Meta-analysis of permeability literature data shows possibilities and limitations of popular methods

Kateřina Storchmannová[1], Martin Balouch[2,3], Jakub Juračka[1,4], František Štěpánek[2], and Karel Berka[1*]

[1] Department of Physical Chemistry, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic.

[2] Department of Chemical Engineering, University of Chemistry and Technology, Prague, Technická 3, 166 28 Prague 6, Czech Republic.

[3] Zentiva k.s., U Kabelovny 130, 102 00 Prague 10, Czech Republic

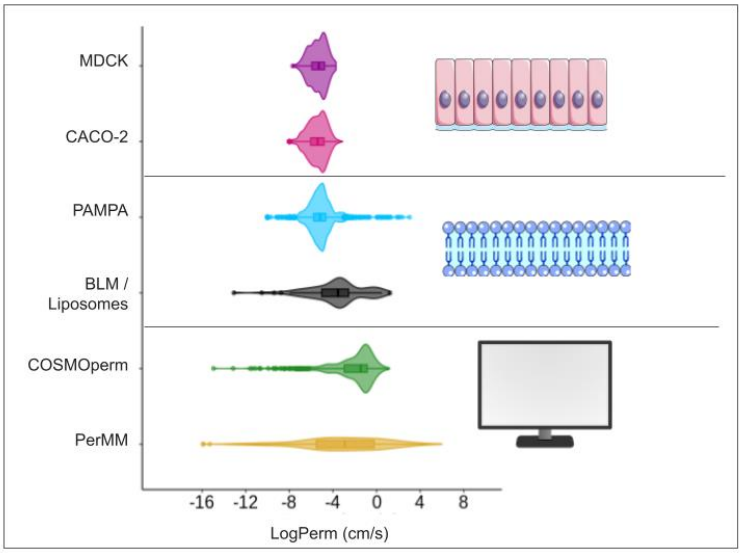
[4] Department of Computer Science, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic.

[*] Corresponding author, e-mail: karel.berka@upol.cz

Abstract

Permeability is an important molecular property in drug discovery, as it co-determines pharmacokinetics whenever a drug crosses the phospholipid bilayer, *e.g.*, into the cell, in the gastrointestinal tract or across the blood-brain barrier. Many methods for the determination of permeability have been developed, including cell line assays, cell-free model systems like PAMPA mimicking, *e.g.*, gastrointestinal epithelia or the skin, as well as the Black lipid membrane (BLM) and sub-micrometer liposomes. Furthermore, many *in silico* approaches have been developed for permeability prediction.

Meta-analysis of publicly available databases for permeability data (MolMeDB and ChEMBL) was performed to establish their usability. Firstly, experimental data can only be measured between thresholds for the lowest and highest permeation rate obtainable within physical boundaries. These thresholds vary strongly between methods. Secondly, computed data do not obey these thresholds but, on the other hand, can produce incorrect results. Thirdly, even for the same method and molecule, there is often a strong discrepancy between individual measured values. These differences are based not only on the statistics but also on the varying approaches and evaluation of the measured data. Thus, when working with in-house measured or published permeability data, we recommend to be cautious with their interpretation.



Introduction

Passive permeability is a critical molecular property studied in drug discovery because of its strong influence on pharmacokinetics. It plays an essential role in the gastrointestinal absorption of oral drugs, penetration of the blood-brain barrier (BBB), and renal reabsorption.[1]

The permeability coefficient ($cm \cdot s^{-1}$) is the quantitative measure of permeability, often presented as a decimal logarithm (logPerm). Numerous methods for their determination have been developed because of the importance of permeability coefficients in pharmaceutical research. We can, in general, divide the approaches into cell-based in vivo experimental assays, membrane-based in vitro experimental essays and in silico approaches.

Among the oldest and best-established methods of permeability measurement are the cell-based colon carcinoma cell line permeability assay (CACO-2)[2] and Madin-Darby Canine Kidney cells (MDCK)[3]. Permeability measurements are realized in transwell plates. Each well is divided into a donor and an acceptor compartment, separated by a membrane. In the case of CACO-2 and MDCK, the membrane consists of a cell monolayer cultured on a solid support. Despite their different origins, the CACO-2 and MDCK are composed of morphologically analogous cells and are widely used as model intestinal membranes[4].

Apart from cell-based experiments, there are several in vitro membrane-based methods. The most often used one is the parallel artificial membrane permeability assay (PAMPA)[5]. The membrane on which PAMPA methods are based is artificially made and chosen depending on the membrane the assay mimics. To this date, many variants of PAMPA have been published. The examples include DS-PAMPA[6], which mimics gastrointestinal absorption, blood-brain barrier PAMPA[7], SkinPAMPA[8], or nasal-PAMPA[9]. Another long-time-known experimental method of permeability measurement is BLM (Black Lipid membrane), first published by Mueller et al.[10]. In this experimental setup, membranes are prepared in the form of very thin lipid films. This method is suitable as a model of more complex natural membranes.[11,12] Unlike CACO-2 or MDCK, which employ a monolayer of complex living cells, PAMPA and BLM methods both use simpler membranes that are unable to effect active transport (influx as well as efflux), paracellular transport, metabolism, or ion-trapping in lysosomes.[13,14]

Experimental methods for the determination of membrane permeability have been supplemented by in-silico approaches, which can be divided into three main categories: molecular dynamics simulations, physics-based computational methods, and machine-learning statistical models.

Molecular dynamics (MD) simulations are *in silico* methods based on time-resolved simulations of complex systems at the atomistic level[15]. We can derive many thermodynamic and kinetic properties of the system from the MD simulations[15]. Thanks to the current level of computational power and MD methods, we can now study the behavior of substances on membranes even at the atomic level[16], but they are so far limited by the quality of membrane force fields[17,18], long time scales necessary for membrane permeation and hysteresis artefacts for advanced sampling methods[19]. Hence, the availability of these data in large quantities is still quite limited, and MD is used more to model how molecules permeate the membranes[20,21].

The PerMM[22] and COSMOperm[23] are examples of physics-based calculated methods. PerMM is based on the solubility diffusion model[24] and the positioning of proteins in membranes (PPM) method[25,26]. PerMM can also calculate the permeability coefficient across four types of membranes (DOPC, BLM, CACO/MDCK, and BBB)[22]. COSMOperm is a mechanistic method for the prediction of membrane permeability based on quantum chemical solubility calculations. Its basis is the calculation of the free energy profile across the membrane. In general, this approach can calculate logPerm on any membrane.[23]

Machine learning (ML) approaches are trained statistically over the existing experimental data, while the data quality and size are extremely important. The QSAR (quantitative structure-activity relationship) model is a mathematical model which identifies statistically significant correlations between the structure of molecules and their properties, such as biological activity[27]. The structure of molecules is described by a variety of descriptors. Choosing the fitting descriptor is one of the key points during the QSAR process. The history of QSAR is long, and countless various permeability QSAR models have been developed, *e.g.* QSAR models for CACO-2 cell permeability[28], intestinal permeability[29], blood-brain barrier (BBB)[30] and skin[31], etc. QSAR models were recently generalized with ML (machine-learning) models. These models are fitted to train data produced by experimental methods. Experimental methods often used for the construction of these models are CACO-2 (e.g., Wang et al.[32] or Frelund et al.[33]) and PAMPA (e.g., Sun et al.[34] or Gousiadou et al.[35]). Sometimes these models are created for a given type of molecule, e.g. cell-penetrating peptides[36] or macrocycles[37]. As their performance can be, in principle, only as good as the original data, we will not discuss them further here.

Because of the importance of permeability and the growing volume of published data obtained by various methods, permeability data are available in well-established cheminformatics databases (*e.g.*, PubChem[38] or ChEMBL[39]). Nevertheless, these databases do not primarily focus on this type of data, unlike the MolMeDB database.

MolMeDB[40] (https://molmedb.upol.cz) is a comprehensive, freely available database of membrane interaction data, including permeation for small molecules. This database stores the manually obtained data from scientific papers as well as the permeability data obtained from ChEMBL by data mining workflow. Currently, there are more than 900,000 interactions for almost 500,000 molecules in MolMeDB. Most of the data is permeability data from 56 theoretical or experimental methods on 48 various membranes.

This paper compares and interprets the results of four experimental methods – PAMPA, CACO-2, MDCK, BLM/Liposomes, and two calculated – PerMM and COSMOperm – available in MolMeDB. However, in order to understand this data properly, we must first understand the methods and their constraints. Therefore, this paper has three main aims: (i) to compare methods with each other to put the logPerm quantity in a real-world context, (ii) to identify and explain the limits of the mentioned methods, and (iii) to put the logPerm quantity in a real-world context.

Methods

Data sources

Data was sourced from MolMeDB and ChEMBL. The data in ChEMBL was fetched by the ChEMBL data web service. For this purpose, the KNIME[41] semi-automatic workflow was created. This workflow fetched information about molecules (SMILES, name, ChEMBL ID), publication (DOI), and interactions. All interactions were converted to decimal logarithms, and all units of interactions were converted to $cm \cdot s^{-1}$. Fetched data are available in MolMeDB. This data is labelled as ChEMBL in the Secondary reference column in MolMeDB. The content of MolMeDB was exported as a .csv file. This is possible on the website https://molmedb.upol.cz/stats/show_all.

Analysis of permeability data - MolMeDB data selection

Data was sourced from the MolMeDB database. The PAMPA method included methods that were referred to as EPAM, EBAMP[42](for apparent PAMPA), and EPAMOL (for insintic PAMPA).The BLM/liposomes included methods EBLM and ELIP in MolMeDB (for more details see: https://molmedb.upol.cz/browse/methods).

