DISSERTATIONES
CHIMICAE
UNIVERSITATIS
TARTUENSIS
186

MARE OJA

Experimental investigation and modelling of pH profiles for effective membrane permeability of drug substances

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
186

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
186

MARE OJA

Experimental investigation and modelling of pH profiles for effective membrane permeability of drug substances

UNIVERSITY OF TARTU
Press

Institute of Chemistry, Faculty of Science and Technology, University of Tartu, Estonia

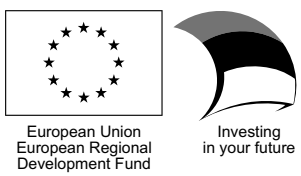
The dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Molecular Engineering on June 11th, 2019, by the Doctoral Committee of the Institute of Chemistry, University of Tartu.

Supervisor: Senior Research Fellow Uko Maran, PhD
Institute of Chemistry, University of Tartu, Estonia

Opponent: Prof. Dr. Johann Gasteiger
Computer Chemistry Center and
Institute of Organic Chemistry,
Department of Chemistry and Pharmacy,
Friedrich Alexander University Erlangen-Nürnberg,
Germany

Commencement: August 27th, 2019 at 10:15, Ravila 14a, Tartu, room 1020

This work has been partially supported by ASTRA project PER ASPERA Graduate School of Functional Materials and Technologies receiving funding from the European Regional Development Fund under project in University of Tartu, Estonia.



ISSN 1406-0299
ISBN 978-9949-03-098-9 (print)
ISBN 978-9949-03-099-6 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following five original papers, which will be denoted by the corresponding Roman numerals in the text:

I. **Oja, M.**; Maran, U. The permeability of an artificial membrane for wide range of pH in human gastrointestinal tract: experimental measurements and quantitative structure-activity relationship. *Mol. Inf.* **2015**, *34*, 493–506.

II. **Oja, M.**; Maran, U. Quantitative structure-permeability relationships at various pH values for acidic and basic drugs and drug-like compounds. *SAR QSAR Environ. Res.* **2015**, *26*, 701–719.

III. **Oja, M.**; Maran, U. Quantitative structure-permeability relationships at various pH values for neutral and amphoteric drugs and drug-like compounds. *SAR QSAR Environ. Res.* **2016**, *27*, 813–832.

IV. **Oja, M.**; Maran, U. pH-permeability profiles for drug substances: experimental detection, comparison with human intestinal absorption and modelling. *Eur. J. Pharm. Sci.* **2018**, *123*, 429–440.

V. **Oja, M.**; Sild, S.; Maran, U. Logistic classification models for pH-permeability profile: predicting permeability classes for the biopharmaceutical classification system. *J. Chem. Inf. Model.* **2019**, *59*, 2442–2455.

Author's contribution:

Paper I: The author was responsible for planning and performing experimental measurements, preparing data for modelling, developing prediction models, analysing results, and writing the manuscript.

Paper II: The author was responsible for planning and performing experimental measurements, preparing data for modelling, developing prediction models, analysing results, and writing the manuscript.

Paper III: The author was responsible for planning and performing experimental measurements, preparing data for modelling, developing prediction models, analysing results, and writing the manuscript.

Paper IV: The author was responsible for planning and performing experimental measurements, preparing data for modelling, developing prediction models, analysing results, and writing the manuscript.

Paper V: The author was responsible for preparing data for modelling, developing classification models, analysing results, and writing the manuscript.

LIST OF ABBREVIATIONS

%HIA	Human intestinal absorption (%)
ABL	Aqueous boundary layer
BCS	Biopharmaceutical classification system
BMLR	Best multiple linear regression
f_u	Fraction of unionised species
GIT	Gastrointestinal tract
log$D$	Logarithm of the octanol-water distribution coefficient
log$D_{highest}$	Highest logarithm of the octanol-water distribution coefficient over pH-s 3, 5, 7.4, and 9
logD_{pH3}	Logarithm of the octanol-water distribution coefficient at pH 3
logD_{pH5}	Logarithm of the octanol-water distribution coefficient at pH 5
log$D_{pH7.4}$	Logarithm of the octanol-water distribution coefficient at pH 7.4
logD_{pH9}	Logarithm of the octanol-water distribution coefficient at pH 9
logP_e	Logarithm of the effective membrane permeability
log$P_{e_highest}$	Highest logarithm of the effective membrane permeability over pH-s 3, 5, 7.4, and 9
logP_{e_pH3}	Logarithm of the effective membrane permeability at pH 3
logP_{e_pH5}	Logarithm of the effective membrane permeability at pH 5
log$P_{e_pH7.4}$	Logarithm of the effective membrane permeability at pH 7.4
logP_{e_pH9}	Logarithm of the effective membrane permeability at pH 9
logP_o	Logarithm of the intrinsic membrane permeability
logP_{ow}	Logarithm of the octanol-water partition coefficient
n	Number of compounds
n_{train}	Number of compounds in the training set
n_{val}	Number of compounds in the validation set
n_{ext}	Number of compounds in the external validation set
PAMPA	Parallel artificial membrane permeability assay
pK_a	Ionisation constant
QSAR	Quantitative structure-activity relationship
R^2_{cv}	Cross-validated coefficient of determination
R^2_{ext}	Coefficient of determination for the external validation set
R^2_{train}	Coefficient of determination for the training set
R^2_{val}	Coefficient of determination for the validation set
s^2	Squared standard error of the estimate
U.S. FDA	U.S. Food and Drug Administration
UV/Vis	Ultraviolet-visible
V_A	Volume of solution in the acceptor section
V_D	Volume of solution in the donor section

INTRODUCTION

Absorption, distribution, metabolism, and excretion (ADME) determine the pharmacokinetic behaviour of drug substances in the human body, in other words, what the human body does with the drug substance. Therefore, unsuitable pharmacokinetic properties are one of the most frequent reasons for drug attrition in the preclinical phase and in the clinical phase I trials [1,2,3,4], and to a lesser extent in the clinical phase II trials [3,4,5]. This suggests that the pharmacokinetic properties must be considered as early as possible and as effectively as possible in the drug discovery process in order to reduce the attrition of drug substance candidates.

Oral administration is a preferable and common way to deliver drugs as tablets, capsules, and solutions to the organism. In the case of orally administrated drugs, absorption in the gastrointestinal tract (GIT) is a significant feature of the ADME profile. Absorption depends on the properties of the GIT and drug substances. One of the significant absorption-related properties for drug substances is permeability through the intestinal epithelium. It is influenced by the various properties in the GIT, and one of such is pH, that is controlled by fasted and fed state [6]. The pH in the GIT fluctuates from very acidic (pH ~2) to basic (pH ~8) [6,7,8,9]. Despite of this the influence of pH to the permeability of drug substance candidates in the early stages of drug discovery is rarely considered.

The permeability can be experimentally studied using *in vivo*, *in situ*, and *in vitro* methods [10]. Since *in vivo* and *in situ* methods are expensive and time-consuming, *in vitro* methods have been widely used to describe permeability in the GIT. *In vitro* methods for detecting permeability are cell-based [11] or cell-free methods [12], such as an artificial membrane [13]. The advantages of the artificial membrane in comparison with the cell-based methods are simplicity, cheapness, and the ability to change easily experimental parameters, like pH.

Transferring the knowledge from the influence of pH in the GIT to the cheminformatics models allows to improve the early phase of drug discovery and development and reduces the attrition of new drug substance candidates. In particular, cheminformatics prediction models can be applied in the early stages of drug discovery to select out the most potential drug substance candidates for oral administration. So far, prediction models for permeability have been usually developed only at neutral or near to neutral pH-s and the wider pH range is not considered [14]. This is partially caused by the fact that the systematic permeability data for the wide pH range are missing in the literature.

Therefore, the aim of the study in this thesis is to involve the GIT pH range in the prediction models of the permeability and thus to improve the application of the models for assessing the permeability in the gastrointestinal tract. This goal is achieved through the systematic experimental permeability measurements and cheminformatics modelling of measured data.

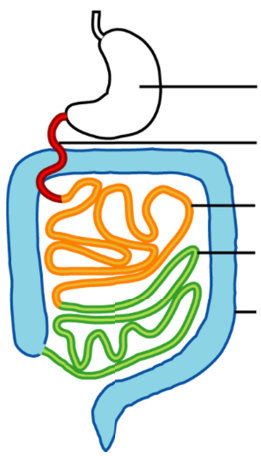
1. LITERATURE OVERVIEW

1.1. Human intestinal absorption

Oral administration is the preferable, cost effective and convenient way to deliver drugs. This makes absorption a key feature in analysing the properties of drug substances (active pharmaceutical ingredient, API) and their candidates in the GIT. Absorption depends on two sets of properties, properties of the GIT and properties of the drug substance.

1.1.1. Properties of the gastrointestinal tract

The GIT (Figure 1) consists of the stomach, the small intestine, which is divided into the duodenum, jejunum, and ileum, and the colon (large intestine). The environment of GIT sections is variable and diverse. Variable parameters in the GIT are mainly pH, surface area, and various compounds and their concentration [6,7,8,9,15,16,17,18,19]. For example the pH range in the GIT (Figure 1) is remarkably wide, changing from acidic (pH ~2–3) to basic (pH ~8–9) [6,7,8,9,15,17,18,19], and is influenced by the fasted and fed state [6,15,17,18,19]. Usually in the upper sections of the GIT (stomach and duodenum) the pH is more acidic in the fasted state than in the fed state (Figure 1). In the bottom sections of the GIT (jejunum and ileum) the pH ranges in the fed and fasted states are similar (Figure 1). The pH is fluctuating in the sections of the GIT and it is highly variable over the test subjects (volunteers) [15,17,19] and thus it is difficult to determine the pH value in each section.



	Area (m^2)	Length (m)	Residence time (h)	Fed pH	Fasted pH
Stomach	3.5	0.25	1.5	5.0 (0.1 h)	1.5–2.0
Duodenum	1.9	0.35	0.5–0.75	4.0–6.5	2.4–6.8
Jejunum	184	2.8	1.5–2.0	4.4–6.6	5.2–6.2
Ileum	276	4.2	5–7	6.8–8.0	6.8–8.4
Colon	1.3	1.5	1–60		5.0–8.0

Figure 1. Properties of the GIT: surface area, length, residence time, and pH for fasted and fed state in the different sections of the GIT [6,7,8,9].

The surface area of the GIT and the length of the section are the largest in the jejunum and ileum (Figure 1). The large surface area of the small intestine (duo-

denum, jejunum, ileum) is caused by the relatively long section length and surface projections (circular folds, villi and microvilli) [8,9,16]. Such intestinal elements like in the small intestine are absent in the colon [8]. According to the literature around 99% of absorption occurs in the jejunum and ileum [7]. Recently, it has also been shown that absorption occurs even in the colon [20], which may be caused by the longest residence time over the intestinal sections (Figure 1).

Other parameters affecting the absorption in the GIT are buffer capacity, osmolarity, surface tension, and the concentrations of bile salts and phospholipids [6,17,18,19]. All these parameters significantly depend on the section of GIT [17,19] and therefore also influence the absorption in the GIT [21,22,23].

It has been concluded that the maximum absorption of drug substance should take place during 3–5 hours in the jejunum and ileum and in the range of pH from 4.5 to 8 [7]. From the literature, it is also known that the changes of pH, area, and other parameters make the absorption of drug substances in the GIT regional and time dependent [20,24].

1.1.2. Properties of the drug substance

The absorption of drug substances is not influenced only by the GIT characteristics, but also by the properties of the drug substance. Therefore, the absorption is highly influenced by the solubility and permeability of drug substances [25,26,27], which in turn is determined by the ionisation of drug substances [28].

Solubility and permeability form the basis of the biopharmaceutical classification system (BCS) [25,27] and are used to classify drug substances into four classes (Table 1). The United States Food and Drug Administration (U.S. FDA) recommends the BCS for selecting suitable biowaivers for *in vivo* bioavailability and/or bioequivalence studies [10]. The U.S. FDA guideline [10] indicates that drug substance is a high soluble, when the maximum dose of drug substance is soluble in 250 ml or less of aqueous media within the range of pH 1–6.8 at 37 °C. The permeability is classified based on bioavailability or absorption in the human GIT (i.e., human intestinal absorption, %HIA), where the cutoff for high permeability is 85% of the administrated dose according to the mass balance determination [10,27].

Table 1. The biopharmaceutical classification system classifies drug substances and their candidates into four groups.

Solubility / **Permeability**	**High solubility**	**Low solubility**
High permeability	Class I	Class II
Low permeability	Class III	Class IV

According to the U.S. FDA guideline [10] the pH is considered in the estimation of solubility up to a certain extent, but not at all for the estimation of permeability. This is despite the high degree of dependence of solubility and

permeability on the fraction of unionised species, i.e. pH-partition hypothesis [29] and solubility-pH equations [7,30]. Unionised compounds have usually higher solubility in the water, and for higher permeability the compound must be unionised. Therefore, it is important to analyse the interplay of permeability and solubility in a wider pH range [31,32]. Solubility-permeability interplay over the pH range in the GIT can be used to determine the suitable pH range for the absorption [31], which in turn can be connected to the sections of the GIT. Solubility is analysed for different pH-s [28,33] and even with simulated intestinal fluids for fasted and fed state [34,35,36]. Considering the solubility dependence of the pH for different chemical classes, it has been proposed that the BCS should include sub-classes of compounds with low solubility [33]. As shown, the effect of pH on solubility has been extensively investigated, while the effect of pH on permeability has been significantly less studied [7,37].

