of drug molecules inside the brain or other tissues. Notably, the log $P_{o/w}$ slopes for lipid/water partitioning (log $K_{b,app}$) are the same as those obtained for transmembrane permeability, indicating a similar contribution of lipophilicity to these processes. We also evaluated if better predictivity could be achieved by inclusion of additional descriptors reflecting hydrogen bonding (N_{HD}, N_{HA}, *TPSA*, *A*, *B*), though none of them resulted in a significant improvement. Similarly to BBB permeability, the predictive model for brain tissue binding contains additional terms accounting for nonlinearity in the partitioning *vs.* lipophilicity relationship. The parameter *β*, that determines the position of the deflection point and optimal log $P_{o/w}$ value, was found to be the same in both systems (*β* = 5). Conversely, the values of the parameter α were markedly different. As mentioned above, log *PS* is characterized by α = −1.25, indicating a sharp fall in permeability after passing the optimal range of lipophilicities. For brain log $K_{b,app}$, this parameter was by its absolute magnitude equal to c_1 effectively canceling out the contribution of the latter. Such kind of apparent saturation of binding could not be determined for tissue affinities obtained from V_{SS} data – all instances of poor distribution in the body among 18 training set compounds with log $P_{o/w}$ > 5 could be reasonably well explained by their strong binding to plasma proteins.

Ionization. The ionization effect constitutes the major qualitative difference between drug binding to brain tissue and permeation across brain endothelium. Since ionized molecules diffuse through the phospholipid bilayer 2–3 orders of magnitude slower than the corresponding neutral species, presence of any functional groups ionized at physiological pH is detrimental for transmembrane permeability. On the contrary, ionization dependence of brain tissue binding is very weak. Only negatively charged ions bind to lipid constituents

of brain tissue noticeably worse than other species, whereas cations and zwitterions are almost indistinguishable from neutral molecules. Such a striking difference between the two brain delivery-related properties can be easily explained by the nature of the modeled processes. To penetrate through the membrane, a molecule must first lose its hydration layer and then pass through the hydrophobic core of the phospholipid bilayer, which is highly energetically unfavorable for any charged species. On the other hand, brain tissue is a non-homogenous target, and once the molecules reach brain interstitial fluid, they may interact with different tissue constituents, leading to similar overall affinities of various ionic species. The molecules bearing no net charge preferentially bind to neutral lipids, whereas additional electrostatic interactions with acidic phospholipids compensate for weaker binding strength of cations[207]. Anions undergo unfavorable electrostatic repulsion from negatively charged surface of phospholipid membranes, that results in lower affinity to lipids. Brain tissue contains appreciable levels of albumin which lowers the free fraction of anionic compounds in brain[208]. Due to the contribution of albumin binding, the presence of negative charge has a lesser effect on brain distribution of acidic drugs than could be expected from lipid binding alone. This is reflected by $\Delta\_/c_1$ value of only -1 log unit.

There was also a clear difference between the two drug distribution systems – BBB and V_{SS}. An appreciable shift of -1.8 units was obtained for tissue binding of mono-anions after the analysis of V_{SS} data. This is understandable as physiological Øie-Tozer equation already accounts for distribution of albumin within the extracellular fluid, and $f_{u,ti}$ term therein represents purely lipid/water partitioning without any additional factors compensating for electrostatic repulsion of anions from phospholipid membranes.

Case study. An additional discussion is necessary with regard to the distribution of basic compounds. As one can see in Table 4.4, the predicted contributions of cations and zwitterions to the overall affinity to tissue are similar to the contribution of neutral species, like in the case of drug distribution to the brain. However, it is generally recognized that basic drugs actually distribute

better than neutral molecules. This statement does not raise any contradictions by itself – indeed, very low volumes of distribution usually originate from low free fractions in plasma, and bases show a tendency to have weaker protein binding. Yet, a number of publications indicate that not only V_{SS} but also the so called "unbound volume of distribution" ($V_{u,SS} = V_{SS}/f_{u,pl}$) of basic compounds is higher, even though this parameter is free of plasma binding effects. For example, in a recent publication by Obach et al.[174], median V_{SS} of bases in a database consisting of 670 drug-like compounds was 2.1 times, and median $V_{u,SS}$ – about 1.7 times larger than that of neutrals. Furthermore, in Ref.[32], the authors present an example how the volume of distribution can be improved by introducing a basic moiety. In this example the initial antibiotic erythromycin (monocation) had $V_{u,SS} = 4.8$ L/kg, while its derivative azithromycin containing an additional amino group (bication) had much larger $V_{u,SS} = 62$ L/kg. Since $V_{u,SS}$ is roughly equivalent to the p$f_{u,ti}$ term considered in the current study, the described examples imply that positively charged species exhibit higher affinity to lipids, which is inconsistent with modeling results presented in Table 4.4.

A more detailed analysis of the data from Refs.[32,174] reveals that in both cases the apparently better tissue distribution of basic drugs can be attributed to log $P_{o/w}$ rather than ionization. If we consider physicochemical properties of drugs included in the database of Obach et al., it becomes obvious that basic drugs are on average significantly more lipophilic than neutrals. As shown in Figure 4.16, (A), larger median $V_{u,SS}$ of bases is compensated by the difference of lipophilicities between these two drug sets. A similar situation is observed in the example with erythromycin, illustrated in Figure 4.16, (B). This compound has experimental log $P_{o/w} = 2.54$. Incorporation of a tertiary amino-group within the 14-membered ring not only introduces another ionizable center but also leads to a substantial increase in lipophilicity – experimental log $P_{o/w}$ of azithromycin is 4.02. According to our model, this 1.5 log unit log $P_{o/w}$ difference translates into 6-fold difference in tissue binding strength, which is in fairly good agreement with given $V_{u,SS}$ values.
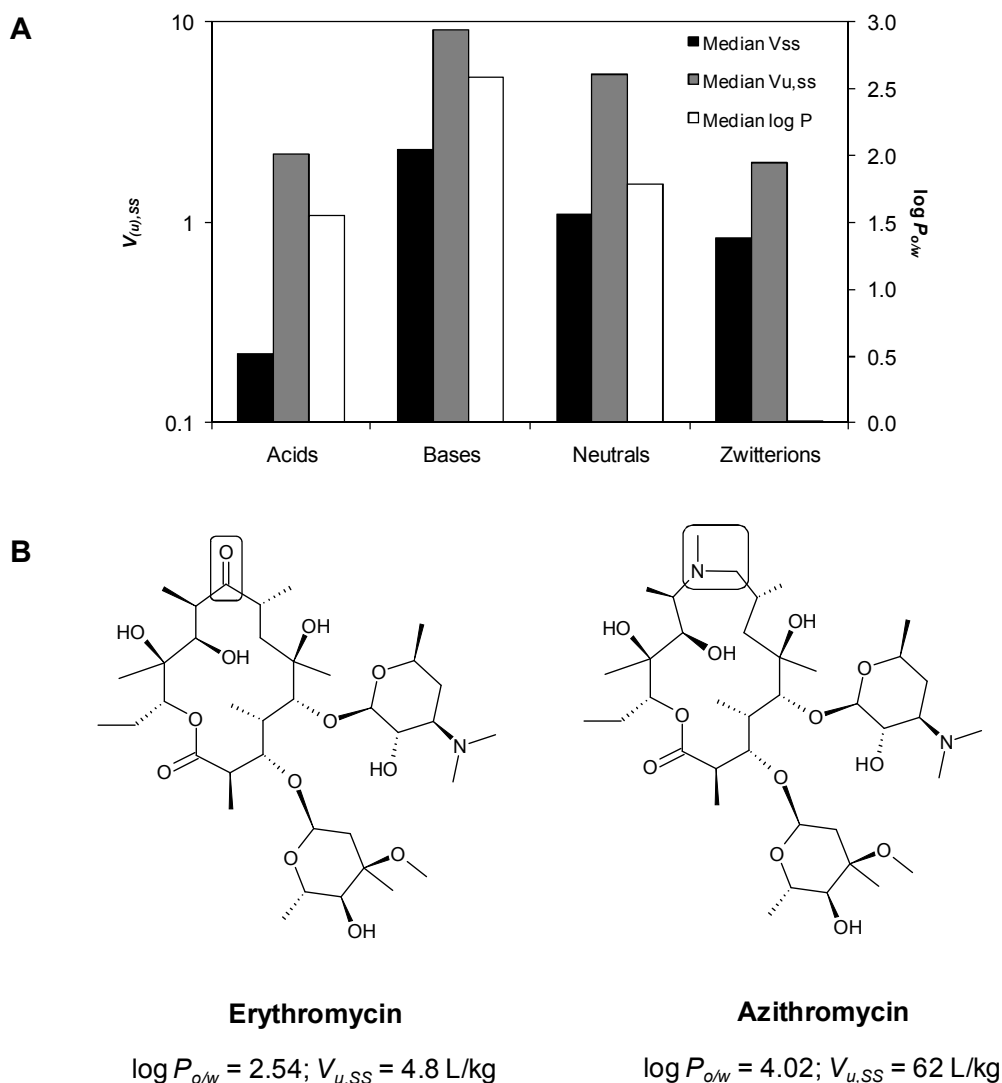
Figure 4.16 Relationship between distribution volume and ionization – a case study: (A) median V_{SS} and $V_{u,SS}$ vs. log $P_{o/w}$ values of drug classes representing different ionization states in the data set analyzed by Obach et al.[174]; (B) improvement in $V_{u,SS}$ of azithromycin compared to erythromycin achieved by expansion of the ring system.

Generalized model. As was the case with absorptive systems, a hypothesis could be raised stating that *in vivo* phospholipid/water partitioning processes should have a similar pattern of ionization dependence. In fact, the tissue binding component in the volume of distribution model (Eq. (3.30)) already provides an example of such generalization since V_{SS} reflects drug distribution into a wide range of tissues. In this model ionization only plays a role in partitioning of anions, which are about 100 times less susceptible to lipid binding than other ionic species, so that we can devise an empirical *diff* ≈ 2 rule for acidic drugs. Brain/plasma partitioning deviates from this rule due to the ne-

119

cessity of additional correction for drug binding to extracellular albumin. Therefore, to test the validity of the generalized model, another endpoint representing lipid/water partitioning unaffected by other processes was needed.

For this purpose, we used the data characterizing *in vitro* binding of drugs to liver microsomes. Non-specific microsomal binding is an important factor to consider in kinetic studies of drug metabolism. Fraction unbound in microsomal incubation ($f_{u,inc}$) determines the difference between the compound's true biotransformation rate and experimentally measured microsomal clearance[223-225]. Austin et al.[224] determined rat $f_{u,inc}$ values for 56 drugs and pointed out the clearly linear relationship between affinity to microsomes ($\log K_{b,inc} = \log (1 - f_{u,inc})/f_{u,inc}$) and lipophilicity. The authors also noticed the lack of ionization dependence for all drug classes except acids and suggested that this effect could be modeled by substituting octanol/water $\log P$ with $\log D$ for acidic drugs (Eq. (4.2)). However, even better correlation could be obtained if microsomal $\log K_{b,inc}$ was described by our ion-specific model applying the *diff* $\approx$ 2 rule outlined in Eq. (4.3):

$$\log K_{b,inc} \sim \begin{cases} \text{acids} : \log D_{o/w} \\ \text{non} - \text{acids} : \log P_{o/w} \end{cases} \tag{4.2}$$

$$\log K_{b,inc} \sim \begin{cases} \text{acids} : \log P_{o/w} - 2 \\ \text{non} - \text{acids} : \log P_{o/w} \end{cases} \tag{4.3}$$
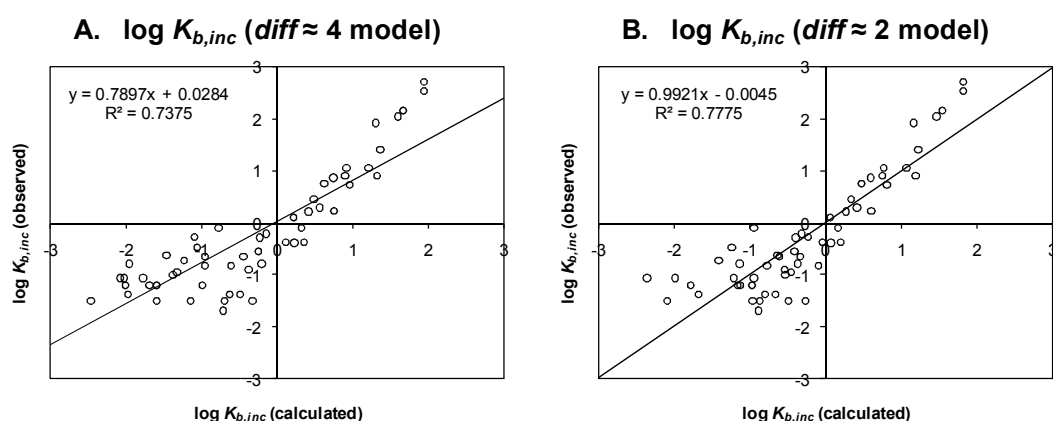
Figure 4.17 Observed *vs.* predicted apparent microsomal binding constants $\log K_{b,inc}$: (A) model obtained by Austin et al. (*RMSE* = 0.63); (B) model derived in the current study (*RMSE* = 0.52). Experimental values taken from Ref.[224]

Observed log $K_{b,inc}$ and the values predicted using both approaches are listed in the Appendix (Table A3), while Figure 4.17 provides a graphical view of the change in predictive power introduced by applying $diff \approx 2$ rule for acids instead of $diff \approx 4$ rule implied by the use of octanol/water log D. If model predictivity is judged solely by R^2, there is very little improvement over $diff \approx 4$ model. In both cases, good overall correlation is ensured by large number of neutrals and bases, that are well described by log $P_{o/w}$. On the other hand, Austin's model shows a clear tendency of underpredicting microsomal binding strength of acidic compounds. This results in a shallower slope of the regression line, while in our model, the regression line coincides with the line of unity, and the observed $RMSE$ of predictions is appreciably lower. These findings support the validity of generalized model approach presented here and demonstrate that its use is not limited to V_{SS} predictions.