In the cases of scatter plots (Fig.2and Fig. S1), a mean logPerm value for molecules that have more than one logPerm value in MolMeDB was calculated. The mean values were calculated according to the rules described in the Colab notebook. Cleveland dot plots (Fig. 4 A and B) is created from median values of logPerm.

For greater inter-comparability of data, we excluded data other than that measured or calculated on the cell membranes, generic membranes, and membranes of the intestine according to the MolMeDB classification system (for more details, see: https://molmedb.upol.cz/browse/membranes). MolMeDB stores permeability coefficients (logPerm) uniformly in the logarithmic form of $cm \cdot s^{-1}$.

Next, the dataset was narrowed down to data for small molecules ($MW \leq 800$ Da). For the computational methods (PerMM and COSMOperm), interactions where the molecules were in a neutral state were included, since only neutral form is usually eligible to penetrate the lipid membrane[43–45]. In the case of experimental methods (CACO-2, MDCK, PAMPA, BLM/liposomes), only interactions for which the pH was between 7.1 and 7.5 were included. Also, molecules with unknown pH were included. Only data pertaining to a temperature of (20−25 °C) were included. 25 °C is the default temperature value in MolMeDB. In the case of the data from ChEMBL, the value of temperature is unknown, and for this reason, the approx. temperature of 25 °C is given in these cases. The resulting dataset contained data on 5483 interactions for 4218 unique molecules (by SMILES).

Analysis was done using KNIME workflow, and figures were created by R programming language version 4.3.2[46] or by Python 3.10.12. The Colab notebook[47] for figures preparation is available on MolMeDB GitHub (https://github.com/MolMeDB/MolMeDB), and the KNIMEworkflow is available on WorkflowHub[48] (https://workflowhub.eu/workflows/772). The UpSet plot was created by UpSetR Shiny App[49].

In the analyzed dataset, data originated from four different experimental methods – cell-based CACO-2 (1415 molecules) and MDCK (402 molecules); membrane-based PAMPA (2592 molecules) and BLM/liposomes (88 molecules); and two computational methods – PerMM (444 molecules) and COSMOperm (504 molecules).

Apparent and intrinsic permeability

In order to compare and analyze values of permeability, we must be able to distinguish between two concepts - apparent permeability and intrinsic (or molecular) permeability.

In a stirred container, the solute concentration is equalized in the bulk of the liquid. However, close to the membrane surface, molecules only move by diffusion rather than by convection. As a solute flows through the membrane, a concentration gradient builds up in close proximity to the membrane, weakening the driving force.[50]

According to equation 1, so-called measurable apparent permeability (p_{app}) is composed of contributions:
p_{UWL}, $f_{neutral}$ and p_M.[51]

$$\frac{1}{p_{app}} = \frac{1}{p_{UWL}} + \frac{1}{(f_{neutral}\, p_M)} \quad (1)$$

where p_{UWL} is the permeation through an unstirred water layer (UWL), $f_{neutral}$ is the fraction of molecule that is in a non-ionized state in the donor compartment, and p_M is the molecule's intrinsic permeability. This equation is plausible for the cases where the permeability of ionized species is negligible. That is often valid, but there are specific examples when this assumption is not fulfilled[52,53]. UWL is a static layer of water directly adjacent to the surface of a membrane, acting as an additional resistance to permeation. This value can vary with different stirring of donor and acceptor compartments, but for the most common experimental setting of CACO-2/MDCK or PAMPA assay, this value is around −3.9.[13]

Intrinsic permeability is often obtained from the calculation based on measuring the permeability scale at different pH levels. This approach can be found in publications by Huque et al.[54], Avdeef et al.[55], or Tsinman et al.[56]. Furthermore, other implementations of this approach showed Velický et al.[57], where the permeabilities are measured at different hydrodynamic regimes and from that, intrinsic permeability can also be calculated.

Here, it can be noted that the apparent permeability is easily calculable from the intrinsic permeability by taking into consideration the fraction of non-ionized molecules (from pK_a) and the permeation rate through an unstirred water layer, which is specific to the experimental setup and can be determined from data for fast-permeating molecules. Conversely, the determination of intrinsic permeability from the apparent one is only feasible when the apparent permeability is not close to the diffusion limit. This relation limits the usefulness of published apparent permeability data for intrinsic permeability models.

Results and discussion

Repeatability of the data

Firstly, we wanted to analyze the repeatability of the data to get the gist of the reported value variability necessary to establish error estimation. For this purpose, permeability measurements for the same molecules were considered using the same method at similar conditions (see Method section). The most prevalent experimental method in the MolMeDB is PAMPA. We have taken seven most studied molecules. They are listed in Table 1, along with the number of measured data points, average permeability, and standard deviation.

Table 1: Molecules with the most data measured by the PAMPA method. Data taken from MolMeDB. P_{intr} - number of intrinsic permeabilities, P_{app} - number of apparent permeabilities

	P_{intr}	P_{app}	Average LogPerm (cm/s)	Standard deviation (cm/s)	Minimum value (cm/s)	Maximum value (cm/s)
Antipyrine	1	6	-5.70	0.34	-6.09	-5.09
Carbamazepine	1	7	-4.83	0.65	-5.8	-3.89
Ketoprofen	1	12	-5.50	0.49	-6.39	-4.32
Naproxen	2	4	-4.38	1.62	-6.2	-1.69
Propranolol	3	14	-4.10	1.48	-6.68	-0.21
Theophylline	3	5	-5.94	0.78	-7.4	-5.07
Verapamil	3	6	-3.74	1.81	-5.19	-0.89

When visualizing the data from these seven molecules measured by PAMPA (Fig. 1), we see several different behaviors of molecules, that can be discerned between intrinsic and apparent permeabilities.

Firstly, there are molecules with very low variance between measured values, such as antipyrine or ketoprofen, within 0.5 log unit. Further molecules, such as carbazepine or theophylline, have reasonable variances within 1 log unit. The Table 1 shows several molecules with variance within 2 log units – hence individual permeability measurements differ by more than 2 magnitudes. However, these large errors can be explained by the mixing of apparent and intrinsic permeability, that Fig. 1 show to be several log units different for those molecules.

Unfortunately, the information on whether intrinsic or apparent permeability is reported is not always properly described in the literature or databases (e.g., we had to implement PAMPA P_{intr} into MolMeDB upon preparation of this manuscript). The lack of distinction between apparent and intrinsic permeability can lead to complications, especially when comparing data from multiple papers using the PAMPA method.

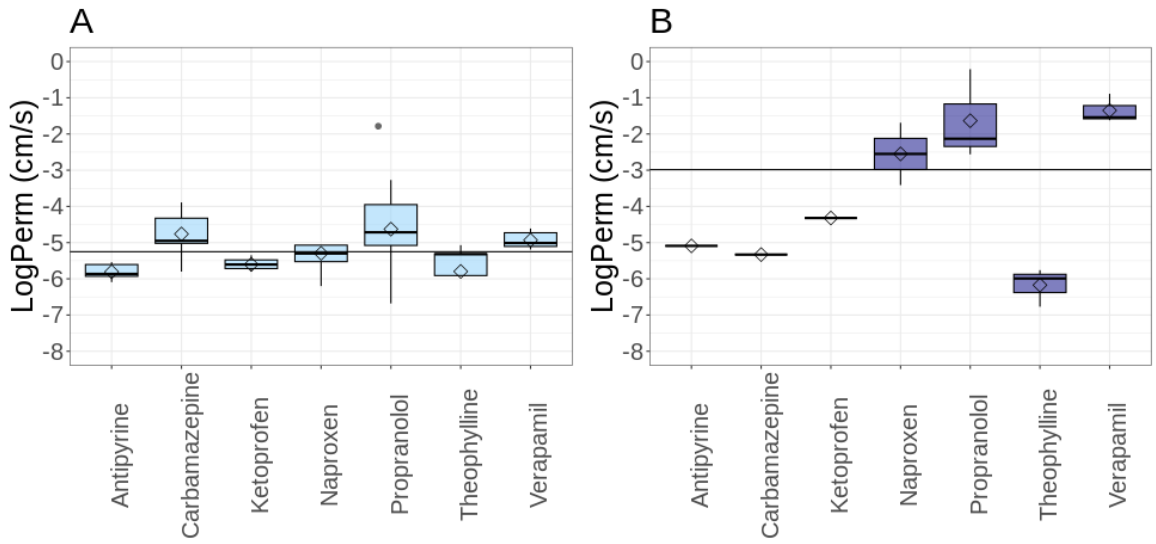


Fig. 1: Box plot of seven molecules with the most data measured by the PAMPA method (in MolMeDB). A – apparent permeability, B – intrinsic permeability. Rhombuses represent mean values of permeability, and the line corresponds to the median value of the dataset. Dots represent outliers.

A similar lack of information affects cell-based methods. There is crucial information about the direction of measurement or membrane transport proteins. There are two possible directions – either from apical to basolateral (from A to B) or from basolateral to apical (from B to A). The difference of permeabilities in direction can be large, e.g. in Colabufo et al.[58]. The direction affects the value of measured permeability because these cells have an asymmetrical expression of pharmacologically relevant proteins that influence molecular transport.[59] These proteins are (i) efflux transporters, e.g., P-glycoprotein (MDR1, ABCB1), MRP2 (ABCC2) or BCRP (ABCG2), and (ii) uptake proteins, e.g., OCT2 (SLC22A2), OATP2B1 (SLCO2B1), or PEPT1 (SLC15A1)[59,60]. In the case of these measurements, not only the direction in which the permeability is measured but also the presence of transporter inhibitors plays a role[33,59]. The inhibitors increase the permeability of substrates of efflux transporters[33]. Gene knockout techniques can also alter the expression of these transporters in the cells[60]. Unfortunately, this information is often not sufficiently reported together with permeability data in the public available databases.