1.2. Permeability in the gastrointestinal tract

Drug substances must cross the intestinal epithelium. The intestinal epithelium consists of cells [9,38], which are tightly attached and form a thin hydrophobic barrier for chemicals with a thickness of about 5 nm, known as a cellular membrane. The cellular membrane is comprised of various amphiphilic phospholipids, cholesterol, and membrane anchored proteins, such as transporters [38]. The intestinal epithelium is additionally coated with a mucous [8], which contains mucin glycoproteins, enzymes, and electrolytes, and they form an additional barrier for the permeability in the GIT. Drug substances can cross the intestinal epithelium using different transport routes.

1.2.1. Transport routes

Drug substances can cross the cellular membrane using passive and/or carrier-mediated transport (Figure 2) [38,39,40]. It has been estimated that around 90% of drug substances are transported passively [41].

Passive transport [38] is the diffusion of drug substances in the cellular membrane from a high concentration region to a low concentration region. Passive transport is not usually saturable, it is not subject to inhibition, and is less sensitive to the stereospecific structure of drug substance [38]. The passive transport is divided into two groups: paracellular and transcellular transport (Figure 2).

Passive paracellular transport refers to the passage of molecules between adjacent cells (Figure 2) [38]. It is most significant in the upper small intestine, which is leakier in comparison with the tighter barriers, like the colon [38]. The paracellular transport route is usually used by the small molecules [7,38,42], hydrophilic compounds [38,43], and small cations [42,44,45].

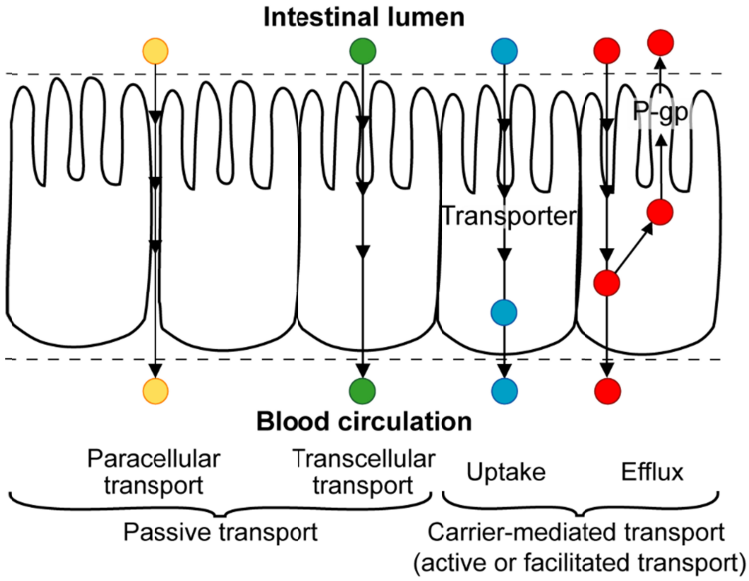


Figure 2. Transport routes in the GIT [39].

Passive transcellular transport [38] is defined as the movement of compounds due to the concentration gradient between two sides of the cellular membranes (Figure 2). Most preferred compounds for the passive transcellular transport are hydrophobic, unionised, and largely desolvated [38]. In addition, this transport route depends on the size of the molecule and on the lipophilicity of the lipid bilayer centre [38]. Passive transcellular transport is highly influenced by the fraction of unionised species based on the pH-partition hypothesis [29], i.e. only unionised species can be moved across the cellular membrane. However, there are also deviations from this hypothesis; recent analyses show that some negatively [46] and positively charged compounds [47], like ion-pairs and naked ions, and permanently charged compounds [48], absorb through the cellular membrane.

Carrier-mediated transport (Figure 2) is a transcellular permeation involving a protein (transporter) that is stereospecific and enantioselective, has a limited capacity, is saturable, and subject to the inhibition [38]. It is usually active transport, which is energy dependent, but it can be also facilitated transport, which is not energy dependent and relies on the concentration gradient of a substrate and transporter protein (Figure 2) [38]. Carrier-mediated transport includes uptake, which moves molecules to the cell, or efflux (like P-gp mediated efflux, Figure 2), which moves molecules out of the cells [38]. In the human body, around 400 different transport proteins have been found including around 10 transporters, which have practical consideration or clinical evidence in the intestinal epithelium [49]. Carrier-mediated transport is important for compounds with low passive permeation, especially for the distribution in the liver and in the brain [38].

1.2.2. Measurement methods

Various *in vivo* (human and animal models), *in situ* (animal models), and *in vitro* (tissue and cell) experimental methods have been developed to describe permeability in the GIT [10].

Human- and animal-based methods are divided between *in vivo* and *in situ* methods. *In vivo* methods include human pharmacokinetic studies, such as mass balance and absolute bioavailability [10,50], and intestinal perfusion in the human or animal subjects [51]. *In situ* methods such as intestinal perfusion study using animal model(s) have also been used to simulate the *in vivo* situation in the GIT [52,53,54]. *In situ* methods are similar to the *in vivo* methods, although in this case animals are anaesthetised, which is also the disadvantage of the method, because it is not well known how anaesthesia will affect absorption process in the GIT [12]. The major concern of *in vivo* and *in situ* methods is ethics, but these methods are also time consuming and costly [55].

In vitro methods have been proposed as alternatives to *in vivo* and *in situ* methods. *In vitro* permeability methods [10] include excised intestinal tissues or monolayer of suitable epithelial cells. Tissue based methods [56,57] mimic closely the *in vivo* situation from an anatomical, biological, and structural point of view, but can be highly influenced by the quality of the tissue (such as an irremovable circular muscle layer) [12]. *In vitro* methods with a monolayer of epithelial cells are considered as valuable alternatives to the human, animal, and tissue-based methods to assess intestinal drug substance permeation. Most used cell lines are the human colorectal carcinoma (Caco-2) [58,59,60] and the Madin-Darby canine kidney (MDCK) [61] cells, which both show good correlation with the permeability in the GIT [62,63]. The cell-based methods have also several drawbacks [12,64,65], such as incompatibility with food components and certain pharmaceutical excipients, the absence of transporters, the lack of mucus layer, time-consuming, expensive, and the final result depends on the cell line.

In the frame of the U.S. FDA guideline [10], *in vivo* methods in the human subjects are the only recommended methods, which can be used to determine both carrier-mediated and passive transport. Animal-based and *in vitro* methods are only recommended for the drug substances that are transported by passive mechanisms.

Considering the drawbacks of the previously mentioned methods, alternatives have been sought to describe passive transport and therefore different cell-free methods have been developed [12]. Cell-free methods include a parallel artificial membrane permeability assay (PAMPA) [13,66,67,68], a phospholipid vesicle-based permeation assay (PVPA) [69], a Permeapad® [70,71], and an artificial membrane insert system (AMI-system) [72]. The oldest and most utilized cell-free based method is the PAMPA, because it is robust and fairly easy to modify and offers the capability of analysing various aspects of absorption, like a wide range of pH.

1.3. Artificial membrane permeability

The artificial membrane builds on the concept that the lipid bilayer is the fundamental structure of the cellular membranes. This concept was used first to form a black lipid membrane (BLM) [73], which was very fragile. To improve the membrane stability, different filter materials [74,75,76] have been introduced to stabilize the lipid membranes.

In 1998, Kansy et al [13] published the first high-throughput and resource-effective artificial membrane method to determine passive transport in the GIT using phospholipid-coated filters, which is called as the PAMPA. Since then the PAMPA has been gaining popularity in estimating the absorption properties of molecules, particularly drug substance candidates. The main advantages of the artificial membrane methods compared to the cell-based methods are readiness for high-throughput, fast membrane preparation, quick concentration detection methods, and tolerance to a wider pH range and higher co-solvent content [66].

The PAMPA system consists of three parts (Figure 3): a donor section (filter plate) that describes the GIT, a membrane solution on a filter that mimic the intestinal epithelium, and an acceptor section (acceptor plate) that describes the blood circulation [13]. Depending on the protocol, the donor and acceptor sections may be interchangeable.

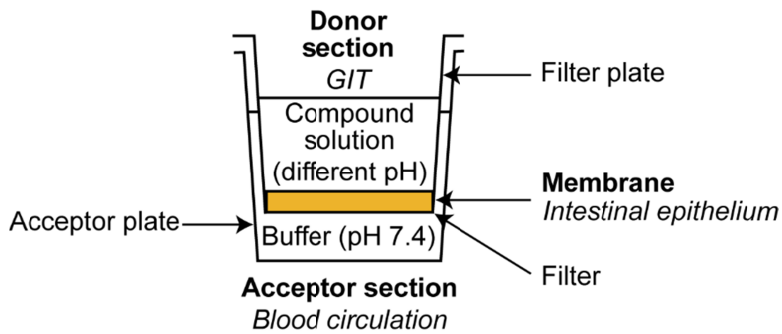


Figure 3. Schematic representation of the PAMPA experimental system: one well in 96-well plate.

1.3.1. Membrane types

The most important part of the PAMPA experimental system is the membrane used to cover the filter. The filter is usually hydrophobic PVDF (polyvinylidene fluoride) [13], although hydrophilic PVDF [66] and polycarbonate filters [77] have been also used (Table 2).

Table 2. The PAMPA experimental systems for describing different permeability barriers in the human body and respective composition of membranes with filter material.

Method	Barrier	Membrane composition	Filter	Ref.
PAMPA-EGG	GIT	1–20% egg lecithin in dodecane	Hydrophobic PVDF	[13]
PAMPA-EGG	GIT	1% egg lecithin in dodecane	Hydrophilic PVDF	[66]
PAMPA-DOPC	GIT	2% dioleoylphosphatidylcholine in dodecane	PVDF	[78]
PAMPA-HDM	GIT	Hexadecane in hexane	Polycarbonate	[77]
PAMPA-BM (bio-mimetic)	GIT	0.8% L-α-phosphatidylcholine, 0.8% L-α-phosphatidylethanolamine, 0.2% L-α-phosphatidylserine, 0.2% L-α-phosphatidylinositol, and 1% cholesterol in 1,7-octadien	Hydrophobic PVDF	[79]
PAMPA-DS (double-sink)	GIT	20% lecithin mixture in dodecane	Hydrophobic PVDF	[80]
Tri-layer PAMPA	GIT	Lipid/oil/lipid tri-layer in hexane	PVDF	[81]
PAMPA-BBB	Blood-brain	20 mg/ml porcine polar brain lipid in dodecane	Hydrophobic PVDF	[82]
PAMPA-Skin	Skin	30% isopropyl myristate and 70% silicone oil	Hydrophobic PVDF	[83]
Skin-PAMPA	Skin	Certramide, free fatty acid, and cholesterol		[84]
Corneal-PAMPA	Cornea	10.7% phosphatidylcholine in a mixture of hexane:dodecane:chloroform		[85]

The membrane composition is one of the most frequently modified parts of the PAMPA system (Table 2). The membrane typically contains two components: phospholipid(s) and organic solvent. The first PAMPA membrane [13] contained lecithin (L-α-phosphatidylcholine) in dodecane, but more sophisticated membrane systems have also been developed, like a double-sink (PAMPA-DS) [80] and bio-mimetic (PAMPA-BM) PAMPA [79]. Complex membrane systems mean more membrane components, added to improve the description of the intestinal epithelium [66]. Additional components however can cause problems with the repeatability of the experiment and make the tests significantly more expensive [86]. In addition to the complexity, the concentrations of the membrane components also play an important role. For example, the concentration of lecithin in the membrane is increased because of the membrane permeability of the polar compounds is reduced at a lower concentration [87]. This is because the polar compounds form strong hydrogen bonds with water and prevent compounds from accessing the membrane. More hydrogen bonding sites in the membrane compensate the energy of breaking strong hydrogen bonds between water and compound, which helps movement of compound into the membrane. The concentration of lecithin in the membrane is increased despite of the fact that high concentration of phospholipids causes accumulation of the compound into the membrane [7].

The compositions of membrane phospholipids do have various membrane fluidity and apparent ion pair effects that influence membrane permeability of drug substances [88]. For example, the PAMPA-DS [87] have been used as the reference system to derive experimental values for the PAMPA-HDM and the PAMPA-DOPC. The correlation between the PAMPA-DS and the PAMPA-HDM/DOPC is usually very good (determination coefficient R^2 around 0.8) and this can be improved by using Abraham descriptors [89], which describe hydrogen bond acceptor and donor properties. This demonstrates that the main difference between membrane compositions is the ability to account for hydrogen bond properties.

The membrane solvent, which is used to dissolve phospholipid(s), has also an important role in the membrane permeability [86]. The most used membrane solvent is dodecane, but also 1,7-octadien has been used (Table 2). To avoid solvent effect, volatile organic solvents, like hexane, has been used for solving membrane components in the PAMPA-HDM and tri-layer PAMPA (Table 2) systems. The solvent is evaporated before the PAMPA experiment and solvent-free membrane is formed.

The PAMPA system is not limited for describing the GIT. Other membrane compositions have been developed to study other barriers in the human body (Table 2). For example, porcine brain lipid in dodecane has been used for describing the blood-brain barrier [82]. More recent modifications of the PAMPA methods are designed for the skin [83,84] and for the cornea [85]. These examples confirm that the PAMPA method has been widely accepted for imitating barriers in the human body and many of which have been widely used in drug discovery.

1.3.2. Section composition

Next to the membrane composition, the composition of acceptor and donor sections can be changed in the PAMPA system [7,68]. These are: pH in the donor section, co-solvent in both sections, etc. The possibility to easily modify different parameters in the PAMPA system provides an opportunity to simulate the different conditions of the GIT.