4.8. CNS access

The ultimate goal of QSAR analysis focusing on BBB penetration is the ability to discriminate between CNS active and inactive molecules. Obviously, that particular task is extraordinarily difficult to accomplish because apart from other things it would require knowing which target(s) in the brain the compound of interest would bind to and what would be the interaction strength. However, it is possible to derive a simpler model focusing solely on brain delivery. Such model would predict whether the analyzed compound is transported into the brain effectively enough for exhibiting CNS activity, or in other words, whether the compound is sufficiently accessible to CNS. An attempt to develop such classification system is presented below.

The theoretical foundation lying behind our classification approach builds up on the findings recently reported by X. Liu and coworkers. Liu et al. applied physiologically based pharmacokinetic (PBPK) models to study the time needed to reach brain equilibrium[226,227]. They suggested that brain equilibration time may be quantified using a new parameter termed intrinsic equilibration half-life ($t_{1/2eq,in}$), defined as the time needed to reach 50% of the brain/plasma

concentration ratio observed at equilibrium. According to their model, $t_{1/2eq,in}$ is inversely related to the product of permeability and unbound fraction in brain ($PS \cdot f_{u,br}$):

$$t_{1/2eq,in} = \frac{\ln 2}{k_{out}} = \frac{V_{br} \ln 2}{PS \cdot f_{u,br}} \qquad (4.4)$$

In Eq. (4.4) k_{out} is the brain elimination rate constant, and V_{br} is the physiological volume of the brain.

Introducing this new parameter allows making a logical connection between quantitative BBB transport characteristics. Indeed, log ($PS \cdot f_{u,br}$) could be viewed as brain/plasma equilibration rate (inverse of equilibration time), while log BB corresponds to the extent of brain delivery when the respective equilibrium is reached. Here we will demonstrate that a linear combination of two aforementioned parameters suffices for devising a simple, yet accurate CNS access classifier.

The analysis was performed using qualitative BBB permeability data taken from the study by Adenot and Lahana[151] (details provided in Section 2.1.2). The values of log BB, log PS, and $f_{u,br}$ for these compounds were calculated using the predictive models described in Sections 4.2 and 4.5. In a preliminary publication[201], we demonstrated that the data points distributed highly unevenly when log BB was plotted against log ($PS \cdot f_{u,br}$). Figure 4.18, (A) illustrates the clear separation of the two compound classes: BBB+ compounds are mostly located above the discrimination line, while BBB– molecules dominate below the line.

Table 4.5 Accuracy testing of CNS Access classifier: contingency table.

Observed\Predicted	BBB+	BBB–	Accuracy
BBB+	1239	72	94.5% (Sensitivity)
BBB–	26	244	90.4% (Specificity)
Accuracy	97.9% (Positive precision)	77.2% (Negative precision)	93.8% (Overall)

Numerically, the discrimination line in Figure 4.18 is defined as follows:

$$S_{BBB} = \log (PS \cdot f_{u,br}) + \log BB = -3.5 \qquad (4.5)$$

Here the sum of two involved quantitative BBB transport characteristics is designated CNS Access Score (S_{BBB}). Compounds obtaining $S_{BBB} \geq$ cut-off are considered sufficiently accessible to CNS, or BBB-penetrating (BBB+), whereas compounds with lower scores are assigned to BBB– category. The optimal cut-off value was determined using linear discriminant analysis technique. As outlined in section 2.3.3, the classification threshold in LDA model depends on prior probabilities that the compound would belong to a particular category. Obviously, using class proportions in the data set as prior probabilities (the default option in R program) would not be a feasible solution as BBB– compounds are clearly underrepresented in the analyzed set. A more reasonable estimate of encountering a brain-penetrant molecule among drug-like compounds could be obtained from raw WDI data used to derive this set[151]. 1889 of the total 4919 compounds in WDI database had neuropsychiatric activity – this would give prior probability $\pi_{BBB+} = 0.384$. After removal of questionable entries, 1335 BBB+ molecules remained ($\pi_{BBB+} = 0.271$). With these estimates, the classification cut-off was quite stable ranging from -3.6 to -3.4, and the mean value of -3.5 was accepted. This choice of S_{BBB} cut-off resulted in a well-balanced model with both sensitivity and specificity higher than 90%. Overall, the proposed classification system was able to assign correct CNS access category for almost 94% compounds in the data set, as shown in Table 4.5. Wilk's U-statistic obtained by the classifier was equal to 0.54. This is much lower than the value reported by Vilar et al.[153] ($U = 0.70$) indicating the superiority of our approach.

Note that the observed sensitivity of classification is better than specificity, so that performance of the classifier well reflects the nature of the data. Despite the efforts employed by the authors of the used data set, several entries where permeable compound is classified BBB– due to absence of CNS activity may still be present. In these circumstances, a higher proportion of false positive predictions is expected since the classifier can only estimate if the compound is

sufficiently accessible to CNS for an effect to occur, but not whether the molecule will indeed be active once it enters the brain.

Theoretical justification. S_{BBB} was deliberately expressed assuming equal weights of both quantitative transport parameters because this allows for an unambiguous theoretical interpretation. Indeed, if brain/plasma equilibration rate and log BB are split into the underlying components, the rearranged equation yields:

$$
\begin{aligned}
S_{BBB} &= \log\left(PS \cdot f_{u,br}\right) + \log\frac{f_{u,pl}}{f_{u,br}} = \\
&= \log PS + \cancel{\log f_{u,br}} + \log f_{u,pl} - \cancel{\log f_{u,br}} = \\
&= \log PS + \log f_{u,pl} = \log\left(PS \cdot f_{u,pl}\right)
\end{aligned}
\tag{4.6}
$$

The brain free fraction is eliminated from the above expression, and it becomes evident that the overall score depends only on BBB permeation rate given by log PS and unbound fraction in plasma. The term log ($PS \cdot f_{u,pl}$) could be viewed as the compound's apparent *in vivo* permeability across BBB, in contrast to simple log PS, which corresponds to the *intrinsic* BBB permeation rate measured *in situ*. Such interpretation is supported by the fact that the former term is used in permeability calculations when brain uptake experiments are conducted in presence of endogenous plasma or plasma proteins specifically added to the perfusion buffer[228]. Since only the unbound drug is available for diffusion, the protein binding effect is accounted by replacing the PS term in Renkin-Crone equation with $f_{u,pl} \cdot PS$:

$$
K_{in} = F\left(1 - e^{-\frac{f_{u,pl}PS}{F}}\right)
\tag{4.7}
$$

As follows from Eq. (4.6), *in vivo* BBB permeability reflected by log ($PS \cdot f_{u,pl}$) unambiguously defines S_{BBB} and is the sole parameter required for the assignment of qualitative categories. However, extracting separate contributions of kinetic and thermodynamic components and visualizing the data on a two-dimensional scale, as demonstrated in Figure 4.18, provides additional useful information, such as the ability to estimate onset and duration of action of CNS drugs.
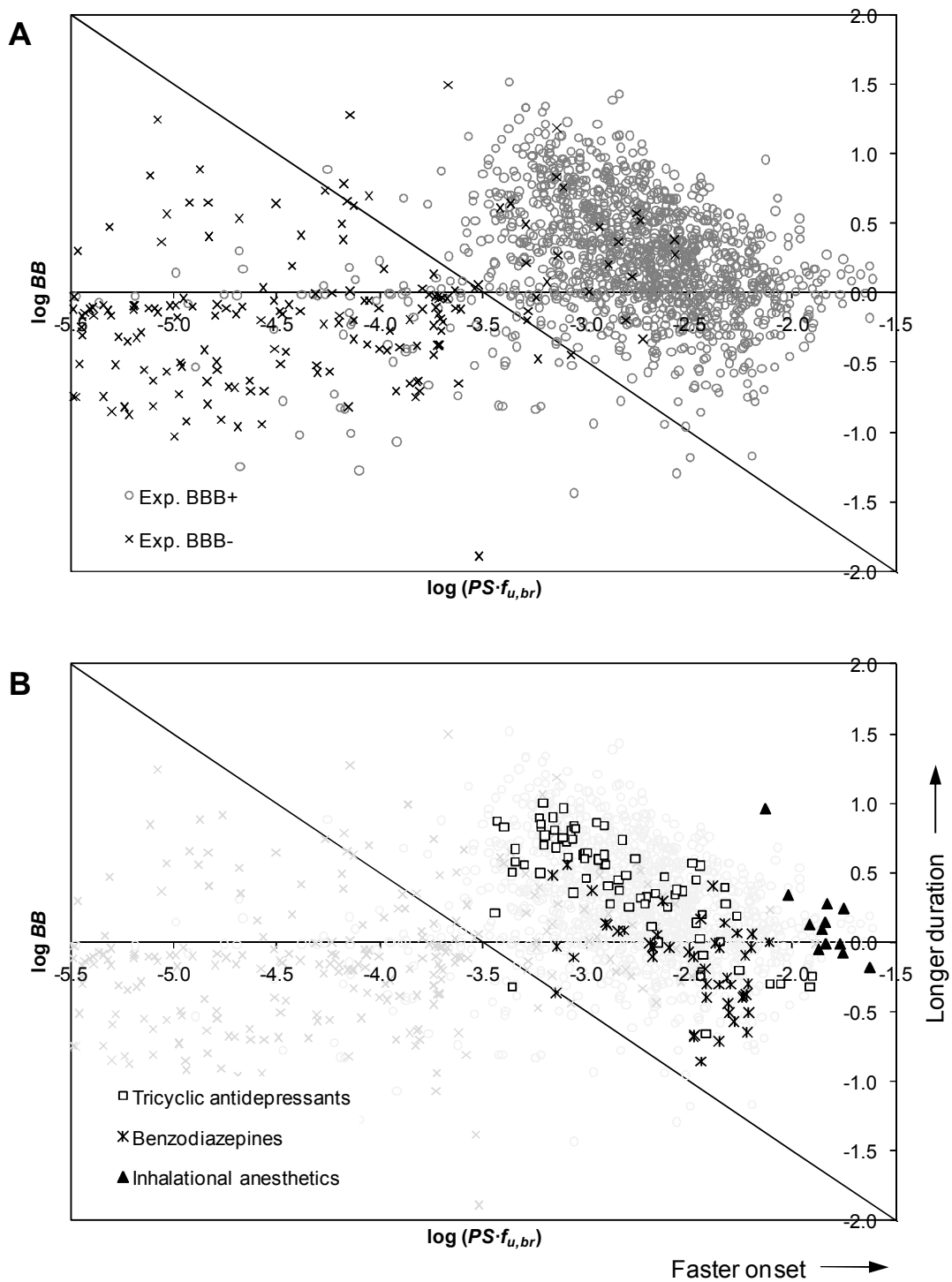
Figure 4.18 Classification of drugs as BBB-penetrating (top right) or BBB-non-penetrating (bottom left) according to predicted values of $\log (PS \cdot f_{u,br})$ and $\log BB$. (A): The entire data set with experimentally assigned classes marked by circles or crosses; (B): the same plot with overlaid data points corresponding to specific drug classes.