Comparison of methods

In this analysis, we have studied the correlation between each pair of permeability methods to see how they can be supplemented with each other if needed. A mean logPerm value for each molecule and method was calculated, and for molecules that were measured/calculated by at least two methods, these data were compared. The comparison of all possible pairs of methods is available in Supporting Information (Fig. S1). Here, we discuss the most prominent and illustrious examples.

The first finding is the strong correlation (R^2 = 0.82) between CACO-2 and MDCK (Fig. 2A). This correlation is expected because both methods are cell-based methods, and their strong correlation was described by Irvine et al.[3]

From the previous section, we know the PAMPA dataset contains a mixture of apparent and intrinsic permeabilities that need to be differentiated if the data are to be compared. Fig. 2B is the correlation

among CACO-2 & MDCK vs intrinsic PAMPA. This correlation is very weak (R^2 = 0.03) because, in contrast to PAMPA, the CACO-2 and MDCK methods can provide only apparent permeability. On the other hand, Fig. 2C shows the correlation (R^2 = 0.44) between CACO-2 & MDCK vs apparent PAMPA. This correlation is unsurprisingly stronger because the PAMPA apparent permeabilities correlate well with CACO-2 and MDCK data. This observation is consistent with the literature because the correlation among these methods was already described in the literature (PAMPA and CACO-2 were described by Zhu et al.[61], MDCK and PAMPA were published by von Richter et al.[62]). However, in addition to the membranes, the cell-based methods CACO-2 and MDCK also have membrane transport proteins, which can influence the transport of molecules through the membrane, and thus lower the correlation. However, the correlation between CACO-2 and PAMPA and MDCK and PAMPA indicates passive diffusion as a dominant transport mechanism in the case of both cell monolayers[11,63]. A good correlation between CACO-2 & MDCK vs apparent PAMPA can indicate a smaller role of membrane transporters in this dataset.

Fig. 2D shows a strong correlation (R^2 = 0.73) between the calculated methods (PerMM and COSMOperm) and BLM / liposomes (abbreviated "BLM"). The BLM method is a very important experimental method because, in contrast to other experimental methods, the BLM is the diffusion-unlimited method. Thus, molecular permeability is easily determined using BLM. The BLM is not widespread, and therefore, we do not have as much data available as the PAMPA method.

The authors of both calculated methods validated these methods against BLM experimental data[22,23]. Bittermann et al.[64] report that COSMOperm has RMSE = 0.62 log units for neutral molecules and RMSE = 0.7 log units for ions. The PerMM method has RMSE = 1.15 log unit by Lomize et al.[22]. Therefore, it is no surprise that these methods correlate strongly. In addition, highly similar BLM datasets were used for validation, and this validation data comprises the majority of the BLM data in the MolMeDB database. We also see that the values are wide ranged because BLM, COSMOperm, and PerMM are methods that are not constrained by diffusion limits and provide intrinsic permeabilities.

Fig. 2E is a correlation between calculated (PerMM and COSMOperm) (R^2 = 0.58). This correlation is not surprising, given what has been said about these methods above. Both methods are unlimited by diffusion limit and can predict intrinsic permeabilities.

Fig. 2F is an example of the weak correlation between diffusion-limited method (PAMPA) and -unlimited method (PerMM). More examples can be found in Supporting information (Fig. 1S H and O). As we can clearly see, PAMPA data are located in the range of values (approx. from −8 to 4 log units), but the PerMM method is in the wider range of values (approx. from −16 to 4). Differences between the logPerm values from PerMM and PAMPA can be huge (several log units), although the PerMM method was successfully evaluated by PAMPA-DS (PAMPA-DS: R^2 = 0.75, RMSE = 1.59 log unit) by Lomize et al. [22]. Our correlation is weaker (R^2= 0.27) than Lomize's because our dataset contains apparent permeabilities and intrinsic permeabilities, whereas Lomize used only intrinsic permeabilities from one source.

Apart from the correlation, it is often more useful to calculate mean absolute error (MAE) for each comparison (Table 2). It shows that the closest pair of methods are both cell-based methods (CACO-2 and MDCK) followed by their pairs with their membrane-based counterpart sharing similar range – apparent PAMPA. The error between BLM and both computational methods is comparable to the MAE in between them, but their similarity to cell-based methods is weak with the largest error. As a negative control, we have tried mean predictor, i.e. we calculated MAE of each method towards the mean average value calculated on its dataset. This value serves as a negative control for the fit to that dataset. If the MAE value for pair is lower or at least similar than MAE to mean predictor, then it can be combined. This comparison has shown that we can combine both cell-based methods (CACO-2 and MDCK) together with apparent PAMPA. Similarly, both computed physics-based methods (COSMOperm and

PerMM) can be used to predict membrane-based BLM method and to some extent also to intrinsic PAMPA.

Table 2. Mean absolute errors (MAE) for each pair of methods. Colors define indicate the ranges – green has MAE range < 1 log units, yellow has MAE range between 1 and 3 log units, and red has MAE range > 3 log units. Diagonal shows MAE of mean predictor, i.e. towards mean average value of each dataset. This value serves as a negative control for the fit to that dataset. If the MAE value for pair is lower or at least similar than MAE to mean predictor, then the datasets can be combined.

MAE	CACO	MDCK	PAMPA app.	PAMPA intr.	BLM	COSMO perm	PerMM
CACO	**0.72**	0.34	0.73	2.81	2.39	2.79	4.02
MDCK	0.34	**0.65**	0.74	2.00	3.49	2.76	3.34
PAMPA app.	0.73	0.74	**0.85**	2.53	3.30	3.56	3.95
PAMPA intr.	2.81	2.00	2.53	**1.59**	3.26	1.78	2.44
BLM	2.39	3.49	3.30	3.26	**1.90**	0.97	1.42
COSMO perm	2.79	2.76	3.56	1.78	0.97	**1.71**	1.98
PerMM	4.02	3.34	3.95	2.44	1.42	1.98	**3.11**

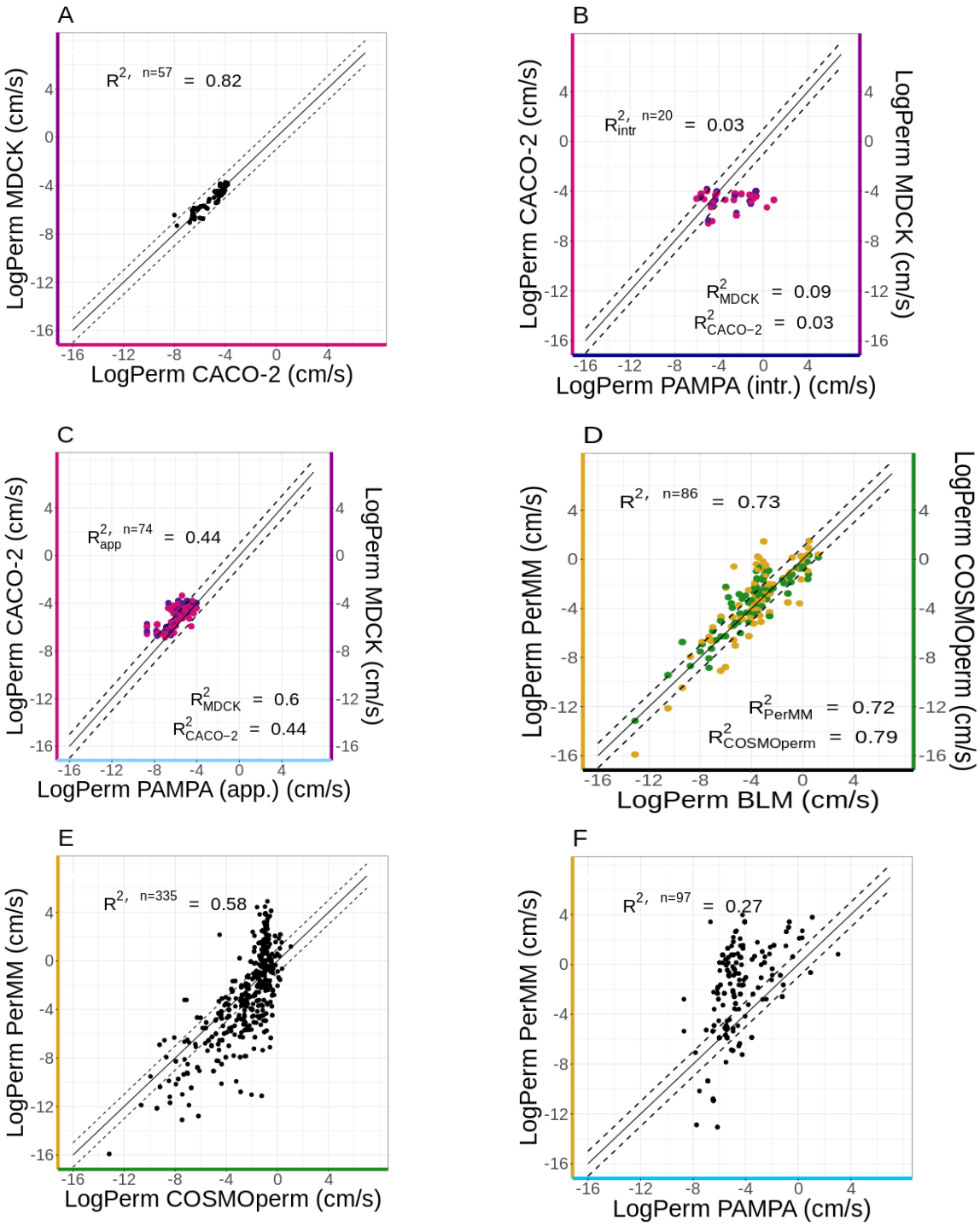


Fig. 2: Mean logPerm values of molecules that are present in overlaps between A: CACO-2 and MDCK datasets; B: CACO-2 & MDCK and intrinsic PAMPA datasets, C: CACO-2 & MDCK and apparent PAMPA datasets D: COSMOperm & PerMM and BLM/Liposomes datasets, E: PerMM and COSMOperm datasets, F: PerMM and PAMPA datasets (apparent and intrinsic permeabilities). The solid line represents the parity of permeation values, and the dashed lines represent a logPerm difference of ±1 between methods. All data is displayed in a log of $cm \cdot s^{-1}$. *N* is the number of unique molecules in overlap, and R^2 is the coefficient of determination.