In the typical experimental setting, the acceptor and the donor sections have solutions with the same pH (pH = 7.4, iso-pH experiment). Using only pH 7.4 in the donor section, will not give reliable membrane permeability information for all drug substances [65], because membrane permeability for about 80% of drug substances is known to be affected by the pH [90]. The human GIT-like PAMPA system (gradient-pH experiment) has the solution with a different pH in the donor section and the pH of the solution in the acceptor section is 7.4 to mimic blood circulation. Surprisingly, the effect of pH to the membrane permeability is rarely considered during the characterization and even less for the modelling of drug substance candidates [66,79,91]. A good example that a wider range of pH values needs to be considered is provided by Velicky et al [92], who measured membrane permeability values in the pH range from 3.5 to

10 for two high absorbed compounds, *Warfarin* and *Verapamil*, which both demonstrate a significant pH-dependence of permeability, represented by the pH-permeability profile.

Partially insoluble compounds can strongly affect membrane permeability values [93] and therefore, different co-solvents have been analysed and used [79,93,94]. Co-solvents should not influence the membrane and the membrane permeability but may influence ionisation constant (pK_a) of drug substances [95]. For example, acetonitrile as a co-solvent has been shown to affect less acidic compounds than basic compounds [95]. The effects of the co-solvent and chemical class have been taken account in correlation analysis between different PAMPA systems using hydrogen bond acidity and basicity parameters defined by Abraham [95]. The most used co-solvents are dimethyl sulfoxide (DMSO), ethanol, and polyethylenglycole (PEG400) [79]. Also, in some cases, when higher concentrations of lecithin has been used (like 20% of lecithin in dodecane), additional compounds are added to the acceptor section to reduce membrane retention [7,96]. Suitable compounds that reduce membrane retention are surfactants, such as sodium lauryl sulfate (SLS) [96], which effectively binds to the drug substances and therefore reduces membrane retention.

1.3.3. Experimental conditions and detection

After adding solutions to the acceptor and donor sections, the filter plate is placed on the acceptor plate and at this point the experiment will start. The incubation time varies and depends on the experimental system and membrane permeability of the drug substance [7,97]. High-permeable compounds need shorter experimental time, because concentration in the donor and acceptor sections should be detected before equilibrium [78]. At the same time, low-permeable compounds need longer experimental time to achieve the concentration in the acceptor section over the limit of detection. Also, the membrane in the PAMPA system has high influence on the incubation time. For example, PAMPA systems with higher lecithin concentration will need shorter experimental time, but then the limiting factor for the membrane permeability is the movement of drug substances to the membrane. In this case, it is important to stir the system [78,98], so that the movement of drug substances near to the membrane is facilitated. For the stirring, the easiest option is the orbital shaker, but it is not influencing homogeneously all wells and stirring depends on the position of the PAMPA system [7,78]. To avoid unequal stirring, the individual-well magnetic stirring in the PAMPA system have been developed, but this system needs specific plates and equipment [98]. With stirring, the test period for high-permeable compounds can be around 30 minutes [7,97].

The concentrations in the donor and/or acceptor sections can be detected with the liquid chromatography-mass spectrometry (LC/MS) or ultraviolet-visible (UV/Vis) spectrometry equipped with microplate reader [78,93]. The advantages of the UV/Vis microplate reader are short measurement time and increased throughput. The problems may occur in detecting the concentration in

the acceptor section for low-permeable compounds [78]. If the concentration in the acceptor section is high, then LC/MS and UV/Vis spectrometry will give comparable results [78]. The limitation of the UV/Vis spectrometry is the detection of the compounds with low extension coefficient and without chromophores. For these compounds, LC/MS can be used, which is also applicable to the mixtures, detection of low concentration in the acceptor section, and to control the stability of the drug substances during the PAMPA experiment [78].

1.3.4. Membrane permeability as process

The membrane permeability process (Figure 4) can be described using characteristic parameters for three sections of the PAMPA system (donor section, membrane, acceptor section). The donor section is described with the initial concentration in the donor section ($C_D(0)$, mol/l), the volume of the solution (V_D, ml), and the concentration in the donor section at a certain time point ($C_D(t)$, mol/l). The movement of the compounds to the membrane is described with the membrane/water apparent partition coefficient ($D_{m/v}$), which is also related to the aqueous boundary layer (ABL). The concentration of the compound in the donor/membrane boundary (C_m^0, mol/l) is decreasing linearly (dC_m/dx) over the thickness of the membrane (h) until arriving to the membrane/acceptor boundary, where the concentration of the compound (C_m^h, mol/l) is significantly lower. This process is described with the diffusivity of the solute in the membrane (D_m, cm^2/s) and the lag time ($\tau_{LAG,}$ s), which is the time when the compound first appears in the acceptor section. The membrane permeability is also influenced by the area of the filter (A, cm^2) and the apparent porosity of the filter (ε_a). The movement of compounds from the membrane/acceptor boundary to the acceptor section is described with the membrane/water apparent partition coefficient ($D_{m/v}$), which is the same for both sides of the membrane when the iso-pH experiment is carried out, but in the case of the gradient-pH experiment, these two values may not be the same. The third part, the acceptor section is described with the initial concentration in the acceptor section ($C_A(0)$, mol/l), the volume of solution (V_A, ml) and the concentration in the acceptor section at a certain time point ($C_A(t)$, mol/l). [7]

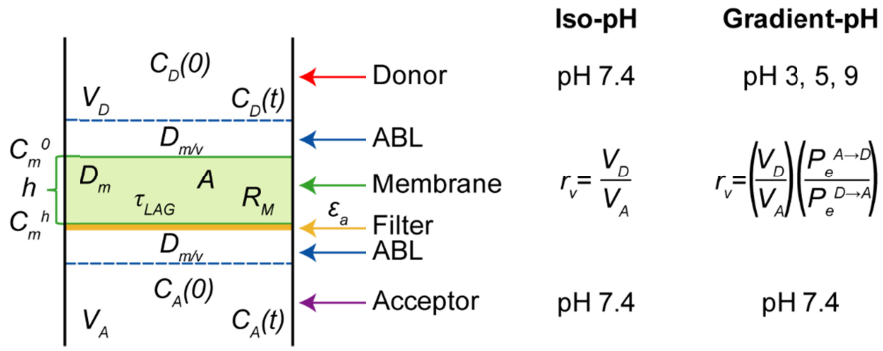


Figure 4. Description and characteristics of the membrane permeability process.

Membrane permeability process is described with the flux (J, mol/(cm^2·s), Eq. (1)), which is derived from the Fick's first law [78].

$$J = D_m \cdot \frac{\mathrm{d}C_m}{\mathrm{d}x} = D_m \cdot \frac{C_m^0 - C_m^h}{h} \tag{1}$$

The concentrations of compounds (C_m^0, C_m^h) within different parts of the membrane are hard to determine. The membrane/water apparent partition coefficient ($D_{m/v}$) can be estimated, and this allows to replace the concentrations in the both sides of the membrane with the concentrations in the donor (C_D) and acceptor (C_A) sections (Eq. (2)). These concentrations can be easily measured using standard detection methods. [7,78]

$$J = D_m \cdot D_{m/v} \cdot \frac{C_D - C_A}{h} \tag{2}$$

To simplify the equation (2), $D_{m/v}$, D_m, and h can be combined to the membrane permeability (P_m, cm/s). (Eq. (3)) [7,78]

$$P_m = \frac{D_m \cdot D_{m/v}}{h} \tag{3}$$

This means that the flux is defined as the relationship between the P_m and the concentration difference for the donor and acceptor sections, which in turn is related to the concentration change in the donor section over the time ($\mathrm{d}C_D(t)/\mathrm{d}t$), the volume in the donor section (V_D), and the area of the filter (A) (Eq. (4)). [78]

$$J = P_m \cdot (C_D - C_A) = -\frac{V_D}{A} \cdot \frac{\mathrm{d}C_D(t)}{\mathrm{d}t} \tag{4}$$

Deriving $\mathrm{d}C_D(t)/\mathrm{d}t$ from the Eq. (4), a new equation (Eq. (5)) is obtained, which is the basic equation for almost all membrane permeability related equations [7].

$$\frac{\mathrm{d}C_D(t)}{\mathrm{d}t} = -\frac{A}{V_D} \cdot P_m \cdot (C_D - C_A) \tag{5}$$

1.3.5. Equations of membrane permeability

The membrane permeability can be described with the different equations, which are taking into account different number of characteristic parameters (Figure 4) [7]. Most used measures are a flux, an apparent membrane permeability, an effective membrane permeability, and an intrinsic membrane permeability.

The **flux (%)** is the simplest way to describe the membrane permeability. It is the ratio between moles in the acceptor section ($n_A(t)$) and moles at the equilibrium state ($n_{eq}(t)$) at a certain time point (Eq. (6)). Moles at the equilibrium state are found using control samples, where membrane is not added. If solutions have been measured with UV/Vis spectroscopy, then the number of moles is equalised with the UV/Vis absorption (OD). [13]

$$flux = \frac{n_A(t)}{n_{eq}(t)} = \frac{OD_A(t)}{OD_{eq}(t)} \tag{6}$$

While the flux takes into account only moles in the acceptor section and at the equilibrium state, then the **apparent membrane permeability** (P_{app}, cm/s) is taking into account also the volume of solution in the donor section (V_D, ml), the membrane area (A, cm^2) and the time of the experiment (t, s). The apparent membrane permeability can be found from a certain time point (Eq. (7)) or from the change of moles in the acceptor section over the experiment ($\Delta n_A(t)/\Delta t$, Eq. (8)). The apparent membrane permeability for a certain time point is found based on the number of moles in the acceptor section at a certain time point ($n_A(t)$) and the initial number of moles in the donor section ($n_D(0)$). [7]

$$P_{app} = \frac{V_D}{A \cdot t} \cdot \frac{n_A(t)}{n_D(0)} \tag{7}$$

$$P_{app} = \frac{V_D}{A \cdot t \cdot n_D(0)} \cdot \frac{\Delta n_A(t)}{\Delta t} \tag{8}$$

The apparent membrane permeability can be also described with a more complex equation (Eq. (9)), which takes account the ratio of the volumes in the donor and acceptor sections ($r_v = V_D/V_A$), the initial concentration in the donor section ($C_D(0)$), the final concentration in the donor ($C_D(t)$) or the acceptor ($C_A(t)$) section, the apparent porosity of the filter (ε_a), and the area of the filter (A) (Figure 4). [7]

$$\begin{aligned} P_{app} &= -\frac{2.303 \cdot V_D}{A \cdot t \cdot \varepsilon_a} \cdot \left(\frac{1}{1 + r_v}\right) \cdot log_{10}\left[-r_v + (1 + r_v) \cdot \frac{C_D(t)}{C_D(0)}\right] = \\ &= -\frac{2.303 \cdot V_D}{A \cdot t \cdot \varepsilon_a} \cdot \left(\frac{1}{1 + r_v}\right) \cdot log_{10}\left[1 - \left(1 + \frac{1}{r_v}\right) \cdot \frac{C_A(t)}{C_D(0)}\right] \end{aligned} \tag{9}$$

Equation with the final concentration in the donor section can be converted to the equation with the final concentration in the acceptor section using the mass-

balance rule. The mass balance rule is connecting the initial concentration in the donor section and the final concentration in the acceptor and donor sections to one equation (Eq. (10)) using the volumes of solutions in the donor and acceptor sections. Usually the equation with the final concentration in the acceptor section is giving results with higher accuracy. [7]

$$V_D \cdot C_D(0) = V_A \cdot C_A(t) + V_D \cdot C_D(t) \tag{10}$$

Compared to the apparent membrane permeability, the **effective membrane permeability** (P_e, cm/s) is additionally taking into account the membrane retention (R_M) and the lag time (τ_{LAG}, s) [7] (Figure 4). Taking account of the membrane retention is important for lipophilic compounds, which tend to accumulate in the membrane [68]. Membrane retention (Eq. (11)) is calculated based on the initial and final concentrations and the volumes of solution in the acceptor and donor sections [7].

$$R_M = 1 - \frac{C_D(t)}{C_D(0)} - \frac{V_A}{V_D} \cdot \frac{C_A(t)}{C_D(0)} \tag{11}$$

The lag time (Eq. (12)) can be found from the empirical relationship with the membrane retention. [7]

$$\tau_{LAG} = (54 \cdot R_M + 1) \cdot 60\text{s} \tag{12}$$

When considering the membrane retention the mass-balance equation is changing (Eq. (13)), and includes the volume of the membrane (V_m) and the concentration of compound in the membrane ($C_m(t)$). [7]

$$V_D \cdot C_D(0) = V_A \cdot C_A(t) + V_D \cdot C_D(t) + V_m \cdot C_m(t) \tag{13}$$

The membrane permeability for an experiment, where in the acceptor and donor sections have the same pH (pH = 7.4), is described with the effective membrane permeability for iso-pH (Eq. (14)). In the iso-pH experiment (Figure 4), the equilibrium state is equal to the volumes ratio in the donor and acceptor sections ($r_v = V_D/V_A$). [7]

$$\begin{aligned} P_e &= -\frac{2.303 \cdot V_D}{A \cdot (t - \tau_{LAG}) \cdot \varepsilon_a} \cdot \left(\frac{1}{1 + r_v}\right) \cdot log_{10}\left[-r_v + \left(\frac{1 + r_v}{1 - R_M}\right) \cdot \frac{C_D(t)}{C_D(0)}\right] = \\ &= -\frac{2.303 \cdot V_D}{A \cdot (t - \tau_{LAG}) \cdot \varepsilon_a} \cdot \left(\frac{1}{1 + r_v}\right) \cdot log_{10}\left[1 - \left(\frac{1 + r_v^{-1}}{1 - \mathrm{R_M}}\right) \cdot \frac{C_A(t)}{C_D(0)}\right] \end{aligned} \tag{14}$$