Onset and duration of action. A few important observations are related to the category of BBB+ drugs. As outlined in Refs.[226,227], $t_{1/2eq,in}$ term reflects the onset of central action. For example, theobromine ($t_{1/2eq,in} \approx 0.1$ h) was experimentally confirmed to act with faster onset than fluoxetine ($t_{1/2eq,in} \approx 1$ h) despite having much lower BBB permeability. This can be explained by stronger brain tissue binding of fluoxetine that compensates for the difference in permeability and prolongs the time needed to reach brain/plasma equilibrium. The predictions produced by our predictive models are consistent with the experimental observations: predicted theobromine log ($PS \cdot f_{u,br}$) = -2.8; fluoxetine log ($PS \cdot f_{u,br}$) = -3.2. Similarly, log BB parameter can be considered to represent duration of action. Indeed, high log BB corresponds to high brain/plasma concentration ratio and is also associated with good affinity to brain lipids. Large concentrations of active compound bound to brain tissue may serve as a depot effectively prolonging drug action.

Based on the above considerations, the proposed CNS access classifier can be used not only for discerning compounds with good brain delivery potential from those restricted to periphery but also for a rough qualitative estimation of the pharmacokinetic profile of CNS drugs in terms of onset and duration of action. Compounds that are close to the discrimination line and have moderate $PS \cdot f_{u,br}$ values would be expected to act with slower onset compared to those concentrated in the right-hand side of the plot. Accordingly, the molecules in the upper part of the plot are likely to exhibit more prolonged effect compared to the drugs with low log BB values. This concept is best illustrated in Figure 4.18, (B) where point swarms representing three different classes of CNS drugs are highlighted:

1. **Tricyclic antidepressants** are usually characterized by very long duration of action (up to several days)[169]. Almost all drugs of this class concentrate in the upper part of the plot with log BB > 0. On the other hand, the majority of tricyclic antidepressants are very lipophilic, resulting in low free fractions in brain and slower onset of action despite their good permeability across BBB.

2. **Benzodiazepines** represent another class of CNS drugs that vary by duration of action (from ultra-short to long acting), but typically have a faster onset[169]. This is well reflected by predicted BBB transport parameters as this drug class is shifted towards higher brain/plasma equilibration rates while spanning a wide range of log BB values.

3. **Inhalational anesthetics** provide an example of a drug class with extremely rapid onset of action[229]. These are usually small molecules predicted to be highly permeable across BBB, yet having a considerable unbound fraction in brain resulting in very high log $(PS \cdot f_{u,br})$ values.

One should keep in mind that Eq. (4.4) is defined under idealized conditions assuming that the drug is administered by constant intravenous infusion and its plasma concentrations remain unchanged. The expression of the actual brain equilibration half-life $(t_{1/2eq})$ after a single dose would differ from Eq. (4.4) by an additional term k_{el} in the denominator, corresponding to plasma terminal elimination rate constant[226]:

$$t_{1/2eq} = \frac{\ln 2}{k_{out} - k_{el}} \tag{4.8}$$

Certain drugs (typically hydrophilic molecules falling into BBB– category) may have $k_{out} < k_{el}$, meaning that their brain/plasma concentration-time curves do not reach a plateau after a single dose administration. Therefore, associating $PS \cdot f_{u,br}$ parameter values with brain/plasma equilibration rate is only meaningful for BBB+ compounds.

Brain exposure. Further analysis involving brain exposure data was conducted to verify the validity of the presented approach. Steady-state unbound brain/plasma partitioning ratio K_p^{uu} introduced in Section 3.3 is recognized as a reliable quantitative characteristic of brain exposure. Generally, this parameter is considered not suitable for QSAR focusing on passive diffusion. It is presumed that if the drug passively equilibrates across BBB membranes, its unbound concentrations in both compartments would be equal ($K_p^{uu} = 1$), and all deviations from unity could be attributed to the involvement of carrier-mediated transport[205]. In spite of this, apparent K_p^{uu} may be less than unity

even for passively transported compounds if the respective molecule is not permeable enough to reach steady-state conditions. For example, hydrophilic carbohydrate molecules bind neither to plasma nor to brain components and according to Eq. (3.22) should distribute evenly between the two compartments. Yet, BBB permeability of such compounds is negligible, which results in K_p^{uu} close to zero and very low apparent log BB. Therefore, K_p^{uu} could be expected to correlate with BBB permeability to a certain extent.

Recent advances in experimental techniques such as brain homogenate binding[176] and brain slice[230] methods enabled determination of K_p^{uu} values for a reasonably large number of drugs. Fridén et al. reported *in vivo* K_p^{uu} for 41 compounds measured in rats using brain slice method[231] (see Table A4 in the Appendix). In order to evaluate how well these data agree with criteria used here for the assignment of qualitative CNS access categories, we correlated experimental log K_p^{uu} with log $(PS \cdot f_{u,pl})$. As shown in Figure 4.19, (A), log PS alone allowed obtaining a reasonably good estimate of the unbound partitioning ratio. The correlation markedly improved when PS term was corrected for the extent of plasma protein binding given by $f_{u,pl}$ (Figure 4.19, (B)).

Several conclusions can be made from these results. First, *in vivo* BBB permeability log $(PS \cdot f_{u,pl})$ serves as a powerful discriminant function because it provides a good estimate of brain exposure. Second, a drug can be considered accessible to CNS if it is exposed to brain well enough for reaching steady-state conditions.

A. log K_p^{uu} vs. log PS B. log K_p^{uu} vs. log $(PS \cdot f_{u,pl})$
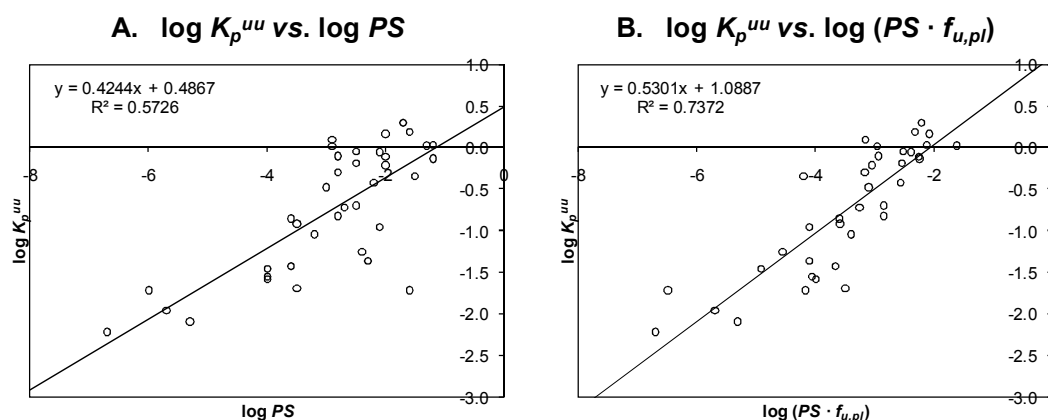
Figure 4.19 Correlation between experimental log K_p^{uu} and BBB permeation rate: (A) intrinsic permeability (log PS); (B) *in vivo* permeability (log $PS \cdot f_{u,pl}$).

CONCLUSIONS

1. GALAS modeling methodology was successfully adapted for predicting the extent of protein binding in plasma (%*PPB*). The resulting predictions were of sufficient quality for use as a descriptor in models of subsequent drug distribution processes.

2. Passive permeability of drug-like compounds across brain endothelium and intestinal epithelium could be described using only physicochemical descriptors that reflect lipophilicity, ionization, hydrogen bonding potential, and molecular size.

3. A similar pattern of permeability-ionization dependence was observed in all considered systems: ionized species permeate through biological membranes 2 to 3 orders of magnitude slower than neutral molecules.

4. The major part of variation in tissue affinity of drugs was explained by differences in their lipophilicity. The observed ionization dependence is only significant for acidic compounds, while bases and zwitterions bind to lipid constituents of tissues as readily as neutral molecules.

5. On the basis of the above findings, non-linear QSAR models were derived for rodent blood-brain permeability (log *PS*), human intestinal absorption (%*HIA*), Caco-2 permeability (log P_e), rodent brain/plasma partitioning ratio (log *BB*), and human steady-state volume of distribution (V_{SS}). Mean square prediction errors of all considered endpoints did not exceed 0.4-0.5 logarithmic units.

6. A linear combination of brain/plasma equilibration rate and equilibrium brain/plasma partitioning ratio allowed classifying drug-like compounds by their accessibility to central nervous system with 94% overall accuracy.

7. Further analysis of the proposed classification model revealed that a drug can be considered accessible to central nervous system if it is permeable enough to reach steady-state in brain.

APPENDIX

Table A1 Observed and predicted parameter values for V_{SS} external validation set (2).

No.	Compound name	log $P_{o/w}$	$f_{u,pl}$	V_{SS}	Obs. p$f_{u,ti}$	Calc. p$f_{u,ti}$
1	Abanoquil	2.92	0.029	6.3	2.72	1.77
2	Adefovir	-1.82	0.96	0.42	-0.23	0.01
3	Alizapride	1.60	0.30	1.6	1.08	1.13
4	Anhydrovinblastine	4.69	0.029	12	3.01	2.65
5	BB 83698	1.03	0.76	1.1	0.48	0.82
6	Bisaramil	3.30	0.37	9	1.78	1.95
7	Bortezomib	2.45	0.17	10	2.16	0.93
8	Bromfenac	3.09	0.0011	0.11	1.35	0.99
9	Bromopride	2.65	0.6	3.1	1.09	1.63
10	Buflomedil	3.08	0.4	1.3	0.86	1.85
11	Bunazosin	1.72	0.06	0.72	1.41	1.19
12	CB 10-277	2.40	0.087	0.25	0.61	0.70
13	Cefcanel	0.70	0.16	0.13	-0.63	0.19
14	Cefetamet	0.00	0.78	0.28	-0.49	0.10
15	Cephaloridine	-2.48	0.8	0.46	-0.06	0.05
16	Chlorazepate	2.14	0.035	0.2	0.83	0.60
17	m-Chlorophenylpiperazine	2.04	0.28	2.5	1.33	1.34
18	Conivaptan	4.43	0.01	0.76	2.22	2.51
19	Dacarbazine	0.06	1	1.2	0.40	0.50
20	4-Desacetylpaclitaxel	3.13	0.030	6.9	2.75	1.87
21	Dexniguldipine	6.90	0.0042	8.4	3.69	3.75
22	Diatrizoic Acid	0.45	1	0.26	-0.82	0.15
23	Dihydroquinidine	3.21	0.22	2.8	1.48	1.91
24	Diprafenone	4.12	0.017	1.2	2.21	2.36
25	Doxifluridine	-1.05	0.61	0.28	-0.31	0.13
26	DP-b 99	9.24	0.0003	1.1	3.98	4.02
27	Edrophonium	-3.01	0.96	1.1	0.37	0.03
28	Eniporide	0.61	0.26	1.1	0.98	0.70
29	Enprofylline	0.23	0.55	0.63	0.33	0.51
30	Epristeride	3.97	0.03	0.54	1.56	1.41
31	Ertapenem	0.53	0.1	0.12	-0.12	0.17
32	Ethacrynic Acid	3.52	0.048	0.26	0.91	1.19
33	Etilefrine	-0.02	0.77	2.1	0.80	0.47
34	Etoricoxib	3.21	0.081	1.5	1.63	1.91
35	Exatecan (acid)	0.27	0.24	0.44	0.51	0.55
36	Fleroxacin	-0.29	0.73	1.6	0.69	0.30
37	Folinic acid	-2.80	0.087	0.25	0.61	0.00
38	Genaconazole	0.88	0.77	0.62	0.16	0.81
39	Gentamicin	-7.45	1	0.33	-0.49	0.00
40	Ginkgolide A	-0.93	0.81	0.62	0.13	0.23
41	3'-Hydroxycotinine	-0.90	0.79	0.85	0.33	0.23
42	3-Hydroxydihydroquinidine	2.14	0.47	6.8	1.55	1.39
43	2-Hydroxyimipramine	4.20	0.36	6.6	1.65	2.40
44	Hydroxyurea	-1.66	1	0.52	-0.10	0.11