Overlaps of the methods

All the above-mentioned methods are well-known, and the permeabilities of small molecules determined by these methods have been published in many publications.

The UpSet plot (Fig. 3) shows overlap between all six methods by molecules (MolMeDB IDs). The biggest overlap is between both calculated methods (176 molecules); the second biggest one is the overlap among PerMM, COSMOperm, and BLM (63 molecules), and all other overlaps are much smaller.

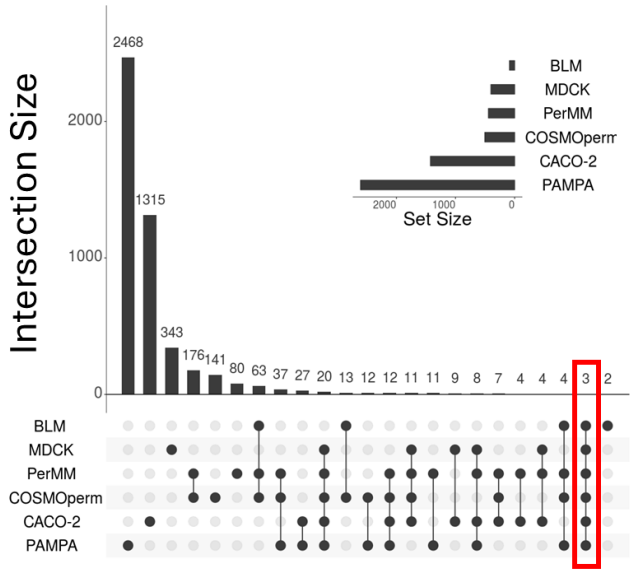


Fig. 3: Upset plot of six methods (created with UpSetR Shiny App). Columns in the graph represent a number of molecules that were measured by the combination of methods shown by black circles only. The first column shows molecules that were measured by PAMPA only, and the last column shows molecules that were measured by PAMPA, calculated by PerMM but which are not present in any other dataset. This figure includes only intersections with more than one molecule. The overlap among all six methods is highlighted by red rectangle.

Overlap among all six methods contains three well-known drugs (hydrocortisone, salicylic acid, and acetylsalicylic acid). Fig. 4A is the Cleveland dot plot of logPerm median values from individual methods for each of these molecules. Fig. 4B shows median permeability values for molecules that are present in all datasets except BLM due to the low number of molecules in the BLM/liposome dataset. The data shown in Fig. 4 underline the phenomena from the previous chapter. MDCK, CACO-2, and apparent PAMPA (purple, pink, and light blue points) often give very similar results. In the case of atenolol, verapamil, or warfarin, the difference between apparent PAMPA (light blue points) and intrinsic PAMPA (navy blue points) can be huge. Also, the figure demonstrates the variability of logPerm values from all calculated methods (green and yellow points).

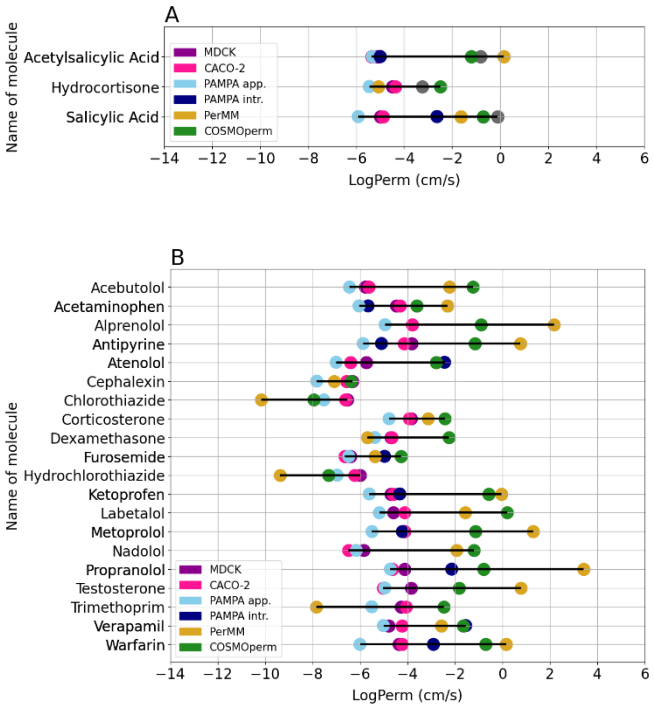


Fig. 4: Cleveland dot plot of median permeability coefficients for molecules in the overlap of MDCK, CACO-2, PAMPA, BLM / Liposomes, COSMOperm, and PerMM datasets.

The limits of permeability of the compared methods

Distributions of values for individual methods have shown that each method has some limits of reported permeability that differ. Fig. 5 shows that almost all CACO-2 and MDCK experimental values fall in a fairly narrow range from −8 to −4 log units. The lower limit is probably due to the duration of the experiments. It is hard to measure slower permeation in experimental conditions within ambient times. The upper limit (around −3.9) corresponds to the diffusion limit through an unstirred water layer (UWL),[13] as explained above. Here, it can be noted that at least 40% of molecules studied using CACO-2 or MDCK assays were close to this diffusion limit. Therefore, a significant amount of measured permeability is, in fact, diffusion of the unstirred water layer, and their intrinsic permeability is thus anywhere between −4 and 4. However, for example Stenberg et al, published CACO-2 data where they tried to reduce the effect of UWL on permeability by stirring[65].

Data from the PAMPA experimental assay (Fig. 5 PAMPA) has a visible peak around logPerm = −4, though higher permeation rates were also measured. This phenomenon is again caused by the above-mentioned mixture of apparent and intrinsic PAMPA permeabilities in the literature. Some publications (e.g. refs. [63,66,67]) report the apparent permeability as a direct experimental value of permeation, which is related to CACO-2 permeability. On the other hand, other publications of PAMPA permeation data (e.g. refs. [54,55]) report the intrinsic permeability, thus mixing permeability values. As can be seen from the figure, the apparent permeability (light blue) peaks between the values -8 and -4, while the intrinsic permeability (navy blue) is shifted to higher values. However, some values of apparent permeability are over −4 threshold. This phenomenon can be explained by e.g. effort to reduce the UWL layer in permeability assay by increasing of stirring speed[68]. Fujikawa et al present e.g. for Desipramine logPerm = −4.77 (0 rpm), logPerm = −4.00 (200 rpm), and logPerm = −3.81 (250 rpm)[67].

Permeabilities measured by the BLM/liposome method are typically higher than −4, likely due to the relatively large membrane area in liposomal systems. Although the BLM/liposome method also provides apparent permeabilities, the UWL, in this case, is significantly smaller than that of CACO-2, MDCK or PAMPA, and its contribution is practically negligible. The lowest experimentally measured permeability value is −13.1 log units for saccharose from Brunner et al. [69]. Its authors say that any value lower than −10 log unit is hardly measurable.

With the computational approaches – PerMM and COSMOperm (Fig. 5 PerMM and COSMOperm), we observe a broad distribution of permeation rates from very slow permeation (logPerm ≤ −8 $cm \cdot s^{-1}$) to very fast permeation (logPerm ≥ −4 $cm \cdot s^{-1}$). This is typical for calculated methods because they have no experimental limits. We can compare the result from the experimental method with the result from the calculation, but only up to the limits of the experimental methods. Beyond these limits, there is no possibility of comparison. The permeabilities obtained by these calculated methods can be categorized as intrinsic permeabilities.

In addition, it is interesting to mention differences in averages of octanol/water partition coefficient (LogP) and molecular weight (MW) for different methods.

CACO-2, MDCK, and PAMPA have considerably higher averages of both values than the other three methods. This is probably caused by the fact that most of the permeating molecules measured in these assays in scientific publications are drug-like molecules, and the molecular weight of around 400 Da and logP around 3 corresponds to a typical drug candidate.

BLM/liposome methods have much lower MW and logP (MW = 159.4 Da, logP = 0.3). This is probably because these methods are not commonly used for extensive drug candidate molecule assays; rather, there are only a few measurements of typical small molecules (e.g., benzoic acid). These data are then often used as training/validation of different computational methods since they are not influenced by the -8 and -4 thresholds that are hard (or even impossible) to overcome for other experimental methods.

Computational methods (COSMOPerm, PerMM) can predict the permeabilities of a wide range of molecules. Their average LogP of around 1.8 and MW of around 260 Da is lower than for experimental methods. This has at least two co-occurring reasons. Firstly, using computational methods, it is possible to calculate small molecules (oxygen, water, carbon dioxide) that are commonly not measured in permeation assays. Secondly, the calculation difficulty scales with molecular weight, and therefore, it is more expensive to calculate molecules with larger MW.