The iso-pH equation cannot be used for describing the system, where in the acceptor and donor sections have different pH-s, i.e. the gradient-pH system. In the gradient-pH system (Figure 4), the equilibrium is not equal to the ratio between the volumes of the donor and acceptor sections and thus the sink asymmetry ratio should be used (r_a, Eq. (15)). The sink asymmetry ratio is the ratio of volumes that accounts for the experimental condition differences (e.g., pH, surfactants) in the acceptor and donor sections. The sink asymmetry ratio takes account $P_e^{A \to D}$, which is equal to the iso-pH membrane permeability value

at pH 7.4. To find $P_e^{D\rightarrow A}$, the Eq. (16) is solved iteratively until both sections have the same value. [7]

$$r_a = \left(\frac{V_D}{V_A}\right) \cdot \left(\frac{P_e^{A\rightarrow D}}{P_e^{D\rightarrow A}}\right) \tag{15}$$

$$P_e^{D\rightarrow A} = -\frac{2.303 \cdot V_D}{A \cdot (t - \tau_{LAG}) \cdot \varepsilon_a} \cdot \left(\frac{1}{1 + r_a}\right) \cdot log_{10}\left[-r_a + \left(\frac{1 + r_a}{1 - R_M}\right) \cdot \frac{C_D(t)}{C_D(0)}\right] \tag{16}$$

Flux, apparent, and effective membrane permeabilities describe the permeability in the membrane and the effect of the aqueous boundary layer (ABL, known also as an unstirred water layer, UWL). [67] The ABL influences the membrane permeability for high-permeable compounds, i.e. such as lipophilic compounds [68]. The ABL exists in the both sides of the membrane and molecules move through the ABL only using diffusion (Figure 4). The ABL in the intestinal epithelium is around 30–100 µm, while in the unstirred PAMPA experiment, the ABL is around 4000 µm [94]. The thickness of the ABL can be reduced using stirring, but it is not possible to absolutely remove it [99]. The influence of the ABL is visible in the high permeability region of the pH-effective membrane permeability profiles, where the effective membrane permeability is not increasing, although all compounds are not yet unionised [67].

To take into account the ABL, the effective membrane permeability (P_e, Eq. (17)) is inversely related to the permeability of the aqueous boundary layer (P_{ABL}) and the membrane permeability (P_m, i.e. membrane resistance) [99].

$$\frac{1}{P_e} = \frac{1}{P_{ABL}} + \frac{1}{P_m} \tag{17}$$

The P_m describes the real membrane permeability at certain pH without the aqueous boundary layer effect. The P_m (Eq. (18)) is related to the **intrinsic membrane permeability** (P_o) and to the ratio between the total sample concentration and the unionised form concentration (it is further in the text used as the fraction of unionised species, f_u). The ratio between P_o and P_m is called a permeability-pH equation, which is equal to $1/f_u$. The intrinsic membrane permeability describes the membrane permeability of the unionised species. [99]

$$P_m = f_u \cdot P_o \tag{18}$$

The calculation of the unionised species fraction depends on the number and type of ionisable groups [99]. For the basic compounds with one ionisable group, the f_u is calculated based on the equation (19) and this requires pK_a and pH values. Other examples can be found in reference [99].

$$f_u = \frac{1}{10^{pK_a - pH} + 1} \tag{19}$$

Replacing P_m, which is unknown variable in the equation (17), into the equation (18), resulted in the equation (20) that can be used to find the P_{ABL} and P_o parameters using the weighted least-squares analysis, where $1/P_e$ is the dependent variables (membrane permeability measurements at different pH-s) and $1/f_u$ is the independent variables (corresponds to respective pH-s). [99]

$$\frac{1}{P_e} = \frac{1}{P_{ABL}} + \frac{1}{P_o \cdot f_u} \tag{20}$$

1.3.6. Using membrane permeability

Methods for determining membrane permeability, like the PAMPA, have been developed with the aim of analysing and evaluating the absorption of drug substances and their candidates in the GIT. Therefore, it is important to evaluate how well the membrane permeability by the PAMPA coincides with what we know about the permeability in the GIT. It has been estimated that approximately 90% of drug substances are passively transported [41], which gives to the PAMPA method and its modifications a relatively wide coverage.

The BCS considers the %HIA as reference for the permeability [10], which is extensively compared with the PAMPA membrane permeability values. Literature witnesses several comparative studies between the membrane permeability and the %HIA at different pH-s: 5.0, 6.2 and 7.4 [100,101]; 5.5, 6.5 and 7.4 [79,102]; 5.5 and 7.4 [66,103]; 6.5 [86]; 6.5 and 7.4 [13]; 6.8 and 4–8 [77]; and 7.4 [104]. These comparisons show that the PAMPA method tends to underestimate the %HIA for compounds with active transport and overestimates the %HIA for low solubility or efflux compounds [100]. The best match with the %HIA have been obtained if the membrane permeability values at different pH-s are analysed together and the highest membrane permeability is selected [77] or the calculated sum of the membrane permeability is used [100].

The comparison of the membrane permeability and %HIA has been done quantitatively using the exponential relationship (Eq. (21)) between the apparent membrane permeability and the human intestinal absorption [79,80,86]. Comparing the membrane permeability at pH 5.5, 6.5, and 7.4 with the human intestinal absorption using the equation (21), shows that stronger correlation exists at acidic pH-s (pH 5.5 and 6.5) compared to the pH 7.4 [79]. This is due to the chemical class-based pH-permeability profiles, which are influenced by the fraction of unionised species based on the pH-partition theory [29].

$$\%\text{HIA} = \left(1 - \mathrm{e}^{a \cdot P_{app}}\right) \cdot 100 \tag{21}$$

A more precise comparison needs a cutoff value for high- and low-permeable compounds, which highly depends on the membrane in the PAMPA system [87]. For example, the membrane permeability described with the flux has been divided into three classes: high 25–100%, moderate 5–25%, and low < 5% [13]. Also, the cutoff value has been determined for the effective membrane

permeability of positively charged compounds [48], where low permeability is defined as $P_e < 0.1 \cdot 10^{-6}$ cm/s ($\log P_e < -7$), moderate permeability as $0.1 \cdot 10^{-6}$ cm/s $\leq P_e \leq 1 \cdot 10^{-6}$ cm/s ($-7 \leq \log P_e \leq -6$), and high permeability as $P_e \geq 1 \cdot 10^{-6}$ cm/s ($\log P_e \geq -6$).

Membrane permeability values have been successfully used along with solubility values to accurately classify compounds according to the BCS [105]. Recently, the PAMPA have been used to measure pH-permeability profiles, which have been successfully combined with the pH-solubility profiles to determine the pH window with optimal conditions for good oral bioavailability [31].

Correlation of PAMPA and Caco-2 membrane permeability values at pH 5.5 and 7.4 [66] and pH 7.3 [106] gives almost ideal correlations and indicates that most of drug substances are mainly transported passively. The deviation from the ideal correlation is caused by the Caco-2 taking account the carrier-mediated transport, which also includes efflux [66].

The PAMPA method has been successfully and widely used in several recent drug discovery research projects and campaigns, leading the way to better absorbed drug candidates. For example, the membrane permeability measurements have been performed to analyse drug substance candidates for disease categories like antitumors [107,108,109,110], diabetes [111], anti-adenoviral therapies [112], anti-inflammatory [110,113,114], metabolic disorder [110], anticoagulants [115], neurological disorders [116], etc. Also, the PAMPA method has been used to study permeability properties for cyclic peptomers, which have typically low cellular permeability [117] and sulfoximines that are rarely studied in drug discovery [118]. Drug design is not the only application as the permeability properties are concern of wider chemical groups. The PAMPA has been used for measuring the pH-permeability profiles for various natural compounds. Good example is natural flavonoids from *Silybum marianum* (L.) Gaertn. and their derivatives from milk thistle [119], which membrane permeability was analysed at pH 5.0, 6.2, and 7.4.

1.3.7. Prediction models for membrane permeability

Although the PAMPA measurements are relatively easy to perform compared with the animal- and cell-based methods, they still have substantial costs. In addition, the experimental measurements can only be performed when the compound is already in the stock. If the compound is not available, like in case of *in silico* virtual screening campaigns, or is not yet synthesised, then cheminformatics methods such as the quantitative structure-activity relationship (QSAR) serves as a suitable alternative to obtain estimates about the membrane permeability.

The first QSAR models for the membrane permeability were published six years after [106] the publication of the PAMPA method [13]. During the past 15 years, a number of QSAR models for the membrane permeability have been published [48,87,106,113,119,120,121,122,123,124,125,126,127,128,129], and

are recently reviewed and analysed [14]. The QSAR models have been derived for different type of PAMPA values (Section 1.3.5), at variety of experimental conditions (Section 1.3.1, 1.3.2, and 1.3.3), and with different modelling methods and descriptors.

The most modelled membrane is 10% (w/v) lecithin in 1,9-decadiene at pH 6.3 [106] and pH 7.3 [106,121,122,125,126]. Another more frequently modelled membrane includes 20% of lecithin in dodecane with the permeability data for pH 7.4 [119,129]. Less modelled membranes (Table 2) are PAMPA-HDM [87], PAMPA-DOPC [87,120], PAMPA-DS [87], PAMPA-BM [123], 10% (w/v) lecithin and 0.5% cholesterol in dodecane [48], 5% lecithin in dodecane [124], trilayer PAMPA [128], and 1% lecithin in dodecane [113]. Also, the QSAR model is derived from the different membrane permeability values measured at the same pH [127]. The membrane permeability QSAR models have been developed for variety of pH-s, such as pH 5.5 [113,123,127], 6.0 [123], 6.3 [106], 6.5 [48,123,124], 7.3 [106,121,122,125,126], and 7.4 [119,123,127,129]. The analysis of used membrane composition and pH-s indicate that the models are mostly for neutral or near neutral pH-s and the composition of membranes is diverse. The lack of QSAR models for more diverse pH values is mainly caused by the absence of high-quality experimental membrane permeability data over a wide range of pH that are measured with the same experimental protocol.

From different measures (Section 1.3.5) used to express membrane permeability, the most frequently modelled is the apparent membrane permeability [106,113,121,122,123,124,125,126,127,128], but also the models for the effective membrane permeability [48,119,129], the flux [113,123], and the intrinsic membrane permeability [87,120] exist. A possible reason for using the apparent membrane permeability can be related to the complexity of the calculation. The apparent membrane permeability is simpler to calculate than the effective membrane permeability, because it does not consider the membrane properties, such as the membrane retention and the lag time (Section 1.3.5).

The modelled datasets are usually small with less than 100 data points and often related to the certain types of compounds: permanently positive charged compounds [48], benzoic acids [124], pyridines and quinolines [124], cyclic peptides [128], and β-hydroxy-β-arylalkanoic acid analogues [113]. Two modelled sets of data are much larger than others. They both describe a PAMPA system at pH 7.4 and membrane is the 20% dodecane solution of lecithin [119, 129]. The first, smaller dataset consists of 251 compounds [119]. The detailed analysis reveals that this dataset describes the membrane permeability ($\log P_m$, i.e. without ABL) and not the effective membrane permeability ($\log P_e$) as stated in the paper [119]. The second dataset is remarkably larger [129], includes more than 4000 data points at pH 7.4. The dataset is not openly available, but the data range provided by the original paper [129] does not correspond to the conventional membrane permeability range. Typical membrane permeability values are negative, while the range presented in the paper [129] is mainly on positive side.

Three main mathematical representations have been used for the membrane permeability QSAR models: multiple linear regression (MLR) [87,106,113,119, 120,121,122,123,124,125,126,127], non-linear regression (NLR) [122,123,128], and partial-least square methods (PLS) [48,106,113,120,121]. The biggest difference between the PLS and MLR models [106], is the number of descriptors in the model. At the same time, the statistical parameters of the models are very similar, suggesting that a larger number of descriptors do not improve the model description and prediction ability [106]. Also, machine learning methods, an artificial-neural network (ANN) [113,127] and a support vector regression (SVR) [129], have been used in development of QSAR models for the membrane permeability. Next to the quantitative models, the qualitative models for the membrane permeability are more an exception. To the best of our knowledge only one classification model has been developed for the membrane permeability and it is using a support vector classification (SVC) approach [129].

Many QSAR models for membrane permeability include the logarithm of the octanol-water partition coefficient ($\log P_{ow}$) [87,106,121,122,123,125], and less the logarithm of the octanol-water distribution coefficient [119,124]. Partition and distribution coefficients describe the lipophilicity of molecules that is largely determined by the molecular size and electrostatic interactions. The second most frequently occurring descriptor in published QSAR models accounts for molecular interactions related to the hydrogen bonds [106, 121,122,124,125,126,127,129] that occurs in the environment when interacting with continuum (for example, water) and in the cellular membrane with lipophilic and electronegative membrane constituents [87]. The third widely used descriptor is the absolute value of the difference between pK_a and pH (|pK_a-pH|) [106,121,122,125], which is included to describe the fraction of unionised species and is mostly used for simple acidic and basic drug substances. Lipophilicity, ionisation, and hydrogen bonds-related descriptors have been combined [106,121,122,125] to provide good prediction models for membrane permeability. Several QSAR models use larger number and complex molecular descriptors, like Abraham parameters [87,120] or descriptors calculated with VolSurf [106,121], FORESEE [48], CODESSA [126], CODESSA PRO [127], QSARmodel [127], OpenEye [129], and Dragon [113] software.