No.	Compound name	log $P_{o/w}$	$f_{u,pl}$	V_{SS}	Obs. p$f_{u,ti}$	Calc. p$f_{u,ti}$
45	Ibandronic Acid	-2.74	0.15	0.55	0.86	0.00
46	Idazoxan	1.31	0.40	3.3	1.29	1.00
47	Imipramine N-oxide	0.77	0.32	1.9	1.13	0.77
48	Iopamidol	-2.20	1	0.28	-0.70	0.06
49	Iothalamic acid	-0.37	0.98	0.17	0.00	0.07
50	Lithium Carbonate	-0.62	1	0.62	0.02	0.05
51	Lovastatin (acid)	3.90	0.043	0.87	1.65	1.37
52	Medroxalol	1.53	0.43	7.9	1.65	1.10
53	Melperone	3.86	0.22	14	2.19	2.23
54	Mercaptopurine, 6-	0.53	0.85	1	0.38	0.54
55	Methylnaltrexone	-1.12	0.86	2.6	0.85	0.19
56	7a,19-Methylnortestosterone	2.99	0.027	0.8	1.81	1.80
57	Metrizoate	1.17	1	0.17	0.00	0.29
58	Napsagatran	0.53	0.22	0.36	0.44	0.65
59	NK 611	0.86	0.013	0.3	1.58	0.80
60	Oxiracetam	-1.62	0.92	0.55	-0.01	0.12
61	Paricalcitol	5.52	0.0016	0.41	2.68	3.06
62	Phenethicillin	2.29	0.25	0.3	0.24	0.65
63	Pirmenol	3.84	0.13	1.4	1.39	2.22
64	Prednisolone	1.62	0.25	0.86	0.87	1.14
65	Quinaprilat	0.24	0.32	0.13	0.00	0.12
66	Rabeprazole	2.48	0.037	0.22	0.90	1.54
67	Repinotan	2.58	0.10	0.21	0.39	1.60
68	Ribostamycin	-7.55	0.92	0.25	-0.80	0.00
69	SarCNU	-0.09	0.65	0.76	0.36	0.45
70	Sisomicin	-6.13	0.15	0.19	0.10	0.00
71	Sitagliptin	1.45	0.62	2.8	1.03	1.06
72	Sulbactam	-0.93	0.62	0.32	-0.20	0.04
73	Sulbenicillin	-1.15	0.5	0.15	0.00	0.03
74	Telavancin	-2.37	0.07	0.115	-0.54	0.02
75	Telithromycin	3.00	0.41	3	1.24	1.81
76	Thiopental	1.27	0.14	1.2	1.29	0.80
77	Tomopenem	-1.64	0.91	0.23	-0.97	0.12
78	Topixantrone	1.12	0.22	57	2.81	0.92
79	Trabectedin	2.37	0.54	25	2.06	1.50
80	Triamterene	1.12	0.42	13	1.88	0.92
81	Trimazosin	1.37	0.01	0.18	1.30	1.03
82	Trospectomycin	-1.70	0.82	0.7	0.20	0.11
83	Tubocurarine	1.14	0.58	0.45	0.10	0.93
84	UK-240,455	0.57	0.13	0.8	1.12	0.49
85	Valspodar	4.15	0.022	1.8	2.29	2.38
86	Vardenafil	2.65	0.05	3	2.16	1.64
87	Verlukast	5.41	0.0004	0.11	1.79	2.11
88	Vorinostat	1.62	0.29	0.5	0.50	1.14
89	Zoledronic Acid	-4.95	0.78	0.6	0.13	0.00
90	Zotarolimus	3.55	0.033	1.3	1.96	2.08

Table A2 Observed *vs.* predicted log *PS* constants and main physicochemical descriptors for external validation set compounds.

No.	Compound name	log $P_{o/w}$	Acid pK_a	Base pK_a	Exp. log *PS*	Calc. log *PS*
1	Amitriptyline	4.92		8.8	-1.45	-1.11
2	Amoxapine	1.89		7.4	-1.53	-1.70
3	Atomoxetine	3.90		9.6	-1.67	-1.34
4	Buproprion	3.21		8.7	-1.77	-1.46
5	Carbamazepine	2.45	14.3	0.1	-1.97	-1.40
6	Citalopram	3.74		9.4	-2.19	-1.32
7	Clozapine	3.32		7.8	-1.60	-1.34
8	Donepezil	3.60		8.7	-1.76	-1.35
9	Doxepin	4.29		8.8	-1.61	-1.16
10	Ergotamine	2.50	9.4	6.7	-2.26	-2.16
11	Ethosuximide	0.38	9.3		-2.65	-2.06
12	Fluphenazine	4.36	14.9	7.9	-2.09	-1.18
13	Gabapentin	1.40	3.8	10.7	-2.74	-2.61
14	Haloperidol	3.36	13.7	8.2	-1.73	-1.38
15	Isocarboxazid	2.40		2	-1.56	-1.55
16	Lamotrigine	2.50		4.9	-2.85	-1.82
17	Loxapine	3.60		6.6	-1.72	-1.09
18	Maprotiline	5.10		10.8	-1.65	-1.21
19	Meprobamate	0.70	11.7		-3.27	-2.76
20	Mesoridazine	3.90		9.4	-2.03	-1.28
21	Metoclopramide	1.80	14.1	9.2	-2.85	-2.76
22	Mirtazapine	2.90		7.2	-1.67	-1.19
23	Olanzapine	3.00		7.8	-1.60	-1.43
24	Pemoline	0.50	10.4		-3.63	-2.27
25	Pergolide	4.00		8.2	-1.42	-1.20
26	Perphenazine	3.82	14.9	7.9	-1.58	-1.25
27	Phenelzine	1.00		7.7	-2.85	-2.40
28	Quetiapine	2.80	14.9	5.5	-1.63	-1.31
29	Risperidone	2.50		8.2	-2.03	-1.75
30	Rizatriptan	1.40		8.6	-3.93	-2.47
31	Selegiline	2.90		6.9	-1.67	-1.11
32	Sertraline	5.10		9.4	-1.31	-1.20
33	Sumatriptan	0.93	10.4	8.6	-4.78	-2.89
34	Tacrine	2.71		9.9	-2.22	-1.90
35	Temazepam	2.19	13.7	1.8	-1.65	-1.42
36	Thioridazine	5.99		9.4	-1.65	-1.34
37	Thiothixene	3.78		8.1	-1.51	-1.26
38	Tiagabine	2.60	3.7	9.3	-2.64	-1.70
39	Trazodone	2.90		6.7	-1.74	-1.26
40	Trifluoperazine	4.90		8.1	-2.10	-1.10
41	Venlafaxine	2.47	15.1	8.4	-2.19	-1.74
42	Zaleplon	1.23		0.5	-2.44	-1.77
43	Ziprasidone	3.80	14.8	7.1	-1.96	-1.17

Table A3 Experimental *vs.* calculated microsomal binding data and physicochemical property values for compounds comprising the data set from Austin et al.[224].

No.	Compound name	Acid pK_a	log $P_{o/w}$	$f_{u,inc}$	log $K_{b,inc}$ (exp.)	log $K_{b,inc}$ (Austin's model)	log $K_{b,inc}$ (current model)
1	Ethoxybenzamide		1.34	0.98	-1.69	-0.71	-0.88
2	Albendazole		3.29	0.56	-0.10	0.32	0.17
3	Alprazolam		1.84	0.82	-0.66	-0.44	-0.61
4	Amiodarone		6.35	0.002	2.70	1.95	1.82
5	Astemizole		5.14	0.012	1.92	1.30	1.17
6	Betaxolol		2.4	0.62	-0.21	-0.15	-0.31
7	Bumetanide	4.5	3.21	0.92	-1.06	-1.77	-0.95
8	Carbamazepine		1.54	0.87	-0.83	-0.60	-0.78
9	Cerivastatin	4.3	4.54	0.65	-0.27	-1.09	-0.23
10	Cinoxacin	4.7	0.59	0.92	-1.06	-3.13	-2.36
11	Clomipramine		5.25	0.038	1.40	1.36	1.23
12	Clozapine		3.6	0.26	0.45	0.49	0.34
13	Colchicine		0.82	0.94	-1.19	-0.99	-1.16
14	Dichloral-phenazone		-0.32	0.94	-1.19	-1.59	-1.78
15	Glipizide	5.9	1.64	0.96	-1.38	-1.97	-1.68
16	Glyburide	5.3	4.29	0.82	-0.66	-0.96	-0.33
17	Indapamide		1.76	0.96	-1.38	-0.49	-0.66
18	Indomethacin	4.5	3.79	0.81	-0.63	-1.47	-0.63
19	Isradipine		3.75	0.34	0.29	0.57	0.42
20	Ketoprofen	4.6	2.71	0.92	-1.06	-2.02	-1.22
21	Losartan	4.1	4.13	0.9	-0.95	-1.32	-0.45
22	Mebendazole		2.9	0.7	-0.37	0.12	-0.04
23	Methocarbamol		0.36	0.84	-0.72	-1.23	-1.41
24	Metyrapone		1.37	0.97	-1.51	-0.69	-0.87
25	Oxaprozin	4.2	4.81	0.87	-0.83	-0.96	-0.09
26	Phensuximide		0.68	0.75	-0.48	-1.06	-1.24
27	Pimozide		5.83	0.007	2.15	1.67	1.54
28	Piroxicam	6.3	0.82	0.92	-1.06	-2.08	-1.99
29	Promethazine		4.38	0.11	0.91	0.90	0.76
30	Propafenone		3.85	0.15	0.75	0.62	0.47
31	Sulfadoxine	5.75	0.93	0.97	-1.51	-2.46	-2.09
32	Sulindac	4.5	2.86	0.86	-0.79	-1.96	-1.14
33	Tamoxifen		6.34	0.003	2.52	1.94	1.82
34	Thioridazine		5.69	0.009	2.04	1.60	1.47
35	Tolmetin	3.5	2.88	0.94	-1.19	-2.01	-1.13
36	Trimeprazine		4.41	0.081	1.05	0.92	0.77
37	Trioxasalen		3.47	0.38	0.21	0.42	0.27
38	Verapamil		4.1	0.37	0.23	0.75	0.61
39	Propranolol		3.15	0.94	-1.19	-1.69	-0.97
40	Imipramine		5.18	0.11	0.91	1.33	1.19

No.	Compound name	Acid pK_a	$\log P_{o/w}$	$f_{u,inc}$	$\log K_{b,inc}$ (exp.)	$\log K_{b,inc}$ (Austin's model)	$\log K_{b,inc}$ (current model)
41	Warfarin	5	3.11	0.71	-0.39	0.23	0.07
42	Chlorpromazine		2.29	0.86	-0.79	-0.21	-0.37
43	Diphenhydramine		4.08	0.12	0.87	0.74	0.60
44	Diltiazem		4.96	0.081	1.05	1.21	1.07
45	Desipramine		3.35	0.7	-0.37	0.36	0.20
46	Amitriptyline		1.22	0.56	-0.10	-0.77	-0.95
47	Quinidine		2.25	0.66	-0.29	-0.23	-0.39
48	Prednisone		1.97	0.89	-0.91	-0.38	-0.54
49	Diazepam		2.22	0.78	-0.55	-0.24	-0.41
50	Methoxsalen		4.49	0.97	-1.51	-1.14	-0.26
51	Triazolam		3.98	0.91	-1.00	-1.38	-0.53
52	Diclofenac	4	3.13	0.97	-1.51	-1.59	-0.96
53	Ibuprofen	4.4	2.07	0.97	-1.51	-0.32	-0.49
54	Tolbutamide	5.27	1.49	0.96	-1.38	-0.63	-0.80
55	Amobarbital		3.15	0.94	-1.19	-1.69	-0.97
56	Hexobarbital		5.18	0.11	0.91	1.33	1.19

Table A4 Experimental brain exposure data (log K_p^{uu}) vs. calculated *in vivo* permeability across blood-brain barrier (log $PS \cdot f_{u,pl}$).