352

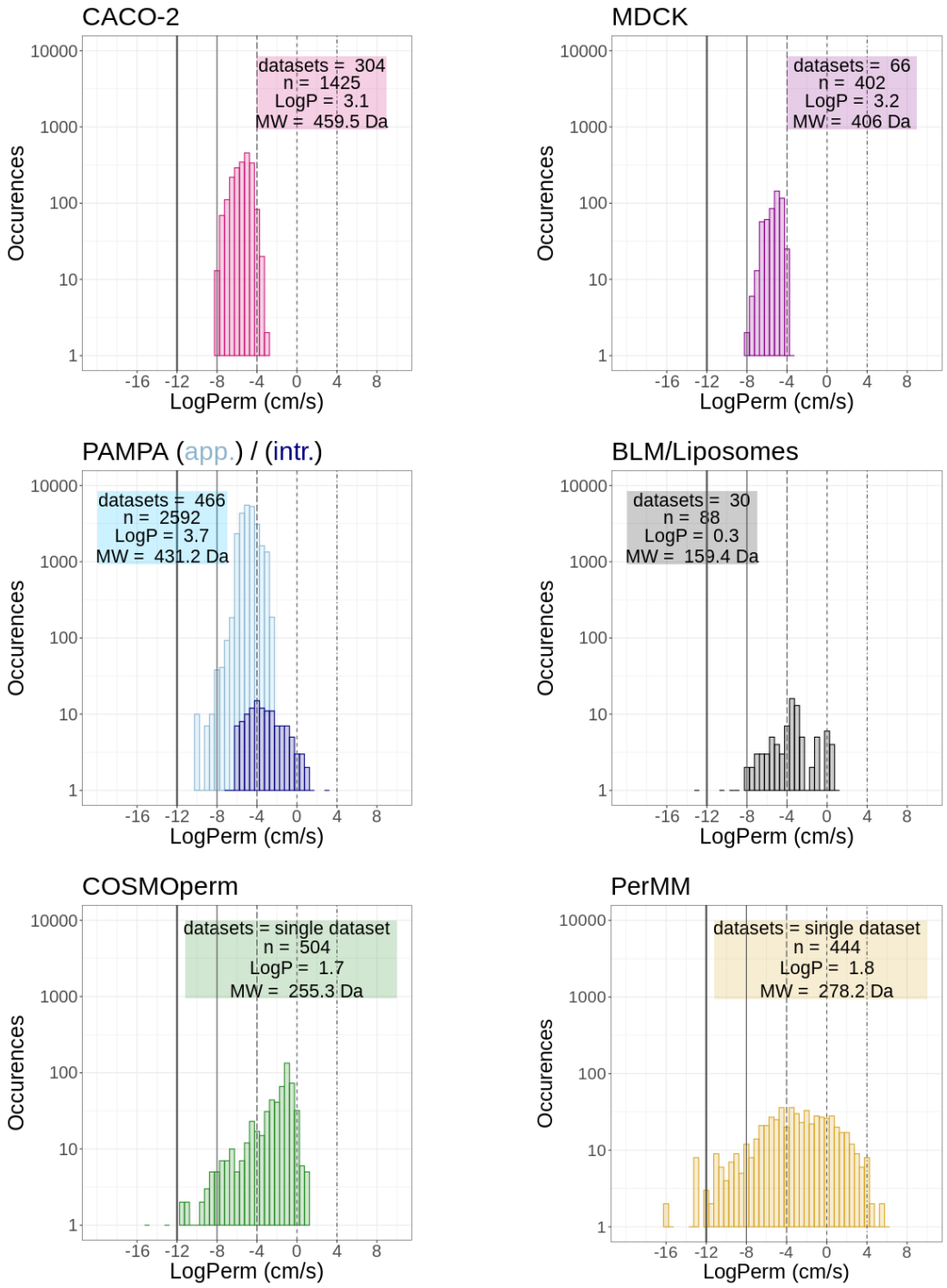


Fig. 5: Distribution of selected LogPerm values (uncharged molecules, 25 °C, smaller than 800 Da) from MolMeDB database according to selected methods. Datasets – number of unique datasets (by primary reference), LogP – mean logP of unique molecules, MW – mean molecular weight of unique molecules, n - number of unique molecules (by SMILES) in combined datasets. All data is displayed in a log of cm/s, bin size = 0.5. The vertical lines emphasize values −12, −8, −4, 0, +4, as discussed above. Occurrences are in log scale.

353

Interpreting permeability coefficients real-world time scales

For the interpretation of the limits we observed within the previous sections, we have designed several simplified boundaries as a multiplication of 4 log units, that can help to explain the limitations of permeability coefficients in real world examples of the time scale for permeation events (Fig. 6).

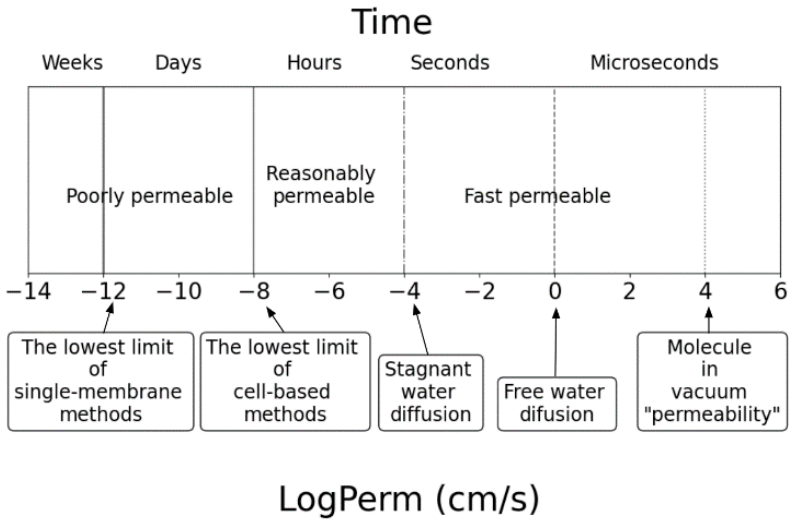


Fig. 6: Illustration of permeability coefficients in the context of timescale

The lowest experimental value of logPerm is −13.1 log units in our dataset (Fig 3 BLM/liposomes)[69]. Hence, the first line of logPerm around −12 log units corresponds to the lowest permeation rates still measurable by single-membrane experimental methods. It is close to the practical limit of the slowest passive permeation that still results in a biologically feasible amount of e.g. highly toxic compounds permeating over a physiologically relevant amount of time. For a permeation area equal to that of an entire human intestine (30 m^2)[70] and 0.1 L as a volume of the intestinal fluid,[71] only 0.25 % of a permeant will have permeated in ten days at such rate with logPerm = −12 log units.

The second line is logPerm around −8 log units. This area corresponds to the lowest limit of cell-based methods (e.g., CACO-2 or MDCK) as well as PAMPA. With the typical measurement setup for these measurements (donor volume 0.5 mL, permeation area 1.4 cm^2),[72] the compound with logPerm = −8 log units will permeate approx. 2% of the permeant in 10 days. The same limit is observable in, e.g. Deur et al. [73] for CACO-2, Chiba et al.[74] for MDCK, or Flaten et al. [75] for PAMPA.

The third important value of logPerm is −4 log units. This represents the value of −3.9 log unit that corresponds to the effect of an unstirred water layer (UWL) as we already discussed earlier in previous section. Hence the typical permeability coefficient values will fall in between logPerm −8 to −4 log units.

Fourth line of logPerm around 0 log units relates to an unrestrained diffusion in water and represents the maximum permeability measurable by experimental methods in water medium. Such logPerm values correspond to the permeation of molecules through water slice of similar thickness to membrane, where there is no energy barrier for permeation and diffusion coefficient is equal to the water self-diffusion coefficient. Then, the homogeneous solubility model for the permeation coefficient is valid:

$$P = \frac{K \cdot D}{L} \qquad (2)$$

where K is the partition coefficient between the membrane and water phase (considered equal to 1, if there is no extra partitioning into the slice due to no energy barrie), D is the diffusion coefficient

(3×10^{-9} m^2/s for water self-diffusion, taken from ref. [76]), and L is the thickness of the membrane (the thinnest experimentally possible membrane with a thickness around 4 nm). Then, P is 75 cm/s, and logPerm is +1.9 log units. Since drugs are bigger molecules than water, they usually have diffusion constant lower by order of magnitude or more (e.g., ibuprofen has $D = 5.5\times10^{-10}$ m^2/s)[77], we have set this simplified limit to 0 log unit.

The rightmost highlighted value of logPerm is around +4 log units. This value of logPerm presents the theoretical upper limit of permeability, describing an unrestrained molecule travelling through a vacuum (together with a spherical chicken from classical physics joke). This was calculated using the formula for the root mean square velocity of a gas molecule (equation 3):

$$v = sqrt(^{3kT}/_{m}) \quad (3)$$

Where k is the Boltzmann constant, T is the temperature (K), and m is the mass of the molecule. If we assume the mass of the molecule as 200 g·mol^{-1} and a temperature of 37 °C, we get a velocity of 6000 m·s^{-1}, which can be considered as a gas permeability limit (logPerm = +5.8 log units of cm·s^{-1}) but only in the case of a negligible thickness of the membrane, area of permeation equal to the projection is of the molecule itself and maximum possible concentration difference. This value is purely hypothetical and does not correspond to the biomembrane permeability in real liquid conditions. It is only stated here as the absolute upper limit of the permeability. Even the 60 times smaller value of permeability (logPerm = +4 log units of cm·s^{-1}) is still purely hypothetical and impossible to get in biomembrane permeation; thus, this limit is used in graphs and discussion below due to the 4 orders of magnitude difference between all other limits.