The validation of the QSAR models for the membrane permeability with an independent dataset is a rarity rather than a routine activity [106,113,126,127]. Typically, internal validation in the form of leave-one-out cross-validation has been used [48,113,119,120,121,122,123,124,125,126,127,128]. This suggests that the developed models for the membrane permeability are not systematically tested. Similarly, the analysis and description of a model's applicability domain for published membrane permeability QSAR models have been rare. Only for three models the applicability domain based on the molecular structure was defined or analysed: two models employed hydrophilic/hydrophobic properties

[122] and one model analysed the applicability domain in the form of a Willams plot [126].

In the conclusion, several QSAR models have been developed for the membrane permeability measured by PAMPA, but most of them have drawbacks, like not validated with an external validation set, small set of data, not taking account full pH range in the GIT, etc. These drawbacks indicate that externally validated QSAR models for different pH are absent in the literature, although these models could be highly usable in the early stage of drug discovery, and also can be utilized in other industrial sectors, such as cosmetics, food, and industrial chemicals to reduce the necessity for the animal testing [130].

2. AIMS OF THE STUDY

Membrane permeability of drug substances has been mostly measured and modelled at neutral pH. Given the diversity of pH in the GIT (Figure 1), it may give a misleading estimation of membrane permeability at an early stage of drug discovery. Thus, the aim of this thesis is to study in detail the effect of pH on the membrane permeability of drug substances. For this purpose, the membrane permeability of drug substances was measured, analysed, and modelled taking into account a wide pH range in the GIT. The study consisted of several inter-related sub-objectives:

- Establish a systematic database of the membrane permeability of drug substances at four pH-s (3, 5, 7.4 and 9) using the PAMPA (Papers **I**–**IV**);
- Assess the ability of an artificial membrane method to describe the absorption of drug substances in the GIT (measured membrane permeability vs. human intestinal absorption) (Paper **IV**);
- Develop quantitative (regression) prediction models for describing the membrane permeability at measured pH-s and at the highest and intrinsic membrane permeability in the complete sets of data, as well as separately for acidic, basic, amphoteric, and neutral drug substances to identify important structural parameters that determine the membrane permeability of drug substances (Papers **I**–**IV**);
- Develop qualitative (classification) prediction models for the evaluation of the membrane permeability of drug substances at all measured pH-s, and for the highest and intrinsic membrane permeability in the complete sets of data and analyse important structural parameters that determine the membrane permeability class of drug substances (Paper **V**);
- Test prediction models with an external blind validation set (Papers **I**–**V**);
- Test the applicability of the developed classification models *via* predicting the permeability classes of the U.S. FDA reference drug substances and demonstrate the use of these models to predict the BCS permeability component (Paper **V**).

3. MATERIALS AND METHODS

3.1. Experimental measurements

3.1.1. Chemicals

Drug substances (for simplicity, the drug substances are referred to as "compounds") were purchased from Sigma-Aldrich Co. and from various providers through MolPort (www.molport.com). The drug substances were used in the experiments without further purification. L-α-phosphatidylcholine (lecithin, 99%), n-dodecane, dimethyl sulfoxide (DMSO), and sodium chloride were purchased from Sigma-Aldrich Co. Potassium dihydrogen phosphate (KH_2PO_4), disodium hydrogen phosphate (Na_2HPO_4), phosphoric acid, and sodium hydroxide were purchased from Merck KGaA. All solutions were prepared using deionised water (Milli-Q, Millipore).

3.1.2. Membrane permeability measurements

The membrane permeability was measured with the PAMPA (Section 1.3). The PAMPA "sandwich" was formed using 96-well filter plate (donor plate, Millipore, MAIPN4550) and acceptor plate (Millipore, MATRNPS50), which were placed on the top of each other (Figure 3) and were separated by a hydrophobic PVDF microfilter (pore size 0.45 μm, apparent porosity of the filter (ε_a) 89%, area of the filter (A) 0.24 cm^2). The PVDF microfilter was coated with the membrane solution, which was selected based on the following criteria:

- The membrane solution should include as few compounds as possible to have a good reproducibility,
- The membrane should not have high retention factor.

The selected membrane solution includes 1% (w/v) lecithin in dodecane, which is similar to the original PAMPA membrane (Table 2). 5 μl of this solution was added to the PVDF microfilter to form an artificial membrane.

Drug substances were dissolved in DMSO and diluted using phosphate buffers with pH-s at 3, 5, 7.4, and 9. These pH-s were selected to have a comparable difference (approximately two pH units) from each other and covered full pH range in the GIT. Phosphate buffer (0.01 M) with exact pH was formed using KH_2PO_4, Na_2HPO_4, phosphoric acid, and sodium hydroxide. The pH was measured with the pH-meter (ETS-D5, IKA® and SevenCompact S210, Mettler Toledo) and three-point calibrated pH-electrode (VWR universal pH electrode and InLab® Basics, Mettler Toledo) and allowed difference from a selected pH for the buffer was ±0.02 pH-units. The constant ionic strength (0.15 M) for the buffer was fixed using sodium chloride. The concentration of drug substances in a solution was 500 μM and including 0.5% (v/v, Papers **II**–**IV**) or 5% (v/v, Paper **I**) of DMSO. DMSO was used to prepare a stock solution and increase the solubility of compounds in the buffer. In case of

limited solubility in the DMSO/water solution, lower concentration was introduced. 0.15 ml (Paper **I**) or 0.2 ml (Papers **II**–**IV**) of drug substance's solution was added to the filter (donor) plate (V_D). The acceptor plate received 0.3 ml of pH 7.4 solution (V_A) with the same concentration of DMSO (0.5% or 5%, v/v) as at the donor plate. The donor plate was covered with the adhesive film and the sandwich was coated with the lid to avoid evaporation. The PAMPA system was incubated at 25 °C.

The time dependence of membrane permeability was measured over 48 hours, because the membrane permeability needs to be calculated before equilibrium between the donor and acceptor sections. Longer experiment time was introduced to get higher concentration in the acceptor section for low-permeable compounds. For this, a full UV/Vis spectrum in each well was measured directly from the acceptor plate for 13 times in the following time points: 2, 4, 6, 8, 10, 12, 24, 26, 28, 30, 32, 34, and 48 hours. The UV/Vis spectra were measured using SpectroStar^{Nano} UV/Vis-microplate reader (version 2.10, BMG LABTECH) with a selected range of wavelength (220–500 nm) and resolution (1 nm). For each time point, the donor plate was removed from the acceptor plate and placed in the plate holder. After measurement, the donor plate was placed onto the acceptor plate and incubated further. This technological innovation was introduced as methodological solution for measuring the time dependence with the PAMPA system in semi-high-throughput manner (Paper **I**). Direct measurements from the acceptor plate are used, because most of drug substances have a maximum UV/Vis absorption at a higher wavelength than the significant UV/Vis absorption spectra of the acceptor plate (range 220–270 nm). At the last time-point (48 hours), the UV/Vis spectra from both sections were measured also using UV/Vis-microplates (Greiner bio-one, UV-Star® Microplate, 655801) to obtain more accurate UV/Vis spectrum of compounds with low absorption wavelength and calculate the membrane retention (R_M, Eq. (11)).

For the analysis, the most precise wavelength for the compound was selected based on the UV/Vis spectra, i.e. usually that with the highest absorption value. For each compound, the calibration curves for pH 3, 5, 7.4, and 9 were measured and used for the calculation of concentration in the acceptor ($C_A(t)$) and donor ($C_D(t)$) sections. Blank measurement was carried out using the buffer solution with the same concentration of DMSO as in the experiment.

The effective membrane permeability for iso-pH (donor and acceptor sections have same pH, i.e. pH = 7.4), was calculated using the equation (14) with $C_A(t)$ (i.e. second half). For the gradient-pH (donor and acceptor sections have different pH-s), the effective membrane permeability was calculated with the equation (22), which is derived from the equation (16) similar way as equation (14) (Paper **I**).

$$P_e^{D\to A} = -\frac{2.303 \cdot V_D}{A \cdot (t - \tau_{LAG}) \cdot \varepsilon_a} \cdot \left(\frac{1}{1 + r_a}\right) \cdot log_{10}\left[1 - \left(\frac{V_A + r_a \cdot V_A}{(1 - R_M) \cdot V_D}\right) \cdot \frac{C_A(t)}{C_D(0)}\right] \quad (22)$$

Equations with $C_A(t)$ (Eq. (14) and (22)) have been used, because these gave more precise membrane permeability value for low-permeable compounds than equations with $C_D(t)$ [7]. These equations allow to calculate the membrane permeability for every time point over a time dependence measurement, where only the UV/Vis spectra in the acceptor section were measured.

Spectral data analysis and membrane permeability calculations were performed by the MARS data analysis software (version 2.40, BMG LABTECH) and the Microsoft Excel, respectively.

3.2. Data series

The data series used in the study are either experimentally measured or derived from the experimental data. An overview of the different data series is given by the relationship between pH and membrane permeability to the basic drug substance *Propranolol* (Figure 5).

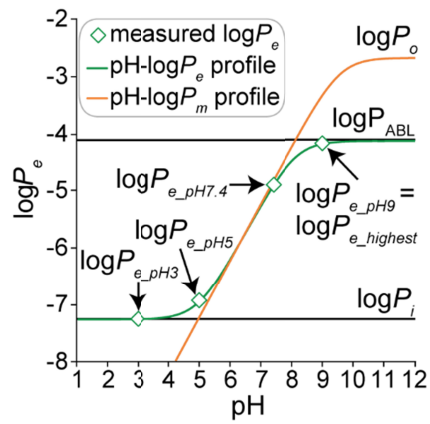


Figure 5. The data series used in the study. Example for *Propranolol*: pK_a = 9.53 [131], logP_o = –2.69, logP_i = –7.25, and logP_{ABL} = –4.11.

Four data series were experimentally measured and they describe **membrane permeability at certain pH** (Papers **I**–**V**, Figure 5: logP_{e_pH3}, logP_{e_pH5}, log$P_{e_pH7.4}$, and logP_{e_pH9}). Measured membrane permeability at different pH-s were analysed as a pH-effective membrane permeability profile (Figure 5: pH-logP_e profile, hereafter referred as a pH-permeability profile), which shows how pH influences the membrane permeability. The membrane permeability of highly permeable compounds (logP_e > –5) is greatly influenced by the permeability of the aqueous boundary layer (Figure 5: logP_{ABL}). For *Propranolol*, the membrane permeability at pH 9 is highly influenced by the ABL (logP_{ABL} = –4.11). The analysis of measured membrane permeability data (Paper **IV**) revealed that for the current data series at acidic pH-s only 11 to 14% of the compounds are affected by the ABL and at neutral and basic pH-s around 30% of the compounds are influenced by the ABL. For the unstirred PAMPA system that is used in the current study the logP_{ABL} is between –4 to –5 and varies

depending on the compound (Article **IV**: Table S1, Table S2). This allows to conclude that although the measured data at various pH values is affected by the ABL, the main descriptive property is the membrane resistance. Additional parameter, which influences the membrane permeability values, is the membrane permeability of the ionised compounds (Figure 5: logP_i). The logP_i is influenced by the limit of detection and quantitation and it varies depending on the compound. If the logP_{ABL} is the upper limit of measured membrane permeability value, then the logP_i is the lower limit of measured membrane permeability value.

Derived data series are the highest, average, and intrinsic membrane permeability. For the **highest membrane permeability** (Figure 5: log$P_{e_highest}$, Papers **I**, **III**, **IV**, and **V**), the maximum membrane permeability value over pH-s 3, 5, 7.4, and 9 was selected. The log$P_{e_highest}$ describes the maximum absorption in the GIT. In the case of *Propranolol*, the highest membrane permeability is equal to the membrane permeability at pH 9 (Figure 5: log$P_{e_highest}$ and logP_{e_pH9}). This data series should be the most accurate, because the highest membrane permeability values are typically more precise than the lower membrane permeability values. The disadvantage of the log$P_{e_highest}$ is that ~36% of measured compounds are influenced by the ABL (Paper **IV**), which means that a considerable number of compounds have similar log$P_{e_highest}$ values.

The **average membrane permeability** over four pH-s (log$P_{e_average}$, Paper **III**) was used for the modelling of neutral drug substances, because in this case the membrane permeability values are similar for all pH-s, i.e. they follow the same molecular interaction pattern common for unionised compounds. The log$P_{e_average}$ was calculated based on 12 experiments, which should be more accurate than the individual membrane permeability values at a certain pH for neutral compounds.

The **intrinsic membrane permeability** (Figure 5: logP_o, Papers **IV** and **V**) describes the membrane permeability of the unionised compounds. The logP_o was calculated using the pCEL-X^{TM} software [132] from the logP_e, logP_{ABL} (Figure **5**), and f_u according to the equation (20). The f_u for the basic compound *Propranolol* with one ionisable group is calculated based on the equation (19) and this requires pK_a and pH values. For the calculation of the f_u values within current study, experimental pK_a values were collected from different sources (mainly the Wiki-pKa^{TM} [131]). When experimental pK_a values were not available, these values were calculated by the JChem for Excel [133]. The logP_o value for *Propranolol* is –2.69, which is around 1.5 units higher than the membrane permeability at pH 9. The logP_o values together with the f_u values (Eq. (18)) can be used to construct a pH-membrane permeability profile for the membrane resistance (Figure 5: pH-logP_m profile), which is not including the ABL compared to the pH-logP_e profile.