No.	Compound name	log $P_{o/w}$	$f_{u,pl}$	log K_p^{uu}	log PS	log $PS \cdot f_{u,pl}$
1	Alprenolol	2.61	0.44	-0.42	-2.2	-2.56
2	Amitriptyline	4.79	0.09	-0.14	-1.2	-2.25
3	Atenolol	0.09	1	-1.59	-4.0	-4.00
4	Baclofen	-0.20	1	-1.70	-3.5	-3.50
5	Bupropion	2.82	0.31	0.30	-1.7	-2.21
6	Codeine	1.09	0.95	-0.05	-2.5	-2.52
7	Delavirdine	2.55	0.016	-1.37	-2.3	-4.10
8	Diazepam	2.87	0.12	0.03	-1.2	-2.12
9	Diphenhydramine	3.87	0.48	0.02	-1.3	-1.62
10	Ethylphenylmalonamide	0.15	0.55	0.10	-2.9	-3.16
11	Gabapentin	-0.40	1	-0.85	-3.6	-3.60
12	Indomethacin	4.30	0.01	-0.96	-2.1	-4.10
13	Lamotrigine	2.12	0.51	-0.06	-2.1	-2.39
14	Levofloxacin	0.00	0.82	-0.92	-3.5	-3.59
15	Loperamide[a]	5.62	0.06	-2.15	-1.5	-2.72
16	M3G	-2.07	1	-1.96	-5.7	-5.70
17	M6G	-1.32	0.98	-2.10	-5.3	-5.31
18	Methotrexate	-1.18	1	-2.22	-6.7	-6.70
19	Metoprolol	2.05	0.9	-0.19	-2.5	-2.55
20	Morphine	0.87	0.9	-0.82	-2.8	-2.85
21	Moxalactam	-1.36	0.32	-1.72	-6.0	-6.49
22	Nadolol	0.91	0.86	-1.43	-3.6	-3.67
23	Nelfinavir	5.16	0.0027	-1.72	-1.6	-4.17
24	Nitrofurantoin[a]	0.22	0.48	-1.96	-3.2	-3.52
25	Norfloxacin	-0.92	0.87	-1.55	-4.0	-4.06
26	Oxprenolol	2.01	0.45	-0.70	-2.5	-2.85
27	Oxycodone	0.50	0.87	0.01	-2.9	-2.96
28	Oxymorphone	0.94	0.73	-0.10	-2.8	-2.94
29	Paclitaxel[a]	3.28	0.05	-2.15	-2.7	-4.00
30	Pindolol	1.96	0.43	-0.30	-2.8	-3.17
31	Propranolol	2.93	0.09	-0.21	-2.0	-3.05
32	Rifampicin	4.69	0.12	-1.46	-4.0	-4.92
33	Salicylic acid	2.43	0.28	-0.72	-2.7	-3.25
34	Saquinavir	4.05	0.007	-1.26	-2.4	-4.55
35	Tacrine	2.69	0.55	-0.11	-2.0	-2.26
36	Thiopental	2.89	0.19	0.18	-1.6	-2.32
37	Thioridazine	6.06	0.002	-0.35	-1.5	-4.20
38	Topiramate	0.37	0.79	-0.48	-3.0	-3.10
39	Tramadol	2.65	0.85	0.16	-2.0	-2.07
40	Verapamil[a]	4.22	0.12	-1.28	-1.5	-2.42
41	Zidovudine	-0.07	0.64	-1.05	-3.2	-3.39

[a] Compound omitted from analysis due to carrier-mediated efflux.

REFERENCES

1. Gleeson MP, Hersey A, Hannongbua S. In-silico ADME models: a general assessment of their utility in drug discovery applications. *Curr Top Med Chem*. 2011;11(4):358-381.

2. Kerns E, Di L. *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*. 1st ed. San Diego: Academic Press; 2008.

3. Reichel A. Addressing central nervous system (CNS) penetration in drug discovery: basics and implications of the evolving new concept. *Chem Biodivers*. 2009;6(11):2030-2049.

4. Sazonovas A, Japertas P, Didziapetris R. Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). *SAR QSAR Environ Res*. 2010;21(1):127-148.

5. Didziapetris R, Dapkunas J, Sazonovas A, Japertas P. Trainable structure-activity relationship model for virtual screening of CYP3A4 inhibition. *J. Comput Aided Mol Des*. 2010;24(11):891-906.

6. Guo Q, Brady M, Gunn RN. A biomathematical modeling approach to central nervous system radioligand discovery and development. *J Nucl Med*. 2009;50(10):1715-1723.

7. van de Waterbeemd H. Which in vitro screens guide the prediction of oral absorption and volume of distribution? *Basic Clin Pharmacol Toxicol*. 2005;96(3):162-166.

8. Smith DA, Di L, Kerns EH. The effect of plasma protein binding on in vivo efficacy: misconceptions in drug discovery. *Nat Rev Drug Discov*. 2010;9(12):929-939.

9. Vuignier K, Schappler J, Veuthey J-L, Carrupt P-A, Martel S. Drug-protein binding: a critical review of analytical tools. *Anal Bioanal Chem*. 2010;398(1):53-66.

10. Hage DS, Jackson A, Sobansky MR, et al. Characterization of drug-protein interactions in blood using high-performance affinity chromatography. *J Sep Sci*. 2009;32(5-6):835-853.

11. Schmidt S, Gonzalez D, Derendorf H. Significance of protein binding in pharmacokinetics and pharmacodynamics. *J Pharm Sci*. 2010;99(3):1107-1122.

12. Yamazaki K, Kanaoka M. Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. *J Pharm Sci*. 2004;93(6):1480-1494.

13. Fournier T, Medjoubi-N N, Porquet D. Alpha-1-acid glycoprotein. *Biochim Biophys Acta*. 2000;1482(1-2):157-171.

14. Avvakumov GV, Cherkasov A, Muller YA, Hammond GL. Structural analyses of sex hormone-binding globulin reveal novel ligands and function. *Mol Cell Endocrinol*. 2010;316(1):13-23.

15. Lin H-Y, Muller YA, Hammond GL. Molecular and structural basis of steroid hormone binding and release from corticosteroid-binding globulin. *Mol Cell Endocrinol*. 2010;316(1):3-12.

16. Kratochwil NA, Huber W, Müller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs: a new approach. *Biochem Pharmacol*. 2002;64(9):1355-1374.

17. Kragh-Hansen U, Chuang VTG, Otagiri M. Practical aspects of the ligand-binding and enzymatic properties of human serum albumin. *Biol Pharm Bull*. 2002;25(6):695-704.

18. Sudlow G, Birkett DJ, Wade DN. The characterization of two specific drug binding sites on human serum albumin. *Mol Pharmacol*. 1975;11(6):824-832.

19. Curry S. Lessons from the crystallographic analysis of small molecule binding to human serum albumin. *Drug Metab Pharmacokinet*. 2009;24(4):342-357.

20. Kosa T, Maruyama T, Otagiri M. Species differences of serum albumins: I. Drug binding sites. *Pharm Res*. 1997;14(11):1607-1612.

21. Hervé F, Caron G, Duché JC, et al. Ligand specificity of the genetic variants of human alpha1-acid glycoprotein: generation of a three-dimensional quantitative structure-activity relationship model for drug binding to the A variant. *Mol Pharmacol*. 1998;54(1):129-138.

22. Nishi K, Ono T, Nakamura T, et al. Structural insights into differences in drug-binding selectivity between two forms of human alpha1-acid glycoprotein genetic variants, the A and F1*S forms. *J Biol Chem*. 2011;286(16):14427-14434.

23. Banker MJ, Clark TH. Plasma/serum protein binding determinations. *Curr Drug Metab*. 2008;9(9):854-859.

24. Howard ML, Hill JJ, Galluppi GR, McLean MA. Plasma protein binding in drug discovery and development. *Comb Chem High Throughput Screen*. 2010;13(2):170-187.

25. Kratochwil NA, Huber W, Müller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs: a new approach. *Biochem Pharmacol*. 2002;64(9):1355-1374.

26. Valko K, Nunhuck S, Bevan C, Abraham MH, Reynolds DP. 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(11):2236-2248.

27. Schmitt W. General approach for the calculation of tissue to plasma partition coefficients. *Toxicol In Vitro*. 2008;22(2):457-467.

28. Berry LM, Roberts J, Be X, Zhao Z, Lin M-HJ. Prediction of V(ss) from in vitro tissue-binding studies. *Drug Metab Dispos*. 2010;38(1):115-121.

29. Toutain PL, Bousquet-Mélou A. Volumes of distribution. *J Vet Pharmacol Ther*. 2004;27(6):441-453.

30. Sui X, Sun J, Wu X, et al. Predicting the volume of distribution of drugs in humans. *Curr Drug Metab*. 2008;9(6):574-580.

31. Yates JWT, Arundel PA. On the volume of distribution at steady state and its relationship with two-compartmental models. *J Pharm Sci*. 2008;97(1):111-122.

32. Smith DA, van de Waterbeemd H, Walker DK, et al. eds. *Pharmacokinetics and Metabolism in Drug Design*.

33. Toutain PL, Bousquet-Mélou A. Plasma terminal half-life. *J Vet Pharmacol Ther*. 2004;27(6):427-439.

34. Oberg T. A QSAR for baseline toxicity: validation, domain of application, and prediction. *Chem Res Toxicol*. 2004;17(12):1630-1637.

35. Anderson N, Borlak J. Drug-induced phospholipidosis. *FEBS Lett*. 2006;580(23):5533-5540.

36. Pelletier DJ, Gehlhaar D, Tilloy-Ellul A, Johnson TO, Greene N. Evaluation of a published in silico model and construction of a novel Bayesian model for predicting phospholipidosis inducing potential. *J Chem Inf Model*. 2007;47(3):1196-1205.

37. Atienzar F, Gerets H, Dufrane S, et al. Determination of phospholipidosis potential based on gene expression analysis in HepG2 cells. *Toxicol Sci*. 2007;96(1):101-114.

38. Pospischil A, Walther P, Dingemanse J. Phospholipidosis in healthy subjects participating in clinical studies: ultrastructural findings in white blood cells. *Exp Toxicol Pathol*. 2010;62(5):567-571.

39. Hanumegowda UM, Wenke G, Regueiro-Ren A, et al. Phospholipidosis as a function of basicity, lipophilicity, and volume of distribution of compounds. *Chem Res Toxicol*. 2010;23(4):749-755.

40. Abbott NJ. 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.

41. Goodwin JT, Clark DE. In Silico Predictions of Blood-Brain Barrier Penetration: Considerations to "Keep in Mind". *J Pharmacol Exp Ther*. 2005;315:477-483.

42. Löscher W, Potschka H. Drug resistance in brain diseases and the role of drug efflux transporters. *Nat Rev Neurosci*. 2005;6(8):591-602.

43. Pangalos MN, Schechter LE, Hurko O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat Rev Drug Discov*. 2007;6(7):521-532.

44. Di L, Kerns EH, Bezar IF, Petusky SL, Huang Y. Comparison of blood-brain barrier permeability assays: in situ brain perfusion, MDR1-MDCKII and PAMPA-BBB. *J Pharm Sci*. 2009;98(6):1980-1991.

45. Cardoso FL, Brites D, Brito MA. Looking at the blood-brain barrier: molecular anatomy and possible investigation approaches. *Brain Res Rev*. 2010;64(2):328-363.

46. Abbott NJ, Patabendige AAK, Dolman DEM, Yusof SR, Begley DJ. Structure and function of the blood-brain barrier. *Neurobiol Dis*. 2010;37(1):13-25.

47. Abbott NJ, Rönnbäck L, Hansson E. Astrocyte-endothelial interactions at the blood-brain barrier. *Nat Rev Neurosci*. 2006;7(1):41-53.

48. Hawkins BT, Davis TP. The blood-brain barrier/neurovascular unit in health and disease. *Pharmacol Rev*. 2005;57(2):173-185.

49. Crone C, Olesen SP. Electrical resistance of brain microvascular endothelium. *Brain Res*. 1982;241(1):49-55.

50. Butt AM, Jones HC, Abbott NJ. Electrical resistance across the blood-brain barrier in anaesthetized rats: a developmental study. *J Physiol (Lond)*. 1990;429:47-62.

51. Butt AM. Effect of inflammatory agents on electrical resistance across the blood-brain barrier in pial microvessels of anaesthetized rats. *Brain Res*. 1995;696(1-2):145-150.

52. McCaffrey G, Staatz WD, Quigley CA, et al. Tight junctions contain oligomeric protein assembly critical for maintaining blood-brain barrier integrity in vivo. *J Neurochem*. 2007;103(6):2540-2555.