Conclusions

In summary, we have meta-analyzed a large amount of permeability data from freely available databases MolMeDB and ChEMBL. Permeability is, among other things, the basis of classifying drug substances into the Biopharmaceutics Classification System (BCS)[78] and this classification forms the basis of product formulation and regulatory approval strategy decisions, hence it is important to have reliable data for permeability. Moreover, permeability as an important pharmacological property is of interest to many researchers who try to create machine learning algorithms for its prediction. The variability between individual measurements, even for the same methods, has shown that efforts should be made to develop robust methods that would enable consistent inter-laboratory values to be measured and stored in FAIR manner, e.g. in MolMeDB database. Next the analysis of individual methods observed their data limits. Meta-analysis in-between the datasets has shown that cell-based methods such as CACO-2, MDCK are comparable with apparent PAMPA, but all of these methods correlate less with calculated physics-based methods (COSMOperm and PerMM) and with single membrane-based BLM/liposomes or intrinsic PAMPA, which are based on molecular permeabilities. This needs to be considered when permeability data are from different methods compared or used in machine-learning approaches. Finally, we have devised a scale with five significant permeability values as a multiplier of 4 log units of cm·s^{-1} to help a comprehensive understanding of the permeability data within their physical context.

Author Contributions

K.S. – data analysis, data curation, visualization, manuscript writing M.B. – data analysis, data interpretation, manuscript writing, J.J. – MolMeDB development, data curation, F.Š. - conceptualization, funding acquisition, manuscript writing, K.B. – conceptualization, funding acquisition, manuscript writing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge the help of Adam Tywoniak in carefully reading the manuscript. The project was supported by the Czech Grant Agency project GAČR 24-11986S. K.S. and K.B. acknowledge support from Palacky University Olomouc Project IGA_PrF_2024_017. K.B. and J.J. acknowledge support from the ELIXIR-CZ infrastructure (Project LM2023055). The graphical abstract also includes icons adapted from Bioicons.com (https://bioicons.com/).

References

(1) Di, L.; Artursson, P.; Avdeef, A.; Benet, L. Z.; Houston, J. B.; Kansy, M.; Kerns, E. H.; Lennernäs, H.; Smith, D. A.; Sugano, K. The Critical Role of Passive Permeability in Designing Successful Drugs. *ChemMedChem* **2020**, *15* (20), 1862–1874. https://doi.org/10.1002/cmdc.202000419.

(2) Hidalgo, I. J.; Raub, T. J.; Borchardt, R. T. Characterization of the Human Colon Carcinoma Cell Line (Caco-2) as a Model System for Intestinal Epithelial Permeability. *Gastroenterology* **1989**, *96* (3), 736–749.

(3) Irvine, J. D.; Takahashi, L.; Lockhart, K.; Cheong, J.; Tolan, J. W.; Selick, H. E.; Grove, J. R. MDCK (Madin-Darby Canine Kidney) Cells: A Tool for Membrane Permeability Screening. *Journal of Pharmaceutical Sciences* **1999**, *88* (1), 28–33. https://doi.org/10.1021/js9803205.

(4) Fedi, A.; Vitale, C.; Ponschin, G.; Ayehunie, S.; Fato, M.; Scaglione, S. In Vitro Models Replicating the Human Intestinal Epithelium for Absorption and Metabolism Studies: A Systematic Review. *Journal of Controlled Release* **2021**, *335*, 247–268. https://doi.org/10.1016/j.jconrel.2021.05.028.

(5) Kansy, M.; Senner, F.; Gubernator, K. Physicochemical High Throughput Screening: Parallel Artificial Membrane Permeation Assay in the Description of Passive Absorption Processes. *J. Med. Chem.* **1998**, *41* (7), 1007–1010. https://doi.org/10.1021/jm970530e.

(6) Bermejo, M.; Avdeef, A.; Ruiz, A.; Nalda, R.; Ruell, J. A.; Tsinman, O.; González, I.; Fernández, C.; Sánchez, G.; Garrigues, T. M.; Merino, V. PAMPA—a Drug Absorption in Vitro Model. *European Journal of Pharmaceutical Sciences* **2004**, *21* (4), 429–441. https://doi.org/10.1016/j.ejps.2003.10.009.

(7) Di, L.; Kerns, E. H.; Fan, K.; McConnell, O. J.; Carter, G. T. High Throughput Artificial Membrane Permeability Assay for Blood–Brain Barrier. *European Journal of Medicinal Chemistry* **2003**, *38* (3), 223–232. https://doi.org/10.1016/S0223-5234(03)00012-6.

(8) Ottaviani, G.; Martel, S.; Carrupt, P.-A. Parallel Artificial Membrane Permeability Assay: A New Membrane for the Fast Prediction of Passive Human Skin Permeability. *J. Med. Chem.* **2006**, *49* (13), 3948–3954. https://doi.org/10.1021/jm060230+.

(9) Henriques, P.; Bicker, J.; Silva, S.; Doktorovová, S.; Fortuna, A. Nasal-PAMPA: A Novel Non-Cell-Based High Throughput Screening Assay for Prediction of Nasal Drug Permeability. *International Journal of Pharmaceutics* **2023**, *643*, 123252. https://doi.org/10.1016/j.ijpharm.2023.123252.

(10) Mueller, P.; Rudin, D. O.; Ti Tien, H.; Wescott, W. C. Reconstitution of Cell Membrane Structure in Vitro and Its Transformation into an Excitable System. *Nature* **1962**, *194* (4832), 979–980. https://doi.org/10.1038/194979a0.

(11) Sugano, K.; Kansy, M.; Artursson, P.; Avdeef, A.; Bendels, S.; Di, L.; Ecker, G. F.; Faller, B.; Fischer, H.; Gerebtzoff, G.; Lennernaes, H.; Senner, F. Coexistence of Passive and Carrier-Mediated Processes in Drug Transport. *Nat Rev Drug Discov* **2010**, *9* (8), 597–614. https://doi.org/10.1038/nrd3187.

(12) Winterhalter, M. Black Lipid Membranes. *Current Opinion in Colloid & Interface Science* **2000**, *5* (3–4), 250–255. https://doi.org/10.1016/S1359-0294(00)00063-7.

(13) Bittermann, K.; Goss, K.-U. Predicting Apparent Passive Permeability of Caco-2 and MDCK Cell-Monolayers: A Mechanistic Model. *PLoS ONE* **2017**, *12* (12), e0190319. https://doi.org/10.1371/journal.pone.0190319.

(14) Berben, P.; Bauer-Brandl, A.; Brandl, M.; Faller, B.; Flaten, G. E.; Jacobsen, A.-C.; Brouwers, J.; Augustijns, P. Drug Permeability Profiling Using Cell-Free Permeation Tools: Overview and Applications. *European Journal of Pharmaceutical Sciences* **2018**, *119*, 219–233. https://doi.org/10.1016/j.ejps.2018.04.016.

(15) Padhi, A. K.; Janežič, M.; Zhang, K. Y. J. Molecular Dynamics Simulations: Principles, Methods, and Applications in Protein Conformational Dynamics. In *Advances in Protein Molecular and Structural Biology Methods*; Elsevier, 2022; pp 439–454. https://doi.org/10.1016/B978-0-323-90264-9.00026-X.

(16) Guo, J.; Bao, Y.; Li, M.; Li, S.; Xi, L.; Xin, P.; Wu, L.; Liu, H.; Mu, Y. Application of Computational Approaches in Biomembranes: From Structure to Function. *WIREs Comput Mol Sci* **2023**, e1679. https://doi.org/10.1002/wcms.1679.

(17) Paloncýová, M.; Fabre, G.; DeVane, R. H.; Trouillas, P.; Berka, K.; Otyepka, M. Benchmarking of Force Fields for Molecule–Membrane Interactions. *J. Chem. Theory Comput.* **2014**, *10* (9), 4143–4151. https://doi.org/10.1021/ct500419b.

(18) Kiirikki, A. M.; Antila, H. S.; Bort, L. S.; Buslaev, P.; Favela-Rosales, F.; Ferreira, T. M.; Fuchs, P. F. J.; Garcia-Fandino, R.; Gushchin, I.; Kav, B.; Kučerka, N.; Kula, P.; Kurki, M.; Kuzmin, A.; Lalitha, A.; Lolicato, F.; Madsen, J. J.; Miettinen, M. S.; Mingham, C.; Monticelli, L.; Nencini, R.; Nesterenko, A. M.; Piggot, T. J.; Piñeiro, Á.; Reuter, N.; Samantray, S.; Suárez-Lestón, F.; Talandashti, R.; Ollila, O. H. S. Overlay Databank Unlocks Data-Driven Analyses of Biomolecules for All. *Nat Commun* **2024**, *15* (1), 1136. https://doi.org/10.1038/s41467-024-45189-z.

(19) Lichtinger, S. M.; Biggin, P. C. Tackling Hysteresis in Conformational Sampling: How to Be Forgetful with MEMENTO. *J. Chem. Theory Comput.* **2023**, *19* (12), 3705–3720. https://doi.org/10.1021/acs.jctc.3c00140.

(20) Paloncýová, M.; DeVane, R. H.; Murch, B. P.; Berka, K.; Otyepka, M. Rationalization of Reduced Penetration of Drugs through Ceramide Gel Phase Membrane. *Langmuir* **2014**, *30* (46), 13942–13948. https://doi.org/10.1021/la503289v.