3.3. Criteria for the classification

The classification of compounds with high and low permeable requires a cutoff value, which depends on the experimental system. The cutoff for high and low permeability (Paper **IV**) was determined from comparison of the %HIA data [134] with the membrane permeability data from the PAMPA measurements. On the %HIA scale, compounds are classified as being high or low absorbed, with a cutoff limit of 85% according to the BCS [10,27]. The $\log P_{e\_highest}$ was used to determine the cutoff value for the membrane permeability because it takes into account the full pH range in the GIT. Two criteria were used to identify the cutoff value for the membrane permeability:

- The majority of low absorbed compounds must be grouped correctly,
- For low-permeable compounds, a small margin for prediction and experimental error (~0.5 $\log P_e$ unit) is provided.

Considering these criteria, the cutoff value for the membrane permeability was set to $\log P_e = -6.20$. According to this, if $\log P_e \geq -6.2$ the compound is a high permeable, and if $\log P_e < -6.2$ the compound is a low permeable. The cutoff was used to divide compounds into high and low permeable for all data series in Papers **IV** and **V**.

3.4. Drug substance molecules

The membrane permeability data were measured for two different sets of drug substances: an initial set of compounds and an external validation set of compounds (Table 3). The **initial set of compounds** was randomly selected, then experimentally measured and used to develop of the prediction models. The **external validation set of compounds** was collected after the development of the prediction models and consisted of randomly selected drug substances that were first predicted and then experimentally measured for verifying the accuracy of the prediction. All drug substances for the measurements were selected on the basis of the following four criteria:

- It can be measured by the detection method (UV/Vis-measurable);
- It is an approved or former drug substance;
- It has no extremely high logarithm of the octanol-water partition coefficient ($\log P_{ow}$) or molecular weight;
- It is available for purchase with an acceptable price.

The compounds were reused in sequential papers (Table 3). The membrane permeability values of some compounds changed somewhat in papers, because they were either repeatedly measured or recalculated to increase the accuracy of the experimental data. For various reasons, some compounds were eliminated in sequential papers: the membrane permeability data were unreliable at a certain pH, failure to calculate the $\log P_o$ values, or the molecular structure could not be unambiguously defined (such as tetracyclines).

Table 3. The number of compounds in the initial and external validation sets in papers and the dynamics of adding drug substances.

Paper	Initial	External validation
I	58	15
II: acids	36 (15 from Paper **I** and 21 new compounds)	1
II: bases	61 (15 from Paper **I** and 46 new compounds)	1
III: neutrals	15 (5 from Paper **I** and 10 new compounds)	1
III: ampholytes	60 (30 from Paper **I** and 30 new compounds)	1
IV and **V**	178 (144 from Papers **I**, **II** and, **III** and 34 new compounds)	60

The sets of compounds have been analysed based on chemical classes, which were defined considering the following rules [90]:

- Acidic compounds have pK_a value(s) for acidic functional group(s) below 12.0;
- Basic compounds have pK_a value(s) for basic functional group(s) over 0.0;
- Neutral compounds are unionised, e.g. do not have or have irrelevant ionisable group (acid functional group with a pK_a value over 12.0 and basic functional group below 0.0, i.e. under physiological conditions, these compounds would not be ionisable);
- Amphoteric compounds have pK_a value(s) for basic functional group(s) over 0.0 and below 12.0 for acidic functional group(s).

The grouping was performed based on the experimental pK_a values from the Wiki-pKa^{TM} database [131] or from scientific publications (see Papers **I**–**IV**). When an experimental pK_a value was not available, the calculated value by the Advanced Chemistry Development (ACD/Labs) Software V11.02 were queried from the SciFinder [135] (Papers **I**–**III**) or was calculated with the JChem for Excel [133] (Paper **IV**).

In Paper **I**, the initial set of compounds included 58 drug substances, which consisted of 18 acids, 10 bases, 24 ampholytes, and 6 neutrals. The comparison with the approved drug substances from the DrugBank [136,137] shows that due to the random selection of compounds, the distribution of chemical classes in the initial set of compounds do not match with that of the set of approved drug substances (Figure 6). The external validation set of compounds included 15 drug substances, which composed of 2 acidic, 10 basic and 3 neutral compounds, that also does not compare well with the set of approved drug substances, because it includes remarkable more basic compounds and no amphoteric compounds (Figure 6).

In Papers **II** and **III**, for the analysis of chemical classes a substantial amount of drug substances was added and measured (Table 3). The distribution of chemical classes in Papers **II** and **III** is similar to the set of approved drug substances (Figure 6). The developed models were externally validated with one additional compound for each chemical class.

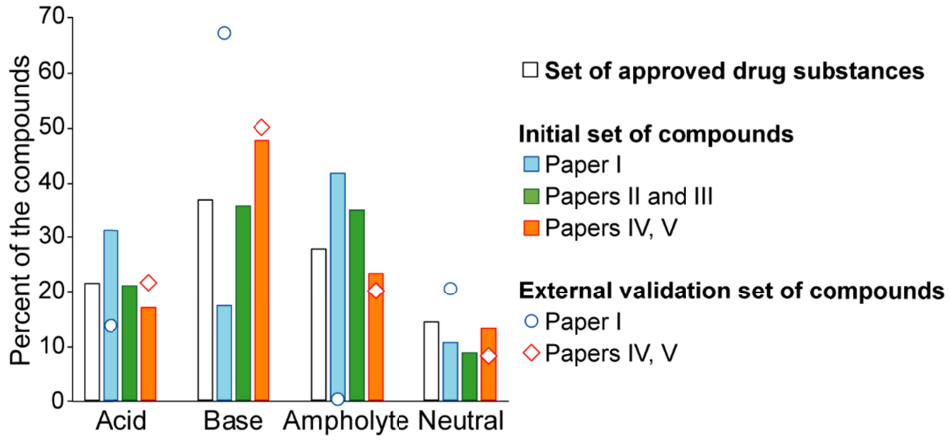


Figure 6. The distribution of chemical classes in the initial and external validation sets of compounds in Papers **I**–**V** and in the set of approved drug substances. All sizes of sets of compounds were normalised to 100 for the comparison purposes.

In Papers **IV** and **V**, the initial set of compounds included 178 drug substances and the external validation set of compounds 60 drug substances (Table 3). The initial set of compounds contained 30 acidic, 84 basic, 41 amphoteric and 23 neutral compounds. The external validation set of compounds consisted of 13 acidic, 30 basic, 12 amphoteric, and 5 neutral drug substances. The distribution of chemical classes for the initial and external validation sets of compounds were in good agreement to the set of approved drug substances (Figure 6).

3.5. U.S. FDA reference drug substances

The U.S. FDA uses 40 reference (model) drug substances [10] for controlling the suitability of permeability measurement methods in detecting high and low permeability classes of drug substance candidates within the BCS. The list includes 11 high, 10 moderate, 10 low, and 5 zero-permeable drug substances, and 4 efflux substrates (Figure 7, Paper **V**: Table 6).

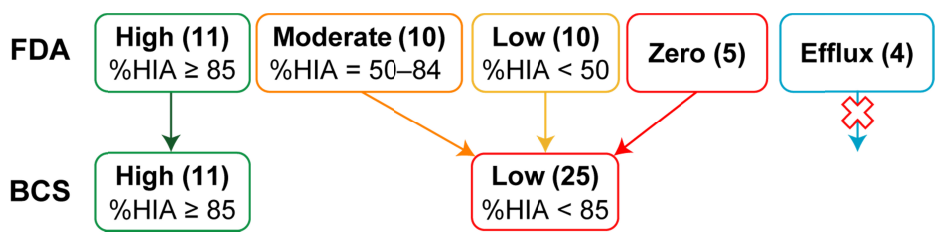


Figure 7. The permeability classes of the U.S. FDA reference drug substances and the conversion to the BCS classes.

The BCS (Section 1.1.2) classifies compounds into two classes and therefore the U.S. FDA reference drug substances were grouped into two classes as

shown in Figure 7. The U.S. FDA reference drug substances were used as a third validation set (in addition to the validation and external validation set) and also provided a practical use case for predicting permeability classes for the BCS. This allowed a more comprehensive testing of the developed classification models (Paper **V**) for their ability to distinguish high- and low-permeable compounds.

Efflux substrates were excluded from the set of the U.S. FDA reference drug substances, because the PAMPA method does not take into account efflux transport. Also, 4 low-permeable compounds (3 polymers and 1 polysaccharide) were excluded. In total, 32 U.S. FDA reference drug substances (11 high- and 21 low-permeable compounds) were used in testing the applicability of the classification models. Comparison of the U.S. FDA reference drug substances with the sets of initial and external validation compounds showed some overlap in compounds. Specifically, the compounds with high permeability in the BCS overlapped to a large extent with the internal and external validation sets of compounds (10 out of 11). The situation is the opposite of low permeability in the BCS, where mainly new drug substances are in the set (15 of 21). Nevertheless, the set of the U.S. FDA reference drug substances provides a valuable comparison of how consistent the estimation of PAMPA measurements and the predictions from the classification models is, when compared with the permeability data measured in the GIT (*in vivo*).

3.6. Cheminformatics modelling

QSAR models describe the relationship (Eq. (23)) between the property and the structure of the molecule, where structural parameters are presented in the form of molecular descriptors (D) [138].

$$Property = f(D) \tag{23}$$

In the current thesis, the property is the membrane permeability measured by the PAMPA and it can be continuous (Section 3.2) or categorical (Section 3.3). Continuous data have been used in the development of quantitative (regression) models (Papers **I**–**IV**) and categorical data for the development of qualitative (classification) models (Papers **IV** and **V**). The process of developing regression and classification models includes similar steps (Figure 8), which are performed using different methodological approaches. All steps are presented in the next sections. The final step for both modelling approaches was organising models and data into the QSAR Data Bank format [139] and archiving to the QsarDB repository [140,141], where all models are openly available and accessible for the prediction.

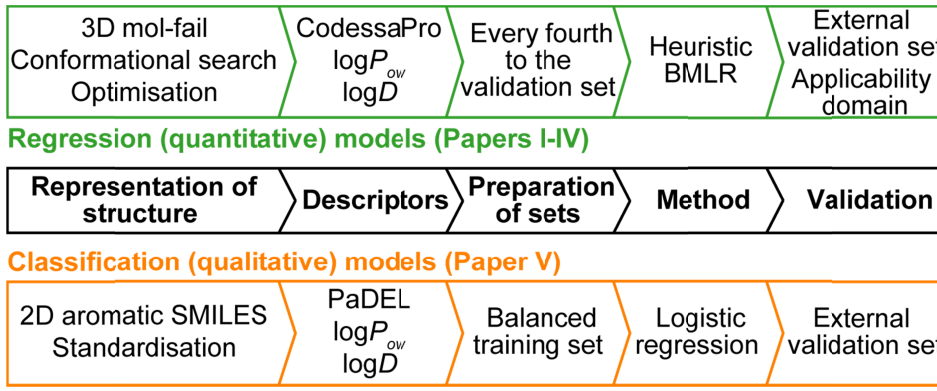


Figure 8. Steps to derive regression and classification models.

3.6.1. Representation of the chemical structure

In Papers **I**–**IV**, the 3D structures of drug substances were downloaded from the PubChem database [142,143]. Selected structures were in unionised, salt- and hydrate-free forms. To achieve a consistent 3D structural representation of all compounds, a conformational space search was performed. This was carried out with the MacroModel [144], as part of the Schrödinger Suite software package. The Merck Molecular Force Field (MMFFs) parameterisation [145,146,147, 148,149,150] and Monte-Carlo Multiple Minimum (MCMM) search method [151,152] were used for the conformational space analysis. The condensed media for the molecules was accounted for using a generalized Born/solvent-accessible surface area (GB/SA) solvation model [153]. Depending on the molecule, the number of steps (conformations) scanned was up to 15000. For each molecule, conformational search variables were set automatically and the conformer with the lowest energy was selected for the next steps. The geometry of the lowest energy conformer of each molecule was further optimised using semi-empirical quantum chemical methods. The AM1 (Austin Model 1) parameterisation [154] was used to characterize molecular structures, and the eigenvector following algorithm [155] was used for the geometry optimisation. Both methods were implemented in the MOPAC 7.0 program [156].

In Paper **V**, the 2D structures of drug substances were represented in SMILES (Simplified Molecular-Input Line-Entry System) [157] notation. For this the structures of compounds were standardised, dominant tautomeric forms of compounds were detected, and the aromatic SMILES were generated with the JChem for Excel [133].

3.6.2. Molecular descriptors

Molecular descriptors describe the chemical structure numerically and can be grouped into the constitutional, topological, geometrical, charge-distribution related, and quantum chemical molecular descriptors. The constitutional

descriptors are derived only from the molecular formula and do not give information about the geometrical and electronical structure. The topological descriptors mainly describe the connections between atoms and in some cases also information about the atomic constitution is included. The geometrical descriptors involve the 3D representation of the molecule and therefore expose the shape and size of the molecule. The charge-distribution related descriptors capture electrostatic interactions in-between the molecules. The quantum-chemical descriptors describe the valence and energy related properties of the molecules and also allow explicitly consider knowledge about the chemical reactivity in the prediction models. [138]

In Papers **I**–**IV**, the MOPAC (MOPAC 7.0 program [156]) calculations were performed for the structures from the conformational space analysis (Section 3.6.1) to gain the geometric, electronic, and energetic parameters for the molecules that were the source for the molecular descriptors. The CODESSA PRO 1.0 software [158,159] was used to calculate the constitutional, topological, geometrical, charge-distribution related, and quantum chemical molecular descriptors. The number of descriptors for each compound depended on their atomic constitution. The descriptors with missing values and the atom- and bond-specific descriptors were excluded from the set of descriptors.