53. Abraham MH. The factors that influence permeation across the blood-brain barrier. *Eur J Med Chem*. 2004;39:235-240.

54. Bodor, Buchwald. Recent advances in the brain targeting of neuropharmaceuticals by chemical delivery systems. *Adv Drug Deliv Rev*. 1999;36(2-3):229-254.

55. Bickel U. How to measure drug transport across the blood-brain barrier. *NeuroRx*. 2005;2:15-26.

56. Mensch J, Oyarzabal J, Mackie C, Augustijns P. In vivo, in vitro and in silico methods for small molecule transfer across the BBB. *J Pharm Sci*. 2009;98(12):4429-4468.

57. Van Bree JB, De Boer AG, Danhof M, Breimer DD. Drug transport across the blood-brain barrier. II. Experimental techniques to study drug transport. *Pharm Weekbl Sci*. 1992;14(6):338-348.

58. Cecchelli R, Berezowski V, Lundquist S, et al. Modelling of the blood-brain barrier in drug discovery and development. *Nat Rev Drug Discov*. 2007;6(8):650-661.

59. Ohno K, Pettigrew KD, Rapoport SI. Lower limits of cerebrovascular permeability to nonelectrolytes in the conscious rat. *Am J Physiol*. 1978;235(3):H299-307.

60. Oldendorf WH. Brain uptake of radiolabeled amino acids, amines and hexoses after arterial injection. *Am J Physiol*. 1971;221:1629-1639.

61. Takasato Y, Rapoport SI, Smith QR. An in situ brain perfusion technique to study cerebrovascular transport in the rat. *Am J Physiol*. 1984;247(3 Pt 2):H484-493.

62. Smith QR, Allen DD. In situ brain perfusion technique. *Methods Mol Med*. 2003;89:209-218.

63. Dagenais C, Rousselle C, Pollack GM, Scherrmann JM. Development of an in situ mouse brain perfusion model and its application to mdr1a P-glycoprotein-deficient mice. *J Cereb Blood Flow Metab*. 2000;20(2):381-386.

64. Young RC, Mitchell RC, Brown TH, et al. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists. *J Med Chem*. 1988;31(3):656-671.

65. Tong W-Q. Molecular and Physicochemical Properties Impacting Oral Absorption of Drugs. In: Krishna R, Yu L, eds. *Biopharmaceutics Applications in Drug Development*. Boston, MA: Springer US; 2008:26-46.

66. Cao X, Yu L, Sun D. Drug Absorption Principles. In: Krishna R, Yu L, eds. *Biopharmaceutics Applications in Drug Development*. Boston, MA: Springer US; 2008:75-100.

67. Avdeef A. Physicochemical profiling (solubility, permeability and charge state). *Curr Top Med Chem*. 2001;1(4):277-351.

68. Avdeef A. *Absoption and Drug Development*. 1st ed. New York: Wiley-Interscience; 2003.

69. Sugano K. Computational oral absorption simulation for low-solubility compounds. *Chem Biodivers*. 2009;6(11):2014-2029.

70. Wiedmann TS, Kamel L. Examination of the solubilization of drugs by bile salt micelles. *J Pharm Sci*. 2002;91(8):1743-1764.

71. Sietsema WK. The absolute oral bioavailability of selected drugs. *Int J Clin Pharmacol Ther Toxicol*. 1989;27(4):179-211.

72. Varma MVS, Obach RS, Rotter C, et al. Physicochemical space for optimum oral bioavailability: contribution of human intestinal absorption and first-pass elimination. *J Med Chem*. 2010;53(3):1098-1108.

73. Turner JR. Intestinal mucosal barrier function in health and disease. *Nat Rev Immunol*. 2009;9(11):799-809.

74. Sugano K, Kansy M, Artursson P, et al. Coexistence of passive and carrier-mediated processes in drug transport. *Nat Rev Drug Discov*. 2010;9(8):597-614.

75. Shimizu M. Interaction between food substances and the intestinal epithelium. *Biosci Biotechnol Biochem*. 2010;74(2):232-241.

76. Anderson JM, Van Itallie CM. Physiology and function of the tight junction. *Cold Spring Harb Perspect Biol*. 2009;1(2):a002584.

77. Zhao YH, Le J, Abraham MH, et al. Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J Pharm Sci*. 2001;90(6):749-784.

78. Toutain PL, Bousquet-Mélou A. Bioavailability and its assessment. *J Vet Pharmacol Ther*. 2004;27(6):455-466.

79. Lennernäs H, Nylander S, Ungell AL. Jejunal permeability: a comparison between the ussing chamber technique and the single-pass perfusion in humans. *Pharm Res*. 1997;14(5):667-671.

80. Lennernäs H. Human intestinal permeability. *J Pharm Sci*. 1998;87(4):403-410.

81. Lennernäs H. Intestinal permeability and its relevance for absorption and elimination. *Xenobiotica*. 2007;37(10-11):1015-1051.

82. Linnankoski J, Mäkelä J, Palmgren J, et al. Paracellular porosity and pore size of the human intestinal epithelium in tissue and cell culture models. *J Pharm Sci*. 2010;99(4):2166-2175.

83. Artursson P, Palm K, Luthman K. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv Drug Deliv Rev*. 2001;46(1-3):27-43.

84. Hubatsch I, Ragnarsson EGE, Artursson P. Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers. *Nat Protoc*. 2007;2(9):2111-2119.

85. Pham The H, González-Álvarez I, Bermejo M, et al. In Silico Prediction of Caco-2 Cell Permeability by a Classification QSAR Approach. *Mol Inf*. 2011;30(4):376-385.

86. Volpe DA. Variability in Caco-2 and MDCK cell-based intestinal permeability assays. *J Pharm Sci*. 2008;97(2):712-725.

87. Korjamo T, Heikkinen AT, Mönkkönen J. Analysis of unstirred water layer in in vitro permeability experiments. *J Pharm Sci*. 2009;98(12):4469-4479.

88. Korjamo T, Heikkinen AT, Waltari P, Mönkkönen J. The asymmetry of the unstirred water layer in permeability experiments. *Pharm Res*. 2008;25(7):1714-1722.

89. Adson A, Raub TJ, Burton PS, et al. Quantitative approaches to delineate paracellular diffusion in cultured epithelial cell monolayers. *J Pharm Sci*. 1994;83(11):1529-1536.

90. Avdeef A, Tam KY. How well can the Caco-2/Madin-Darby canine kidney models predict effective human jejunal permeability? *J Med Chem*. 2010;53(9):3566-3584.

91. Avdeef A. Leakiness and size exclusion of paracellular channels in cultured epithelial cell monolayers-interlaboratory comparison. *Pharm Res*. 2010;27(3):480-489.

92. Pardridge WM. Drug targeting to the brain. *Pharm Res*. 2007;24(9):1733-1744.

93. Habgood MD, Begley DJ, Abbott NJ. Determinants of passive drug entry into the central nervous system. *Cell Mol Neurobiol*. 2000;20(2):231-253.

94. Thomae AV, Wunderli-Allenspach H, Krämer SD. Permeation of aromatic carboxylic acids across lipid bilayers: the pH-partition hypothesis revisited. *Biophys J*. 2005;89(3):1802-1811.

95. Thomae AV, Koch T, Panse C, Wunderli-Allenspach H, Krämer SD. Comparing the lipid membrane affinity and permeation of drug-like acids: the intriguing effects of cholesterol and charged lipids. *Pharm Res*. 2007;24(8):1457-1472.

96. Krämer SD, Lombardi D, Primorac A, Thomae AV, Wunderli-Allenspach H. Lipid-bilayer permeation of drug-like compounds. *Chem Biodivers*. 2009;6(11):1900-1916.

97. Scherrer RA. Multi-pH QSAR: II. Regression Analysis Sensitive Enough to Determine the Transition-State pKa of Human Buccal Absorption. *Mol Inf*. 2011;30(2-3):251-255.

98. Fagerholm U. The highly permeable blood-brain barrier: an evaluation of current opinions about brain uptake capacity. *Drug Discov Today*. 2007;12(23-24):1076-1082.

99. Ohtsuki S, Terasaki T. Contribution of carrier-mediated transport systems to the blood-brain barrier as a supporting and protecting interface for the brain; importance for CNS drug discovery and development. *Pharm Res*. 2007;24(9):1745-1758.

100. Klaassen CD, Aleksunes LM. Xenobiotic, bile acid, and cholesterol transporters: function and regulation. *Pharmacol Rev*. 2010;62(1):1-96.

101. Demel MA, Krämer O, Ettmayer P, Haaksma EEJ, Ecker GF. Predicting ligand interactions with ABC transporters in ADME. *Chem Biodivers*. 2009;6(11):1960-1969.

102. Scherrmann J-M. Transporters in absorption, distribution, and elimination. *Chem Biodivers*. 2009;6(11):1933-1942.

103. Zhou S-F. Structure, function and regulation of P-glycoprotein and its clinical relevance in drug disposition. *Xenobiotica*. 2008;38(7-8):802-832.

104. Didziapetris R, Japertas P, Avdeef A, Petrauskas A. Classification analysis of P-glycoprotein substrate specificity. *J Drug Target*. 2003;11(7):391-406.

105. Ni Z, Bikadi Z, Rosenberg MF, Mao Q. Structure and function of the human breast cancer resistance protein (BCRP/ABCG2). *Curr Drug Metab*. 2010;11(7):603-617.

106. Aller SG, Yu J, Ward A, et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*. 2009;323(5922):1718-1722.

107. Tsuji A. Impact of transporter-mediated drug absorption, distribution, elimination and drug interactions in antimicrobial chemotherapy. *J Infect Chemother*. 2006;12(5):241-250.

108. Ballatori N, Li N, Fang F, et al. OST alpha-OST beta: a key membrane transporter of bile acids and conjugated steroids. *Front Biosci*. 2009;14:2829-2844.

109. Tsuji A. Small molecular drug transfer across the blood-brain barrier via carrier-mediated transport systems. *NeuroRx*. 2005;2(1):54-62.

110. Hawkins R, O'Kane R, Simpson I, Viña J. Structure of the blood-brain barrier and its role in the transport of amino acids. *J Nutr*. 2006;136(1 Suppl):218S-26S.

111. Koepsell H, Lips K, Volk C. Polyspecific organic cation transporters: structure, function, physiological roles, and biopharmaceutical implications. *Pharm Res*. 2007;24(7):1227-1251.

112. Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Prediction of Blood-Brain Barrier Penetration by Drugs. In: Jain KK, ed. *Drug Delivery to the Central Nervous System*.Vol 45. 1st ed. Neuromethods. New York: Springer; 2010:63-83.

113. Cisternino S, Rousselle C, Dagenais C, Scherrmann JM. Screening of multidrug-resistance sensitive drugs by in situ brain perfusion in P-glycoprotein-deficient mice. *Pharm Res*. 2001;18(2):183-190.

114. Vlaming MLH, Lagas JS, Schinkel AH. Physiological and pharmacological roles of ABCG2 (BCRP): recent findings in Abcg2 knockout mice. *Adv Drug Deliv Rev*. 2009;61(1):14-25.

115. Kruh GD, Belinsky MG, Gallo JM, Lee K. Physiological and pharmacological functions of Mrp2, Mrp3 and Mrp4 as determined from recent studies on gene-disrupted mice. *Cancer Metastasis Rev*. 2007;26(1):5-14.

116. Estrada E, Uriarte E, Molina E, Simón-Manso Y, Milne GWA. An integrated in silico analysis of drug-binding to human serum albumin. *J Chem Inf Model*. 2006;46(6):2709-2724.

117. Wichmann K, Diedenhofen M, Klamt A. Prediction of blood-brain partitioning and human serum albumin binding based on COSMO-RS sigma-moments. *J Chem Inf Model*. 2007;47:228-233.

118. Deeb O, Hemmateenejad B. ANN-QSAR model of drug-binding to human serum albumin. *Chem Biol Drug Des*. 2007;70(1):19-29.

119. Colmenarejo G. In silico prediction of drug-binding strengths to human serum albumin. *Med Res Rev*. 2003;23(3):275-301.

120. Rodgers SL, Davis AM, Tomkinson NP, van de Waterbeemd H. QSAR modeling using automatically updating correction libraries: application to a human plasma protein binding model. *J Chem Inf Model*. 2007;47(6):2401-2407.

121. Weaver S, Gleeson MP. The importance of the domain of applicability in QSAR modeling. *J Mol Graph Model*. 2008;26(8):1315-1326.