(21) Paloncýová, M.; Vávrová, K.; Sovová, Ž.; DeVane, R.; Otyepka, M.; Berka, K. Structural Changes in Ceramide Bilayers Rationalize Increased Permeation through Stratum Corneum Models with Shorter Acyl Tails. *J. Phys. Chem. B* **2015**, *119* (30), 9811–9819. https://doi.org/10.1021/acs.jpcb.5b05522.

(22) Lomize, A. L.; Pogozheva, I. D. Physics-Based Method for Modeling Passive Membrane Permeability and Translocation Pathways of Bioactive Molecules. *J. Chem. Inf. Model.* **2019**, *59* (7), 3198–3213. https://doi.org/10.1021/acs.jcim.9b00224.

(23) Schwöbel, J. A. H.; Ebert, A.; Bittermann, K.; Huniar, U.; Goss, K.-U.; Klamt, A. COSMO *Perm:* Mechanistic Prediction of Passive Membrane Permeability for Neutral Compounds and Ions and Its pH Dependence. *J. Phys. Chem. B* **2020**, *124* (16), 3343–3354. https://doi.org/10.1021/acs.jpcb.9b11728.

(24) Diamond, J. M.; Katz, Y. Interpretation of Nonelectrolyte Partition Coefficients between Dimyristoyl Lecithin and Water. *J. Membrain Biol.* **1974**, *17* (1), 121–154. https://doi.org/10.1007/BF01870176.

(25) Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. Anisotropic Solvent Model of the Lipid Bilayer. 2. Energetics of Insertion of Small Molecules, Peptides, and Proteins in Membranes. *J. Chem. Inf. Model.* **2011**, *51* (4), 930–946. https://doi.org/10.1021/ci200020k.

(26) Lomize, A. L.; Pogozheva, I. D.; Lomize, M. A.; Mosberg, H. I. Positioning of Proteins in Membranes: A Computational Approach. *Protein Science* **2006**, *15* (6), 1318–1333. https://doi.org/10.1110/ps.062126106.

(27) Sprous, D. G.; Palmer, R. K.; Swanson, J. T.; Lawless, M. QSAR in the Pharmaceutical Research Setting: QSAR Models for Broad, Large Problems. *CTMC* **2010**, *10* (6), 619–637. https://doi.org/10.2174/156802610791111506.

(28) Pham-The, H.; Cabrera-Pérez, M. Á.; Nam, N.-H.; Castillo-Garit, J. A.; Rasulev, B.; Le-Thi-Thu, H.; Casañola-Martin, G. M. In Silico Assessment of ADME Properties: Advances in Caco-2 Cell Monolayer Permeability Modeling. *CTMC* **2019**, *18* (26), 2209–2229. https://doi.org/10.2174/1568026619666181130140350.

(29) Egan, W. J.; Lauri, G. Prediction of Intestinal Permeability. *Advanced Drug Delivery Reviews* **2002**, *54* (3), 273–289. https://doi.org/10.1016/S0169-409X(02)00004-2.

(30) Vilar, S.; Sobarzo-Sanchez, E.; Santana, L.; Uriarte, E. Ligand and Structure-Based Modeling of Passive Diffusion through the Blood-Brain Barrier. *CMC* **2018**, *25* (9), 1073–1089. https://doi.org/10.2174/0929867324666171106163742.

(31) Barratt, M. D. Quantitative Structure-Activity Relationships for Skin Permeability. *Toxicology in Vitro* **1995**, *9* (1), 27–37. https://doi.org/10.1016/0887-2333(94)00190-6.

(32) Wang, N.-N.; Dong, J.; Deng, Y.-H.; Zhu, M.-F.; Wen, M.; Yao, Z.-J.; Lu, A.-P.; Wang, J.-B.; Cao, D.-S. ADME Properties Evaluation in Drug Discovery: Prediction of Caco-2 Cell Permeability Using a Combination of NSGA-II and Boosting. *J. Chem. Inf. Model.* **2016**, *56* (4), 763–773. https://doi.org/10.1021/acs.jcim.5b00642.

(33) Fredlund, L.; Winiwarter, S.; Hilgendorf, C. In Vitro Intrinsic Permeability: A Transporter-Independent Measure of Caco-2 Cell Permeability in Drug Design and Development. *Mol. Pharmaceutics* **2017**, *14* (5), 1601–1609. https://doi.org/10.1021/acs.molpharmaceut.6b01059.

(34) Sun, H.; Nguyen, K.; Kerns, E.; Yan, Z.; Yu, K. R.; Shah, P.; Jadhav, A.; Xu, X. Highly Predictive and Interpretable Models for PAMPA Permeability. *Bioorganic & Medicinal Chemistry* **2017**, *25* (3), 1266–1276. https://doi.org/10.1016/j.bmc.2016.12.049.

(35) Gousiadou, C.; Doganis, P.; Sarimveis, H. Development of Artificial Neural Network Models to Predict the PAMPA Effective Permeability of New, Orally Administered Drugs Active against the Coronavirus SARS-CoV-2. *Netw Model Anal Health Inform Bioinforma* **2023**, *12* (1), 16. https://doi.org/10.1007/s13721-023-00410-9.

(36) Holton, T. A.; Pollastri, G.; Shields, D. C.; Mooney, C. CPPpred: Prediction of Cell Penetrating Peptides. *Bioinformatics* **2013**, *29* (23), 3094–3096. https://doi.org/10.1093/bioinformatics/btt518.

(37) Williams-Noonan, B. J.; Speer, M. N.; Le, T. C.; Sadek, M. M.; Thompson, P. E.; Norton, R. S.; Yuriev, E.; Barlow, N.; Chalmers, D. K.; Yarovsky, I. Membrane Permeating Macrocycles: Design Guidelines from Machine Learning. *J. Chem. Inf. Model.* **2022**, *62* (19), 4605–4619. https://doi.org/10.1021/acs.jcim.2c00809.

(38) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2023 Update. *Nucleic Acids Research* **2023**, *51* (D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956.

(39) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Research* **2019**, *47* (D1), D930–D940. https://doi.org/10.1093/nar/gky1075.

(40) Juračka, J.; Šrejber, M.; Melíková, M.; Bazgier, V.; Berka, K. MolMeDB: Molecules on Membranes Database. *Database* **2019**, *2019*, baz078. https://doi.org/10.1093/database/baz078.