In Paper **V**, the PaDEL-Descriptor software [160,161] was used to calculate 1D and 2D molecular descriptors from SMILES (Section 3.6.1) with "detect aromaticity" and "standardize nitro groups" settings. The descriptors with near zero variance, highly correlated descriptors (R=0.9999), autocorrelation, and atom and fragment specific descriptors were removed from the set of descriptors.

The logarithm of the octanol-water partition (logP_{ow}) and distribution (log$D$) coefficients have been added to the set of descriptors (Papers **I**–**V**), because they are often associated with the membrane permeability (Section 1.3.7). The logP_{ow} describes the partition properties between octanol and water for unionised compounds. Experimental logP_{ow} values were retrieved from the PhysProp database [162]. For the compounds without experimental logP_{ow} values, predicted logP_{ow} values were used. Different logP_{ow} calculators were evaluated for amphoteric (Paper **III**), neutral (Paper **III**), and 142 drug substances (Paper **V**). The logP_{ow} from the XlogP3 software (version 3.2.2) [163,164] resulted in satisfactory correlations in all these comparisons. In Papers **I**–**III**, the XlogP3 (or XlogP3-AA) [163,164] values from the PubChem database [142,143] were collected and in Papers **IV** and **V**, the XlogP3 values were calculated using the XlogP3 software (version 3.2.2) [163,164].

As the logP_{ow} characterises the properties of unionised compounds, the log$D$ should be more suitable for describing the membrane permeability at different pH-s, since it takes into account ionisation of drug substances. Experimental log$D$ values are rarely found in the literature and are almost non-existent in a wide range of pH. Thus, calculated log$D$ values were used instead of experimental ones. In Papers **II**–**V**, the log$D$ values for pH-s 3, 5, 7.4, and 9 were calculated with the JChem for Excel [133]. In Papers **III**–**V**, the highest

log$D$ (log$D_{highest}$) value for a compound corresponded to the pH where the membrane permeability was highest.

3.6.3. Training and validation sets

In Papers **I**–**IV**, the initial set of compounds was sorted for each series of data according to the experimental values (highest to lowest). Every fourth compound was selected for a validation set, and the remaining compounds formed a training set. The sizes of the training (n_{train}), validation (n_{val}), and external validation (n_{ext}) sets used in Papers **I**–**IV** are shown in Table 4.

Table 4. The sizes of the training, validation and external validation sets for Papers **I**–**IV**.

Paper	Training set	Validation set	External validation set
I	44	14	15
II: acids	28	8	1
II: bases	46	15	1
III: neutrals	12	3	1
III: ampholytes	46	14	1
IV	134	44	60

In Paper **V**, the division of compounds into the training and validation sets was adapted from Paper **IV**. For the regression analysis, order according to the experimental data ensured a similar distribution of the membrane permeability values in the training and validation sets. However in the case of classification, such training sets at pH 3, pH 5, log$P_{e_highest}$, and logP_o were unbalanced (Table 5). This influences the development and results of classification models [165,166,167], because the large variance in the number of compounds in any of the classes for the training set causes overly optimistic accuracy estimates due to the bias toward the majority class and poor predictive accuracy of the minority class.

Table 5. The number of compounds and the distribution of high- and low-permeable compounds in the initial set of compounds and final training, validation, and external validation sets within the data series for the classification.

Data series	Initial		Training			Validation			External		
	H	L	Total	H	L	Total	H	L	Total	H	L
pH 3	44	134	67	33	34	111	11	100	60	15	45
pH 5	58	120	89	44	45	89	14	75	60	21	39
pH 7.4	90	88	134	68	66	44	22	22	60	40	20
pH 9	90	88	134	68	66	44	22	22	60	40	20
log$P_{e_highest}$	110	68	107	56	51	71	54	17	60	45	15
logP_o	115	63	91	44	47	87	71	16	60	47	13

Therefore, before the development of the classification models, the training sets from Paper **IV** were balanced. The distribution of high- and low-permeable

compounds in the initial set of compounds (Table 5) were used for the selection of appropriate ratios for making a balanced training set. The selection was made while keeping the same order based on the experimental values as it was in the original training sets (Paper **IV**). For example (Figure 9), the initial set of compounds at pH 3 included 134 low- and 44 high-permeable compounds (Table 5), i.e. three-times more low- than high-permeable compounds. This means that from the original training set (Paper **IV**), every second and third low-permeable compound was moved to the validation set and every first low-permeable compound together with high-permeable compounds form the new training set (Paper **V**).

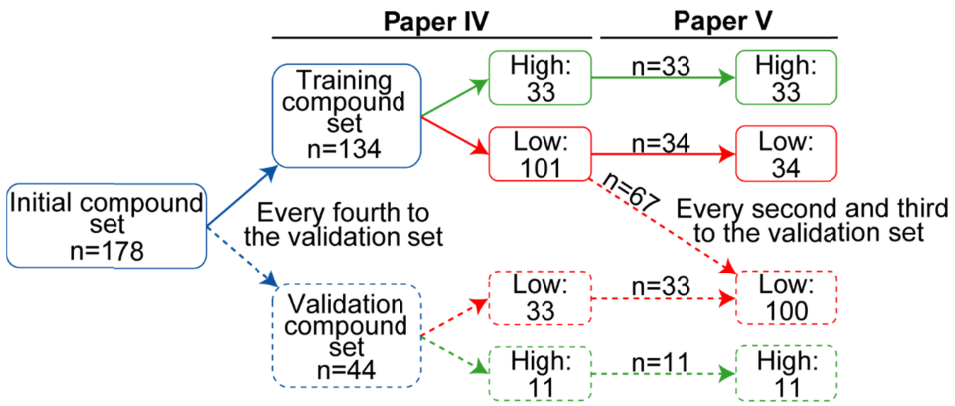


Figure 9. The formation of training and validation sets in Paper **V** using the training and validation sets from Paper **IV**: example for the data series at pH 3 (n – number of compounds).

For pH 5 every second low-permeable compound in the original training set (Paper **IV**) was selected to the balanced training set and the remaining low-permeable compounds were moved to the validation set. For the highest membrane permeability every third and for the intrinsic membrane permeability every second high-permeable compound from the original training set was moved to the validation set and the remaining compounds formed the balanced training set. Final sizes of the training and validation sets with the distribution of high- and low-permeable compounds are shown in Table 5.

3.6.4. Methods for the regression models

Quantitative structure-activity relationship (QSAR) is obtained using (multiple) linear least squares regression method between experimental property ($\log P_e$) and molecular descriptors (D_1, D_2, D_3 …) with weights (α_1, α_2, α_3, ...) for each descriptor and the intercept (b) (Eq. (24)) [138].

$$\log P_e = b + \alpha_1 \cdot D_1 + \alpha_2 \cdot D_2 + \alpha_3 \cdot D_3 + \ldots \quad (24)$$

The QSAR models (Papers **I**–**IV**) were developed using methods implemented in the CODESSA PRO 1.0 software [158,159]. The heuristic stepwise forward selection method was used to develop one-parameter models. The best multiple linear regression (BMLR) approach was used to develop multi-parameter regression models. The BMLR follows a stepwise forward selection algorithm to add significant descriptors to the multiple linear model. During the BMLR procedure, the set of descriptors is cleaned of non-significant descriptors ($R^2 < 0.1$) and descriptors with missing values. The descriptor scales are normalised and centred automatically, and the result is given in natural scales. The selection of models is made based on the statistical significance and non-collinearity criteria ($R^2<0.6$) of the selected descriptors.

The **quality of the models** was assessed by the coefficient of determination for the training set (R^2_{train}), the cross-validated (leave-one-out) coefficient of determination (R^2_{cv}), the squared standard error of the estimate (s^2), the coefficient of determination for the validation set (R^2_{val}) and the coefficient of determination for the external validation set (R^2_{ext}). [138]

The **applicability domain** of the regression models was assessed and analysed using the Williams plot [168,169]. This plot is based on the leverage (h_i) and the standardised prediction error (r') of each compound and generates a graphical representation of structurally different and statistically deviating compounds. Compounds with leverage values larger than the critical leverage value (h*) can be considered structurally different and thus may result in unreliable predictions [170,171]. Standardised residuals indicate moderate outliers (±2r', 95% confidence limit) and strong outliers (±3r', 98% confidence limit) according to the Gaussian distribution.

3.6.5. Methods for the classification models

Two types of classification approaches have been used: the classification of predictions from the regression models (Paper **IV**) and the classification of high- and low-permeable compounds using the logistic regression method (Paper **V**).

The cutoff classification (Paper **IV**) is based on a match between predicted membrane permeability from the QSAR models (Section 3.6.4) and experimental membrane permeability, which both are converted to the classes of high and low permeable using the cutoff value for the membrane permeability (Section 3.3).

In Paper **V**, the classification models for the membrane permeability were developed with the logistic regression using the *glm* function in the R software (version 3.3.2) [172]. The logistic regression [173] is a statistical method for the binary classification problems, where the dependent variable is the probability of a binary event, calculated as the logit function of the linear combination of the independent variables. The probability (P) of outcome (Eq. (25)) being high ($P \leq 0.5$) or low ($P > 0.5$) permeable is calculated using the intercept (b) and the

coefficients (a_1, a_2, … a_n) of the variables (X_1, X_2, … X_n) in the logistic regression (z).

$$P = \frac{e^{(b + a_1 \cdot X_1 + a_2 \cdot X_2 \ldots a_n \cdot X_n)}}{1 + e^{(b + a_1 \cdot X_1 + a_2 \cdot X_2 \ldots a_n \cdot X_n)}} = \frac{e^z}{1 + e^z} \tag{25}$$

"Positive" (class 0, i.e. $P < 0.5$) was defined as high-permeable compounds and "negative" (class 1, i.e. $P > 0.5$) was defined as low-permeable compounds. The logistic regression was selected because the method's ability to estimate probability of the predicted response and due to its simple mathematical representation, which makes the models easily usable.

The **confusion matrix** [174] is used to analyse classification results (Papers **IV** and **V**). The confusion matrix (Table 6) groups compounds based on experimental and predicted classes into four groups: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Each row in the confusion matrix shows how many compounds are in the specific class according to the predicted classes and each column shows how many compounds are in the specific class according to the experimental classes.

Table 6. The confusion matrix and the performance characteristics calculated from the confusion matrix elements.

Experimental / **Predicted**	**High permeability**	**Low permeability**
High permeability	True positive (TP)	False positive (FP)
Low permeability	False negative (FN)	True negative (TN)
	Sensitivity=TP/(TP+FN)	**Specificity**=TN/(TN+FP)
	Accuracy=(TP+TN)/(TP+TN+FP+FN)	

The model performance was described with the sensitivity, specificity, and accuracy (Table 6). The sensitivity measures the proportion of correctly predicted high-permeable compounds. The specificity measures the proportion of correctly predicted low-permeable compounds. The accuracy measures the proportion of correctly predicted responses by the model.

3.6.6. Decision tree

The use of a single logistic classification model to predict the membrane permeability classes can be inaccurate, because one model can misclassify the compound. Since pH influences the permeability in the GIT, the logistic classification models were combined into a simple decision tree and tested to predict the BCS permeability classes with the U.S. FDA reference drug substances (Section 3.5).

The decision tree is a method that is based on the decision(s) to classify compounds based on certain criterion(s). The decision tree's criterion in Paper **V** was the number of high permeability predictions over six models

(#HighPrediction). To find the threshold for the criterion the following rules were considered:

- The decision cannot be made only on the basis of one or two high permeability predictions;
- The decision cannot be made based on only high membrane permeability predictions of the $\log P_{e\_highest}$ and/or $\log P_o$ models.

According to these rules, the #HighPrediction value must be greater than two (#HighPrediction > 2) in order to classify compound as a high permeable. The predicted class from the decision tree was compared to the BCS permeability class, and the match was described with the accuracy, sensitivity, and specificity (Section 3.6.5).

4. RESULTS AND DISCUSSION

4.1. Analysis of the experimental membrane permeability

The membrane permeability is usually measured at a single time point and one pH. However, to calculate the membrane permeability with high accuracy, the information about the time dependence is required. In order to draw the right conclusions about the absorption of drug substances in the GIT, the pH dependence of the membrane permeability must be considered. Therefore, the time and pH dependence of the membrane permeability for 274 drug substances was measured and investigated (Papers **I**–**IV**).

4.1.1. Time- and pH-dependent membrane permeability

Experimental measurement results of basic compound *Propranolol* are used to explain the time and pH dependence of the membrane permeability (Figure 10). It can be observed that the time dependence in the acceptor section is different for selected pH-s in the donor section (Figure 10: a). For acidic pH-s (3 and 5), the concentration in the acceptor section is not changing remarkably over 48 hours, but for pH 7.4 and 9, a noticeable time dependence exists. The time dependence for pH 7.4 and 9 shows that the system reaches to the equilibrium (i.e. concentration is not changing anymore) before 48 hours and membrane permeability values are calculated before the equilibrium. Time dependence for pH 3 and 5 shows the importance to measure membrane permeability for a longer period (48 hours) to be able to calculate the membrane permeability values also for low-permeable compounds. Based on the time dependence, membrane permeability values were calculated at 48 hours for pH-s 3 and 5 and at 4 hours for pH 9 using the gradient-pH equation (Eq. (22)) and at 8 hours for pH 7.4 using the iso-pH equation (Eq. (14)).