122. Lombardo F, Obach RS, Shalaeva MY, Gao F. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. *J Med Chem*. 2002;45(13):2867-2876.

123. Lombardo F, Obach RS, Shalaeva MY, Gao F. Prediction of human volume of distribution values for neutral and basic drugs. 2. Extended data set and leave-class-out statistics. *J Med Chem*. 2004;47(5):1242-1250.

124. Hollósy F, Valkó K, Hersey A, et al. 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(24):6958-6971.

125. Rodgers T, Rowland M. Mechanistic approaches to volume of distribution predictions: understanding the processes. *Pharm Res*. 2007;24(5):918-933.

126. Sui X, Sun J, Li H, et al. Prediction of volume of distribution values in human using immobilized artificial membrane partitioning coefficients, the fraction of compound ionized and plasma protein binding data. *Eur J Med Chem*. 2009;44(11):4455-4460.

127. Berellini G, Springer C, Waters NJ, Lombardo F. In silico prediction of volume of distribution in human using linear and nonlinear models on a 669 compound data set. *J Med Chem*. 2009;52(14):4488-4495.

128. Poulin P, Theil F-P. Development of a novel method for predicting human volume of distribution at steady-state of basic drugs and comparative assessment with existing methods. *J Pharm Sci*. 2009;98(12):4941-4961.

129. Poulin P, Ekins S, Theil F-P. A hybrid approach to advancing quantitative prediction of tissue distribution of basic drugs in human. *Toxicol Appl Pharmacol*. 2011;250(2):194-212.

130. Poulin P, Theil FP. A priori prediction of tissue:plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery. *J Pharm Sci*. 2000;89(1):16-35.

131. Levin VA. Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability. *J Med Chem*. 1980;23(6):682-684.

132. Hansch C, Björkroth JP, Leo A. Hydrophobicity and central nervous system agents: on the principle of minimal hydrophobicity in drug design. *J Pharm Sci*. 1987;76:663-687.

133. Pardridge WM, Triguero D, Yang J, Cancilla PA. Comparison of in vitro and in vivo models of drug transcytosis through the blood-brain barrier. *J Pharm Exp Ther*. 1990;253:884-891.

134. Murakami H, Takanaga H, Matsuo H, Ohtani H, Sawada Y. Comparison of blood-brain barrier permeability in mice and rats using in situ brain perfusion technique. *Am J Physiol Heart Circ Physiol*. 2000;279(3):H1022-1028.

135. Luco JM, Marchevsky E. QSAR Studies on Blood-Brain Barrier Permeation. *Curr Comp-Aided Drug Des*. 2006;2:31-35.

136. Gratton JA, Abraham MH, Bradbury MW, Chadha HS. Molecular factors influencing drug transfer across the blood-brain barrier. *J Pharm Pharmacol*. 1997;49(12):1211-1216.

137. Abraham MH. The permeation of neutral molecules, ions, and ionic species through membranes: brain permeation as an example. *J Pharm Sci*. 2011;100(5):1690-1701.

138. Liu X, Tu M, Kelly RS, Chen C, Smith BJ. Development of a computational approach to predict blood-brain barrier permeability. *Drug Metab Dispos*. 2004;32(1):132-139.

139. Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Ionization-specific prediction of blood-brain permeability. *J Pharm Sci*. 2009;98(1):122-134.

140. Dagenais C, Avdeef A, Tsinman O, Dudley A, Beliveau R. P-glycoprotein deficient mouse in situ blood-brain barrier permeability and its prediction using an in combo PAMPA model. *Eur J Pharm Sci*. 2009;38(2):121-137.

141. Garg P, Verma J. In silico prediction of blood brain barrier permeability: an Artificial Neural Network model. *J Chem Inf Model*. 2006;46(1):289-297.

142. Abraham MH, Ibrahim A, Zhao Y, Acree WE. A data base for partition of volatile organic compounds and drugs from blood/plasma/serum to brain, and an LFER analysis of the data. *J Pharm Sci*. 2006;95(10):2091-2100.

143. Abraham MH, Acree WE, Leo AJ, Hoekman D, Cavanaugh JE. Water-solvent partition coefficients and Delta Log P values as predictors for blood-brain distribution; application of the Akaike information criterion. *J Pharm Sci*. 2010;99(5):2492-2501.

144. Fu X-C, Wang G-P, Shan H-L, Liang W-Q, Gao J-Q. Predicting blood-brain barrier penetration from molecular weight and number of polar atoms. *Eur J Pharm Biopharm*. 2008;70(2):462-466.

145. Zhang L, Zhu H, Oprea TI, Golbraikh A, Tropsha A. QSAR modeling of the blood-brain barrier permeability for diverse organic compounds. *Pharm Res*. 2008;25(8):1902-1914.

146. Kortagere S, Chekmarev D, Welsh WJ, Ekins S. New predictive models for blood-brain barrier permeability of drug-like molecules. *Pharm Res*. 2008;25(8):1836-1845.

147. Chen Y, Zhu Q-J, Pan J, Yang Y, Wu X-P. A prediction model for blood-brain barrier permeation and analysis on its parameter biologically. *Comput Methods Programs Biomed*. 2009;95(3):280-287.

148. Fan Y, Unwalla R, Denny RA, et al. Insights for predicting blood-brain barrier penetration of CNS targeted molecules using QSPR approaches. *J Chem Inf Model*. 2010;50(6):1123-1133.

149. Norinder U, Haeberlein M. Computational approaches to the prediction of the blood-brain distribution. *Adv Drug Deliv Rev*. 2002;54(3):291-313.

150. Lobell M, Molnár L, Keserü GM. Recent advances in the prediction of blood-brain partitioning from molecular structure. *J Pharm Sci*. 2003;92(2):360-370.

151. Adenot M, Lahana R. Blood-brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J Chem Inf Comput Sci*. 2004;44(1):239-248.

152. Zhao YH, Abraham MH, Ibrahim A, et al. Predicting penetration across the blood-brain barrier from simple descriptors and fragmentation schemes. *J Chem Inf Model*. 2007;47(1):170-175.

153. Vilar S, Chakrabarti M, Costanzi S. Prediction of passive blood-brain partitioning: straightforward and effective classification models based on in silico derived physicochemical descriptors. *J Mol Graph Model*. 2010;28(8):899-903.

154. Fagerholm U. Prediction of human pharmacokinetics--gastrointestinal absorption. *J Pharm Pharmacol*. 2007;59(7):905-916.

155. Linnankoski J, Ranta V-P, Yliperttula M, Urtti A. Passive oral drug absorption can be predicted more reliably by experimental than computational models--fact or myth. *Eur J Pharm Sci*. 2008;34(2-3):129-139.

156. Willmann S, Lippert J, Schmitt W. From physicochemistry to absorption and distribution: predictive mechanistic modelling and computational tools. *Expert Opin Drug Metab Toxicol*. 2005;1(1):159-168.

157. Saitoh R, Sugano K, Takata N, et al. Correction of permeability with pore radius of tight junctions in Caco-2 monolayers improves the prediction of the dose fraction of hydrophilic drugs absorbed by humans. *Pharm Res*. 2004;21(5):749-755.

158. Obata K, Sugano K, Saitoh R, et al. Prediction of oral drug absorption in humans by theoretical passive absorption model. *Int J Pharm*. 2005;293(1-2):183-192.

159. Iyer M, Tseng YJ, Senese CL, Liu J, Hopfinger AJ. Prediction and mechanistic interpretation of human oral drug absorption using MI-QSAR analysis. *Mol Pharm*. 2007;4(2):218-231.

160. Deconinck E, Ates H, Callebaut N, Van Gyseghem E, Vander Heyden Y. Evaluation of chromatographic descriptors for the prediction of gastro-intestinal absorption of drugs. *J Chromatogr A*. 2007;1138(1-2):190-202.

161. Deconinck E, Coomans D, Vander Heyden Y. Exploration of linear modelling techniques and their combination with multivariate adaptive regression splines to predict gastro-intestinal absorption of drugs. *J Pharm Biomed Anal*. 2007;43(1):119-130.

162. Hou T, Wang J, Zhang W, Xu X. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. *J Chem Inf Model*. 2007;47(1):208-218.

163. Talevi A, Goodarzi M, Ortiz EV, et al. Prediction of drug intestinal absorption by new linear and non-linear QSPR. *Eur J Med Chem*. 2011;46(1):218-228.

164. Nakao K, Fujikawa M, Shimizu R, Akamatsu M. QSAR application for the prediction of compound permeability with in silico descriptors in practical use. *J Comput Aided Mol Des*. 2009;23(5):309-319.

165. Di Fenza A, Alagona G, Ghio C, et al. Caco-2 cell permeability modelling: a neural network coupled genetic algorithm approach. *J Comput Aided Mol Des*. 2007;21(4):207-221.

166. Castillo-Garit JA, Marrero-Ponce Y, Torrens F, Rotondo R. Atom-based stochastic and non-stochastic 3D-chiral bilinear indices and their applications to central chirality codification. *J Mol Graph Model*. 2007;26(1):32-47.

167. Gozalbes R, Jacewicz M, Annand R, Tsaioun K, Pineda-Lucena A. QSAR-based permeability model for drug-like compounds. *Bioorg Med Chem*. 2011;19(8):2615-2624.

168. Dollery C ed. *Therapeutic drugs*. 2nd ed. London: Churchill Livingstone; 1999.

169. Hardman, J.G., Goodman Gillman, A., Limbird, L.E. eds. *Goodman & Gillman's The pharmacological basis of therapeutics*. 9th ed. New York: McGraw Hill; 1996.

170. Moffat A, Jackson, JV, Moss, MS, Widdop, B eds. *Clarke's isolation and identification of drugs in pharmaceuticals, body fluids, and postmortemmaterial*. 2nd ed. London: The Pharmaceutical Press; 1986.

171. *Physicians' desk reference*. 48 and 51 editions. Medical Economics Data Production Company; 1994-1997.

172. *USP DI – Volume I. Drug Information for the Health Care Professional*. 23rd ed. Greenwood Village: Thomson MICROMEDEX; 2003.

173. Hall LM, Hall LH, Kier LB. QSAR modeling of beta-lactam binding to human serum proteins. *J Comput Aided Mol Des*. 2003;17(2-4):103-118.

174. Obach RS, Lombardo F, Waters NJ. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. *Drug Metab Dispos*. 2008;36(7):1385-1405.

175. Summerfield SG, Read K, Begley DJ, et al. Central nervous system drug disposition: the relationship between in situ brain permeability and brain free fraction. *J Pharmacol Exp Ther*. 2007;322:205-213.

176. Kalvass JC, Maurer TS. Influence of nonspecific brain and plasma binding on CNS exposure: implications for rational drug discovery. *Biopharm Drug Dispos*. 2002;23(8):327-338.

177. Summerfield SG, Lucas AJ, Porter RA, et al. Toward an improved prediction of human in vivo brain penetration. *Xenobiotica*. 2008;38(12):1518-1535.

178. 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(6):2147-2160.

179. Adenot M, Lahana R. Blood-brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J Chem Inf Comput Sci*. 2004;44(1):239-248.

180. Zmuidinavicius D, Didziapetris R, Japertas P, Avdeef A, Petrauskas A. Classification structure-activity relations (C-SAR) in prediction of human intestinal absorption. *J Pharm Sci*. 2003;92(3):621-633.

181. Zhao YH, Abraham MH, Le J, et al. Rate-limited steps of human oral absorption and QSAR studies. *Pharm Res*. 2002;19(10):1446-1457.

182. Zhao YH, Abraham MH, Hersey A, Luscombe CN. Quantitative relationship between rat intestinal absorption and Abraham descriptors. *Eur J Med Chem*. 2003;38(11-12):939-947.

183. Linnankoski J, Mäkelä JM, Ranta V-P, Urtti A, Yliperttula M. Computational prediction of oral drug absorption based on absorption rate constants in humans. *J Med Chem*. 2006;49(12):3674-3681.

184. Yu LX, Amidon GL. A compartmental absorption and transit model for estimating oral drug absorption. *Int J Pharm*. 1999;186(2):119-125.

185. Reynolds DP, Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Ionization-specific analysis of human intestinal absorption. *J Pharm Sci*. 2009;98(11):4039-4054.

186. Platts JA, Butina D, Abraham MH, Hersey A. Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach. *J Chem Inf Comput Sci*. 1999;39(5):835-845.