(41) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME - the Konstanz Information Miner: Version 2.0 and Beyond. *SIGKDD Explor. Newsl.* **2009**, *11* (1), 26–31. https://doi.org/10.1145/1656274.1656280.
(42) 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. *International Journal of Pharmaceutics* **2002**, *241* (2), 241–251. https://doi.org/10.1016/S0378-5173(02)00240-5.
(43) Berka, K.; Hendrychová, T.; Anzenbacher, P.; Otyepka, M. Membrane Position of Ibuprofen Agrees with Suggested Access Path Entrance to Cytochrome P450 2C9 Active Site. *J. Phys. Chem. A* **2011**, *115* (41), 11248–11255. https://doi.org/10.1021/jp204488j.
(44) Cramariuc, O.; Rog, T.; Javanainen, M.; Monticelli, L.; Polishchuk, A. V.; Vattulainen, I. Mechanism for Translocation of Fluoroquinolones across Lipid Membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012**, *1818* (11), 2563–2571. https://doi.org/10.1016/j.bbamem.2012.05.027.
(45) Lundborg, M.; Wennberg, C.; Lindahl, E.; Norlén, L. Simulating the Skin Permeation Process of Ionizable Molecules. *J. Chem. Inf. Model.* **2024**, *64* (13), 5295–5302. https://doi.org/10.1021/acs.jcim.4c00722.
(46) R Core Team. R: A Language and Environment for Statistical Computing, 2023. https://www.R-project.org/.
(47) Google. *Welcome to Colaboratory!* https://colab.research.google.com/gist/lzhou1110/2a30a81cb8c175514ed627bc18016774/hello-colaboratory.ipynb.
(48) Goble, C.; Soiland-Reyes, S.; Bacall, F.; Owen, S.; Williams, A.; Eguinoa, I.; Droesbeke, B.; Leo, S.; Pireddu, L.; Rodríguez-Navas, L.; Fernández, J. M.; Capella-Gutierrez, S.; Ménager, H.; Grüning, B.; Serrano-Solano, B.; Ewels, P.; Coppens, F. Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory. **2021**. https://doi.org/10.5281/ZENODO.4605654.
(49) Lex, A.; Gehlenborg, N.; Strobelt, H.; Vuillemot, R.; Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Visual. Comput. Graphics* **2014**, *20* (12), 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248.
(50) Barry, P. H.; Diamond, J. M. Effects of Unstirred Layers on Membrane Phenomena. *Physiological Reviews* **1984**, *64* (3), 763–872. https://doi.org/10.1152/physrev.1984.64.3.763.
(51) Palumbo, P.; Picchini, U.; Beck, B.; Van Gelder, J.; Delbar, N.; DeGaetano, A. A General Approach to the Apparent Permeability Index. *J Pharmacokinet Pharmacodyn* **2008**, *35* (2), 235–248. https://doi.org/10.1007/s10928-008-9086-4.
(52) Goswami, T.; Li, X.; Jasti, B. R. Effect of Lipophilicity and Drug Ionization on Permeation Across Porcine Sublingual Mucosa. *AAPS PharmSciTech* **2017**, *18* (1), 175–181. https://doi.org/10.1208/s12249-016-0479-1.
(53) Avdeef, A.; Artursson, P.; Neuhoff, S.; Lazorova, L.; Gråsjö, J.; Tavelin, S. Caco-2 Permeability of Weakly Basic Drugs Predicted with the Double-Sink PAMPA Method. *European Journal of Pharmaceutical Sciences* **2005**, *24* (4), 333–349. https://doi.org/10.1016/j.ejps.2004.11.011.
(54) Huque, F. T. T.; Box, K.; Platts, J. A.; Comer, J. Permeability through DOPC/Dodecane Membranes: Measurement and LFER Modelling. *European Journal of Pharmaceutical Sciences* **2004**, *23* (3), 223–232. https://doi.org/10.1016/j.ejps.2004.07.009.
(55) Avdeef, A.; Tsinman, O. PAMPA—A Drug Absorption in Vitro Model. *European Journal of Pharmaceutical Sciences* **2006**, *28* (1–2), 43–50. https://doi.org/10.1016/j.ejps.2005.12.008.
(56) Tsinman, O.; Tsinman, K.; Sun, N.; Avdeef, A. Physicochemical Selectivity of the BBB Microenvironment Governing Passive Diffusion—Matching with a Porcine Brain Lipid Extract Artificial Membrane Permeability Model. *Pharm Res* **2011**, *28* (2), 337–363. https://doi.org/10.1007/s11095-010-0280-x.
(57) Velický, M.; Tam, K. Y.; Dryfe, R. A. W. In Situ Artificial Membrane Permeation Assay under Hydrodynamic Control: Correlation between Drug in Vitro Permeability and Fraction Absorbed in Humans. *European Journal of Pharmaceutical Sciences* **2011**, *44* (3), 299–309. https://doi.org/10.1016/j.ejps.2011.08.007.
(58) Colabufo, N. A.; Berardi, F.; Cantore, M.; Perrone, M. G.; Contino, M.; Inglese, C.; Niso, M.; Perrone, R.; Azzariti, A.; Simone, G. M.; Porcelli, L.; Paradiso, A. Small P-Gp Modulating Molecules: SAR Studies on Tetrahydroisoquinoline Derivatives. *Bioorganic & Medicinal Chemistry* **2008**, *16* (1), 362–373. https://doi.org/10.1016/j.bmc.2007.09.039.
(59) Sun, H.; Chow, E. C.; Liu, S.; Du, Y.; Pang, K. S. The Caco-2 Cell Monolayer: Usefulness and Limitations. *Expert Opinion on Drug Metabolism & Toxicology* **2008**, *4* (4), 395–411. https://doi.org/10.1517/17425255.4.4.395.
(60) Volpe, D. A. Drug-Permeability and Transporter Assays in Caco-2 and Mdck Cell Lines. *Future Med. Chem.* **2011**, *3* (16), 2063–2077. https://doi.org/10.4155/fmc.11.149.
(61) Zhu, C.; Jiang, L.; Chen, T.-M.; Hwang, K.-K. A Comparative Study of Artificial Membrane Permeability Assay for High Throughput Profiling of Drug Absorption Potential. *European Journal of Medicinal Chemistry* **2002**, *37* (5), 399–407. https://doi.org/10.1016/S0223-5234(02)01360-0.
(62) Von Richter, O.; Glavinas, H.; Krajcsi, P.; Liehner, S.; Siewert, B.; Zech, K. A Novel Screening Strategy to Identify ABCB1 Substrates and Inhibitors. *Naunyn-Schmied Arch Pharmacol* **2009**, *379* (1), 11–26. https://doi.org/10.1007/s00210-008-0345-0.
(63) Kerns, E. H.; Di, L.; Petusky, S.; Farris, M.; Ley, R.; Jupp, P. Combined Application of Parallel Artificial Membrane Permeability Assay and Caco-2 Permeability Assays in Drug Discovery. *Journal of Pharmaceutical Sciences* **2004**, *93* (6), 1440–1453. https://doi.org/10.1002/jps.20075.
(64) Bittermann, K.; Spycher, S.; Endo, S.; Pohler, L.; Huniar, U.; Goss, K.-U.; Klamt, A. Prediction of Phospholipid–Water Partition Coefficients of Ionic Organic Chemicals Using the Mechanistic Model COSMO *Mic. J. Phys. Chem. B* **2014**, *118* (51), 14833–14842. https://doi.org/10.1021/jp509348a.

(65) Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and Computational Screening Models for the Prediction of Intestinal Drug Absorption. *J. Med. Chem.* **2001**, *44* (12), 1927–1937. https://doi.org/10.1021/jm001101a.

(66) Teksin, Z. S.; Seo, P. R.; Polli, J. E. Comparison of Drug Permeabilities and BCS Classification: Three Lipid-Component PAMPA System Method versus Caco-2 Monolayers. *AAPS J* **2010**, *12* (2), 238–241. https://doi.org/10.1208/s12248-010-9176-2.

(67) Fujikawa, M.; Nakao, K.; Shimizu, R.; Akamatsu, M. QSAR Study on Permeability of Hydrophobic Compounds with Artificial Membranes. *Bioorganic & Medicinal Chemistry* **2007**, *15* (11), 3756–3767. https://doi.org/10.1016/j.bmc.2007.03.040.

(68) Avdeef, A.; Nielsen, P. E.; Tsinman, O. PAMPA—a Drug Absorption in Vitro Model: 11. Matching the in Vivo Unstirred Water Layer Thickness by Individual-Well Stirring in Microtitre Plates. *European Journal of Pharmaceutical Sciences* **2004**, *22* (5), 365–374. https://doi.org/10.1016/j.ejps.2004.04.009.

(69) Brunner, J.; Graham, D. E.; Hauser, H.; Semenza, G. Ion and Sugar Permeabilities of Lecithin Bilayers: Comparison of Curved and Planar Bilayers. *J. Membrain Biol.* **1980**, *57* (2), 133–141. https://doi.org/10.1007/BF01868999.

(70) Helander, H. F.; Fändriks, L. Surface Area of the Digestive Tract – Revisited. *Scandinavian Journal of Gastroenterology* **2014**, *49* (6), 681–689. https://doi.org/10.3109/00365521.2014.898326.

(71) Schiller, C.; Frohlich, C.-P.; Giessmann, T.; Siegmund, W.; Monnikes, H.; Hosten, N.; Weitschies, W. Intestinal Fluid Volumes and Transit of Dosage Forms as Assessed by Magnetic Resonance Imaging. *Aliment Pharmacol Ther* **2005**, *22* (10), 971–979. https://doi.org/10.1111/j.1365-2036.2005.02683.x.

(72) Hubatsch, I.; Ragnarsson, E. G. E.; Artursson, P. Determination of Drug Permeability and Prediction of Drug Absorption in Caco-2 Monolayers. *Nat Protoc* **2007**, *2* (9), 2111–2119. https://doi.org/10.1038/nprot.2007.303.

(73) Deur, C.; Agrawal, A. K.; Baum, H.; Booth, J.; Bove, S.; Brieland, J.; Bunker, A.; Connolly, C.; Cornicelli, J.; Dumin, J.; Finzel, B.; Gan, X.; Guppy, S.; Kamilar, G.; Kilgore, K.; Lee, P.; Loi, C.-M.; Lou, Z.; Morris, M.; Philippe, L.; Przybranowski, S.; Riley, F.; Samas, B.; Sanchez, B.; Tecle, H.; Wang, Z.; Welch, K.; Wilson, M.; Yates, K. N-(6,7-Dichloro-2,3-Dioxo-1,2,3,4-Tetrahydroquinoxalin-5-Yl)-N-Alkylsulfonamides as Peripherally Restricted N-Methyl-d-Aspartate Receptor Antagonists for the Treatment of Pain. *Bioorganic & Medicinal Chemistry Letters* **2007**, *17* (16), 4599–4603. https://doi.org/10.1016/j.bmcl.2007.05.083.

(74) Chiba, J.; Iimura, S.; Yoneda, Y.; Watanabe, T.; Muro, F.; Tsubokawa, M.; Iigou, Y.; Satoh, A.; Takayama, G.; Yokoyama, M.; Takashi, T.; Nakayama, A.; Machinaga, N. Synthesis and Biological Evaluation of Benzoic Acid Derivatives as Potent, Orally Active VLA-4 Antagonists. *Bioorganic & Medicinal Chemistry* **2007**, *15* (4), 1679–1693. https://doi.org/10.1016/j.bmc.2006.12.006.

(75) Flaten, G. E.; Kottra, G.; Stensen, W.; Isaksen, G.; Karstad, R.; Svendsen, J. S.; Daniel, H.; Svenson, J. In Vitro Characterization of Human Peptide Transporter hPEPT1 Interactions and Passive Permeation Studies of Short Cationic Antimicrobial Peptides. *J. Med. Chem.* **2011**, *54* (7), 2422–2432. https://doi.org/10.1021/jm1015704.

(76) Holz, M.; Heil, S. R.; Sacco, A. Temperature-Dependent Self-Diffusion Coefficients of Water and Six Selected Molecular Liquids for Calibration in Accurate 1H NMR PFG Measurements. *Phys. Chem. Chem. Phys.* **2000**, *2* (20), 4740–4742. https://doi.org/10.1039/b005319h.

(77) Derrick, T. S.; McCord, E. F.; Larive, C. K. Analysis of Protein/Ligand Interactions with NMR Diffusion Measurements: The Importance of Eliminating the Protein Background. *Journal of Magnetic Resonance* **2002**, *155* (2), 217–225. https://doi.org/10.1006/jmre.2002.2513.

(78) Amidon, G. L.; Lennernäs, H.; Shah, V. P.; Crison, J. R. A Theoretical Basis for a Biopharmaceutic Drug Classification: The Correlation of in Vitro Drug Product Dissolution and in Vivo Bioavailability. *Pharmaceutical Research* **1995**, *12* (3), 413–420. https://doi.org/10.1023/A:1016212804288.