187. Ertl P, Rohde B, Selzer P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contribution and Its Application to the Prediction of Drug Transport Properties. *J Med Chem*. 2000;43:3714-3717.

188. Abraham MH. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem Soc Rev*. 1993;22(2):73.

189. Abraham MH, McGowan JC. The Use of Characteristic Volumes to Measure Cavity Terms in Reversed Phase Liquid-Chromatography. *Chromatographia*. 1987;23:243-246.

190. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. *Multi- and Megavariate Data Analysis. Principles and Applications*. UMETRICS AB; 2001.

191. Wold S, Eriksson L, Trygg J, Kettaneh N. The PLS method -- partial least squares projections to latent structures-- and its applications in industrial RDP (research, development, and production). UMETRICS AB; 2004. Available at: http://www.umetrics.com/pdfs/events/prague%200408%20__%20PLS_text_wold.pdf.

192. Kubinyi H. Quantitative structure--activity relationships. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. *J Med Chem*. 1977;20:625-629.

193. Migliavacca E. Applied introduction to multivariate methods used in drug discovery. *Mini Rev Med Chem*. 2003;3(8):831-848.

194. Ritz C, Streibig JC eds. *Nonlinear Regression with R*. New York, NY: Springer New York; 2009.

195. Fox J. Nonlinear Regression and Nonlinear Least Squares. Appendix to An R and S-PLUS Companion to Applied Regression. 2002. Available at: http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-nonlinear-regression.pdf.

196. Haerdle W, Simar L. *Applied Multivariate Statistical Analysis*. 2nd ed. Berlin: Springer; 2003.

197. Van Belle G, Fisher LD, Heagerty PJ, Lumley T. *Biostatistics – a methodology for the health sciences*. 2nd ed. New Jersey: Wiley-Interscience; 2004.

198. Čekanavičius V, Murauskas G. *Statistika ir jos taikymai. II dalis*. Vilnius: TEV; 2004.

199. Japertas P, Didziapetris R, Petrauskas A. Fragmental Methods in the Design of New Compounds. Applications of The Advanced Algorithm Builder. *Quant Struct-Act Relat*. 2002;21:23-37.

200. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.r-project.org/.

201. Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Ionization-specific QSAR models of blood-brain penetration of drugs. *Chem Biodivers*. 2009;6(11):2050-2054.

202. Sugano K, Takata N, Machida M, Saitoh K, Terada K. Prediction of passive intestinal absorption using bio-mimetic artificial membrane permeation assay and the paracellular pathway model. *Int J Pharm*. 2002;241(2):241-251.

203. Sugano K, Nabuchi Y, Machida M, Aso Y. Prediction of human intestinal permeability using artificial membrane permeability. *Int J Pharm*. 2003;257(1-2):245-251.

204. Takano R, Sugano K, Higashida A, et al. Oral absorption of poorly water-soluble drugs: computer simulation of fraction absorbed in humans from a miniscale dissolution test. *Pharm Res*. 2006;23(6):1144-1156.

205. Hammarlund-Udenaes M, Fridén M, Syvänen S, Gupta A. On the rate and extent of drug delivery to the brain. *Pharm Res*. 2008;25(8):1737-1750.

206. Jeffrey P, Summerfield SG. Challenges for blood-brain barrier (BBB) screening. *Xenobiotica*. 2007;37(10-11):1135-1151.

207. Rodgers T, Leahy D, Rowland M. Physiologically based pharmacokinetic modeling 1: predicting the tissue distribution of moderate-to-strong bases. *J Pharm Sci*. 2005;94(6):1259-1276.

208. Rodgers T, Rowland M. Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *J Pharm Sci*. 2006;95(6):1238-1257.

209. Maurer TS, Debartolo DB, Tess DA, Scott DO. Relationship between exposure and nonspecific binding of thirty-three central nervous system drugs in mice. *Drug Metab Dispos*. 2005;33(1):175-181.

210. Kalvass JC, Maurer TS, Pollack GM. Use of plasma and brain unbound fractions to assess the extent of brain distribution of 34 drugs: comparison of unbound concentration ratios to in vivo p-glycoprotein efflux ratios. *Drug Metab Dispos*. 2007;35(4):660-666.

211. Waters NJ, Lombardo F. Use of the Øie-Tozer model in understanding mechanisms and determinants of drug distribution. *Drug Metab Dispos*. 2010;38(7):1159-1165.

212. Ermondi G, Lorenti M, Caron G. Contribution of ionization and lipophilicity to drug binding to albumin: a preliminary step toward biodistribution prediction. *J Med Chem*. 2004;47(16):3949-3961.

213. Jantratid E, Prakongpan S, Dressman JB, et al. Biowaiver monographs for immediate release solid oral dosage forms: cimetidine. *J Pharm Sci*. 2006;95(5):974-984.

214. Artursson P, Karlsson J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun*. 1991;175(3):880-885.

215. Neuhoff S, Ungell A-L, Zamora I, Artursson P. pH-dependent bidirectional transport of weakly basic drugs across Caco-2 monolayers: implications for drug-drug interactions. *Pharm Res*. 2003;20(8):1141-1148.

216. Vilar S, Chakrabarti M, Costanzi S. Prediction of passive blood-brain partitioning: straightforward and effective classification models based on in silico derived physicochemical descriptors. *J Mol Graph Model*. 2010;28(8):899-903.

217. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46:3-26.

218. Ozaki S, DeWald DB, Shope JC, Chen J, Prestwich GD. Intracellular delivery of phosphoinositides and inositol phosphates using polyamine carriers. *Proc Natl Acad Sci USA*. 2000;97(21):11286-11291.

219. Sugano K, Hamada H, Machida M, Ushio H. High throughput prediction of oral absorption: improvement of the composition of the lipid solution used in parallel artificial membrane permeation assay. *J Biomol Screen*. 2001;6(3):189-196.

220. Neubert R. Ion pair transport across membranes. *Pharm Res*. 1989;6(9):743-747.

221. Degim IT. New tools and approaches for predicting skin permeability. *Drug Discov Today*. 2006;11(11-12):517-523.

222. Audus KL. Controlling drug delivery across the placenta. *Eur J Pharm Sci*. 1999;8(3):161-165.

223. Obach RS. Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: An examination of in vitro half-life approach and nonspecific binding to microsomes. *Drug Metab Dispos*. 1999;27(11):1350-1359.

224. Austin RP, Barton P, Cockroft SL, Wenlock MC, Riley RJ. The influence of nonspecific microsomal binding on apparent intrinsic clearance, and its prediction from physicochemical properties. *Drug Metab Dispos*. 2002;30(12):1497-1503.

225. Hosea NA, Collard WT, Cole S, et al. Prediction of human pharmacokinetics from preclinical information: comparative accuracy of quantitative prediction approaches. *J Clin Pharmacol*. 2009;49(5):513-533.

226. Liu X, Smith BJ, Chen C, et al. Use of a physiologically based pharmacokinetic model to study the time to reach brain equilibrium: an experimental

analysis of the role of blood-brain barrier permeability, plasma protein binding, and brain tissue binding. *J Pharmacol Exp Ther*. 2005;313(3):1254-1262.

227. Liu X, Chen C, Smith BJ. Progress in brain penetration evaluation in drug discovery and development. *J Pharmacol Exp Ther*. 2008;325(2):349-356.

228. Parepally JMR, Mandula H, Smith QR. Brain uptake of nonsteroidal anti-inflammatory drugs: ibuprofen, flurbiprofen, and indomethacin. *Pharm Res*. 2006;23(5):873-881.

229. Becker DE, Rosenberg M. Nitrous oxide and the inhalation anesthetics. *Anesth Prog*. 2008;55(4):124-130; quiz 131-132.

230. Fridén M, Ducrozet F, Middleton B, et al. Development of a high-throughput brain slice method for studying drug distribution in the central nervous system. *Drug Metab Dispos*. 2009;37(6):1226-1233.

231. Fridén M, Winiwarter S, Jerndal G, et al. 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(20):6233-6243.

LIST OF PUBLICATIONS

Articles in peer-reviewed journals:

1. Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Ionization-specific prediction of blood-brain permeability. *J Pharm Sci*. 2009;98(1):122-134.
2. Reynolds DP, Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Ionization-specific analysis of human intestinal absorption. *J Pharm Sci*. 2009;98(11):4039-4054.
3. Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Ionization-specific QSAR models of blood-brain penetration of drugs. *Chem Biodivers*. 2009;6(11):2050-2054.
4. 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(6):2147-2160.

Book chapter:

Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Prediction of Blood-Brain Barrier Penetration by Drugs. In: Jain KK, ed. *Drug Delivery to the Central Nervous System*.Vol 45. 1st ed. Neuromethods. New York: Springer; 2010:63-83.

Conference abstracts and materials:

Oral presentations:

1. Lanevskij K, Japertas P, Didziapetris R, Petrauskas A. Ionization-Specific Prediction of Blood-Brain Permeability and Distribution of Drugs. LogP2009, Zürich, Feb. 8-11, 2009.
2. Lanevskij K, Japertas P, Didziapetris R. Ionization-Specific Prediction of Passive Permeability of Drugs Across Caco-2 Monolayers. Memtrans Workshop, Saarbrücken, Sep. 17-18, 2009.

Selected poster presentations:

1. Petrauskas A, Lanevskij K, Japertas P, Didziapetris R. Passive Transport of Mono-Protic Drugs' Ions. LogP2009, Zürich, Feb. 8-11, 2009.
2. Lanevskij K, Japertas P, Didziapetris R. Classification of Drugs by CNS Activity Based on QSAR Models of the Rate and Extent of Brain Delivery. 16th North American ISSX Meeting, Baltimore, MD, Oct. 18–22, 2009.
3. Lanevskij K, Japertas P, Didziapetris R. Trainable QSAR Model of Plasma Protein Binding and its Application for Predicting Volume of Distribution. The 240th ACS National Meeting, Boston, MA, Aug. 22-26, 2010.
4. Japertas P, Sazonovas A, Lanevskij K, Kassam K. Approach to quick lead optimization including physicochemical and ADME profiling. The 240th ACS National Meeting, Boston, MA, August 22-26, 2010.

SUMMARY IN LITHUANIAN (REZIUMĖ)

Šiame darbe pristatomi mechanistiniai kiekybinio struktūros ir aktyvumo ryšio modeliai, skirti vaistinių junginių savybių, charakterizuojančių jų absorbciją ir pasiskirstymą organizme prognozavimui. Nagrinėjama keletas parametrų, apibūdinančių paprastos difuzijos per biologines membranas greitį, taip pat termodinaminės konstantos, aprašančios vaistų pasiskirstymą tarp kraujo plazmos ir audinių.

Ląstelinių barjerų pralaidumas buvo modeliuojamas netiesinėmis lygtimis, siejančiomis paprastos difuzijos greitį su vaistų fizikocheminėmis savybėmis, tokiomis kaip lipofiliškumas, jonizacija, vandenilinių ryšių sudarymo potencialas ir molekulių dydis. Žarnyno epitelio pralaidumo modeliuose atsižvelgta į junginių difuziją tiek tarpląsteliniu, tiek viduląsteliniu keliu. Nustatyta, kad smegenų endotelyje ir žarnyno epitelyje stebima panašaus pobūdžio difuzijos greičio priklausomybė nuo jonizacijos – katijonai ir anijonai difunduoja atitinkamai 2 ir 3 eilėmis lėčiau už neutralias molekules.

Pademonstruota, kad analizuojant vaistų pasiskirstymo tarp audinių ir kraujo duomenis, būtina paversti pradines eksperimentines vertes kitais dydžiais, atspindinčiais vaistų jungimosi prie plazmos ir audinių komponentų stiprumą. Vaistų giminingumas audiniams gali būti aprašytas jų lipofiliškumu, o neigiama jonizacijos įtaka stebima tik rūgštiniams junginiams.

Visi sukurti modeliai užtikrina gerą koreliaciją tarp eksperimentinių ir suskaičiuotų savybių reikšmių, o vidutinė kvadratinė prognozavimo klaida patikrinamuosiuose duomenų rinkiniuose siekia 0,4-0,5 logaritminio vieneto. Taip pat parodyta, kad vaistų pernašos per hematoencefalinę užtvarą kiekybinių parametrų tiesinė kombinacija leidžia 94% tikslumu klasifikuoti vaistus pagal jų prieinamumą centrinei nervų sistemai.