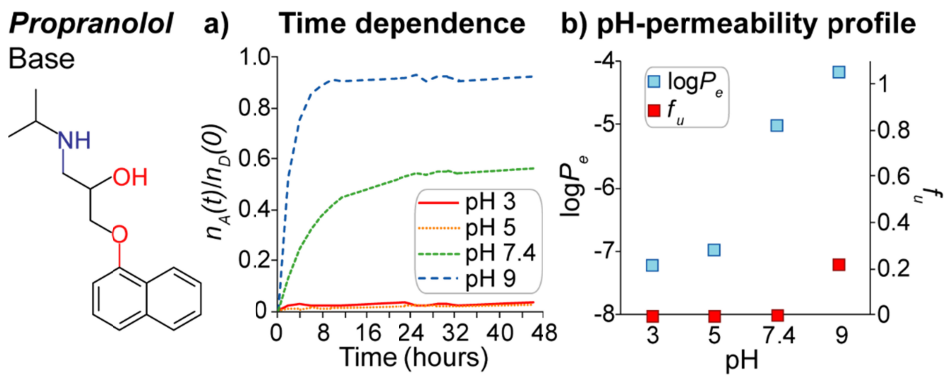


Figure 10. The time and pH dependence of the membrane permeability for *Propranolol*: (a) time vs. the number of moles in the acceptor section at a certain time point ($n_A(t)$) divided with the initial number of moles in the donor section ($n_D(0)$) and (b) pH vs. logP_e and pH vs. the fraction of unionised species (f_u).

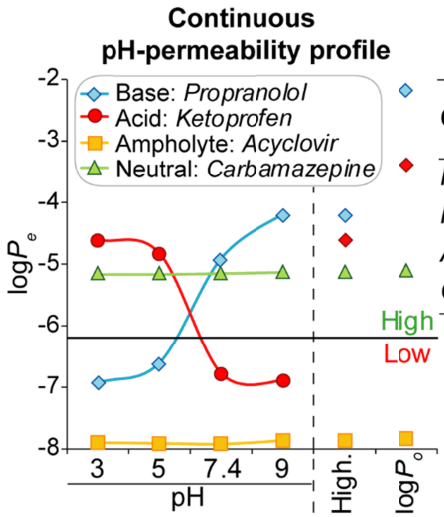
Propranolol has a very noticeable pH dependence for the membrane permeability and the f_u (Figure 10: b). *Propranolol* is practically fully ionised at acidic pH-s, where experimental membrane permeability values are similarly low, although these values should be different. This is due to the fact that the small amount of ionised species can permeate cross the membrane ($\log P_i$, Section 3.2) and the membrane permeability values are influenced by the concentrations close to the limit of detection in the acceptor section. An increase in the membrane permeability at pH 7.4 shows that very small increase of the f_u compared to acidic pH-s can drastically influence the membrane permeability. The highest membrane permeability ($\log P_{e\_highest} = -4.2$) is measured at pH 9, where around 20% of species are unionised. This shows that the membrane permeability is highly influenced by the fraction of unionised species, which vary significantly for ionisable compounds over the pH range in the GIT.

The PAMPA measurements are usually performed using only one time point, although it is important to calculate the membrane permeability before the equilibrium. This makes the time dependence measurements important in order to precisely calculate the membrane permeability values for both high- and low-permeable compounds. Also analysing only one pH, which is usually neutral pH, does not give enough information about the pH dependence of the membrane permeability, which is needed for the estimation of absorption in the GIT.

4.1.2. pH-permeability profiles for chemical classes

It is evident that the pH-permeability profiles vary in different chemical classes. To compare the profiles of different chemical classes, four U.S. FDA reference drug substances for the BCS permeability classes [10] were selected (Figure 11): *Propranolol*, *Ketoprofen*, *Acyclovir*, and *Carbamazepine*. The pH-permeability profiles can be both continuous and categorical (Figure 11). For the second case experimental membrane permeability data were converted to the corresponding high and low permeability classes using the cutoff value of $\log P_e = -6.2$ (Section 3.3). The conversion allows to compare the BCS permeability classes with the experimental membrane permeability classes. The comparison indicates if the membrane permeability measured by the PAMPA can be used to predict the BCS permeability classes.

As already shown for *Propranolol* (Section 4.1.1), the membrane permeability for **basic drug substances** significantly depends on pH. The basic compound (Figure 11) has a higher membrane permeability at basic pH-s (pH 9) and a lower membrane permeability at acidic pH-s. Indeed, *Propranolol* is assigned as high permeable by the BCS and therefore the membrane permeability at pH 7.4 and 9 must be considered. This shows that basic compounds permeate at basic pH-s and in the GIT regions where neutral and basic pH-s are present.



Categorical pH-permeability profile

Compounds	BCS	pH 3	pH 5	pH 7.4	pH 9	High.	$\log P_o$
Propranolol	H	L	L	H	H	H	H
Ketoprofen	H	H	H	L	L	H	H
Acyclovir	L	L	L	L	L	L	L
Carbamazepine	H	H	H	H	H	H	H

Figure 11. The continuous and categorical pH-permeability profiles (H – High, L – low, High. – $\log P_{e\_highest}$) and comparison with the BCS permeability classes.

Acidic drug substances have higher membrane permeability values for acidic pH-s (pH 3 and 5) and lower membrane permeability values for basic pH-s, like *Ketoprofen* (Figure 11). Notably, *Ketoprofen*'s membrane permeability at acidic pH values is influenced by the ABL, i.e. membrane permeability values are similar. *Ketoprofen* is assigned as high permeable by the BCS and it is correctly classified based on the membrane permeability data only on acidic pH-s. This shows that acidic drug substances permeate across the intestinal epithelium at acidic conditions. This can occur despite the fact that the area of the region in the GIT with pH ~3 is rare.

Amphoteric drug substances possess both acidic and basic properties and therefore form a complex chemical class. The membrane permeability and the fraction of unionised species for amphoteric compounds are highly influenced by the pH in the donor and acceptor sections and pK_a values of the molecule. Most amphoteric compounds do not exist in a fully unionised form in the GIT and because of this their membrane permeability values are typically lower than for other chemical classes. For example, *Acyclovir* is a low-permeable compound with minimal pH dependence of the membrane permeability. The BCS assigns *Acyclovir* as a low-permeable compound, which is coherent with the experimental pH-permeability profile.

Neutral drug substances are unionised in the range of pH from 3 to 9 and therefore the membrane permeability does not depend on the pH. The example compound *Carbamazepine* (Figure 11) is classified as high-permeable compound ($\log P_e \geq -6.2$) for all pH-s over the pH-permeability profile. Indeed, the BCS assigns *Carbamazepine* as a high-permeable compound and the same result was obtained from the membrane permeability measurements.

From the above, it can be concluded that the compounds in different chemical classes do have characteristic pH-permeability profiles. The maximum membrane permeability can be at different pH values depending on the chemical class. The comparison between the BCS permeability classes and the experimental class-based pH-permeability profiles shows that the membrane permeability measured at the correct pH (i.e. highest membrane permeability) matches well with the BCS permeability classes (Figure 11). This indicates that using the highest membrane permeability value over selected pH-s ($\log P_{e\_highest}$) is suitable property to estimate absorption in the GIT.

The membrane permeability without ionisation is described with the intrinsic membrane permeability ($\log P_o$), which characterises the maximum membrane permeability for the compound. Both $\log P_{e\_highest}$ and $\log P_o$ correspond correctly to the BCS permeability classes for all four U.S. FDA reference compounds (Figure 11). The comparison of the $\log P_{e\_highest}$ and the $\log P_o$ for *Ketoprofen* and *Propranolol* illustrates the difference between the two values, showing that the $\log P_o$ values are more than one unit higher compared to the $\log P_{e\_highest}$. The reason for the difference is that the $\log P_o$ does not include the ABL and the ionisation of the compound, which both decrease the membrane permeability value. Four U.S. FDA reference compounds show that the membrane permeability measured by the PAMPA over a wide pH range can be used to estimate the BCS permeability classes.

4.2. Membrane permeability vs human intestinal absorption

Four U.S. FDA reference compounds show an exceptionally good match between the BCS permeability class and the membrane permeability measured by the PAMPA (Section 4.1.2). A wider comparison of the %HIA and measured membrane permeability was performed using 138 compounds (Paper **IV**) [134]. Most of these compounds are highly absorbed and include both passively and actively transported compounds. For this comparison, the cutoff value for the %HIA was 85% according to the BCS [10,27] and the cutoff value for the membrane permeability (Paper **IV**) was $\log P_e = -6.20$ according to the match between the %HIA and the highest membrane permeability (Figure 12: a, Section 3.3).

The prediction of high absorbed compounds (%HIA ≥ 85%, Figure 12: b,c) is influenced by the pH and chemical classes. The number of correctly classified high absorbed compounds increases when the pH increases (Figure 12: b). This is because of basic compounds, which have higher permeability at basic pH-s. Deeper analysis reveals that the estimation capability of high absorbed acidic compounds also depends on pH: the number of correctly classified high absorbed compounds decrease when the pH increases. The match between the membrane permeability and the %HIA for high absorbed amphoteric and neutral compounds does not significantly depend on pH, but they reveal the limitation of describing certain compounds with the PAMPA system. For

amphoteric compounds, most misclassified compounds (Figure 12: c) are sulphonamides and fluoroquinolones and for neutral compounds steroids. The best match for high absorbed compounds was obtained for the log$P_{e_highest}$ and the logP_o, where more than 80 high-permeable compounds corresponds to the high absorbed compounds (i.e. correctly classified) and less than 30 high-permeable compounds are indicated to be low absorbed (i.e. misclassified). This indicates that it is important to consider the correct pH for all chemical classes, especially for basic and acidic compounds in order to improve the estimation of high absorbed compounds in the GIT. The reason of mismatch for 30 compounds (Figure 12: c) can be either the active transport, which is not described by the PAMPA system, or the underestimated membrane permeability by the PAMPA system.

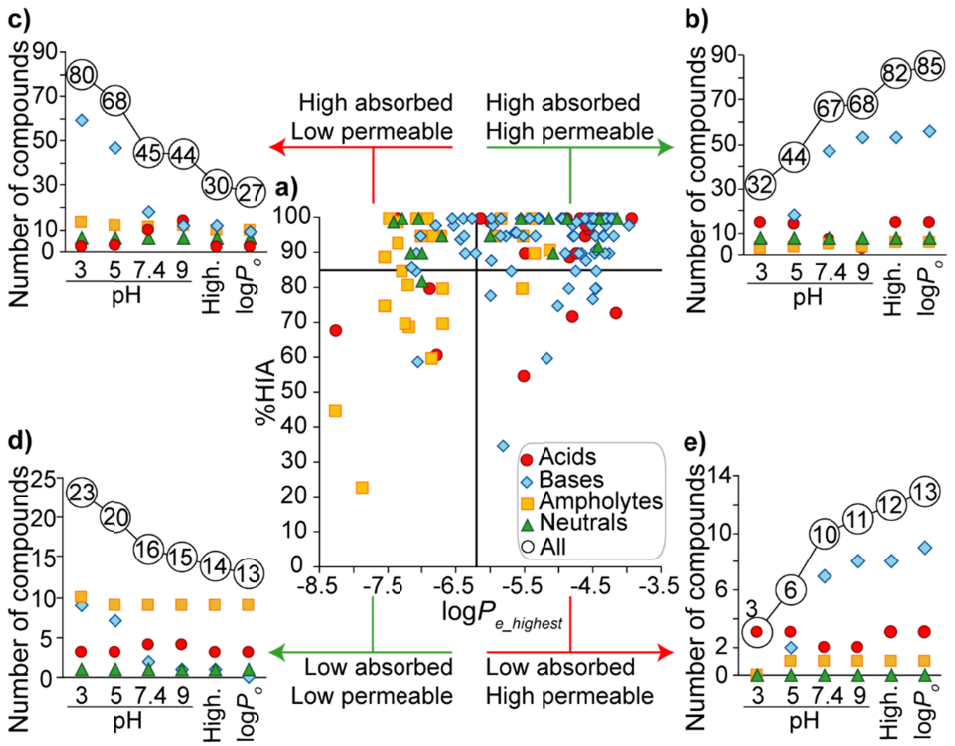


Figure 12. The match between the human intestinal absorption and the membrane permeability: (a) %HIA vs. log$P_{e_highest}$ (High.) based on chemical classes and (b–e) the number of compounds in each quarter depending on the data series and chemical classes.

The number of low absorbed compounds (%HIA < 85%, Figure 12: d,e) in the analysed set of compounds was significantly lower than the number of high absorbed compounds. Low absorbed compounds mainly match correctly at specific pH-s (Figure 12: d), which can be explained by low solubility at the corresponding pH to the highest membrane permeability. For example at pH 3,

only three low absorbed acidic compounds are defined as high permeable (Figure 12: e), while at pH 9 eleven compounds are mismatched, mostly basic compounds. This is also shown with the $\log P_{e\_highest}$ and the $\log P_o$, where ~13 compounds have correct match and ~13 compounds are mismatched between the membrane permeability and the %HIA. The compound may be highly permeable, but if solubility is low at specific pH, then compound's absorption may not be high in the GIT. This directs towards the simultaneous analysis of the pH-permeability and the pH-solubility profiles to spot the optimal pH for the absorption [31]. Amphoteric and neutral compounds show very good agreement between the membrane permeability and the %HIA for every data series. Another reason for the mismatch of low absorbed compounds is an efflux, which is not described by the PAMPA system.

From the above, one can conclude that the membrane permeability measured by the PAMPA is comparable to the %HIA when considering the full pH range in the GIT. Most problematic are compounds with the limited solubility and the active transport and some specific compound types. The comparison of the PAMPA values of different chemical classes at various pH-s shows the importance of pH to the %HIA and the clear necessity to measure, analyse, and predict the pH-permeability profiles for drug substance candidates to reveal the possible absorption in the GIT.

4.3. QSAR models for chemical classes

According to the pH-permeability profiles (Section 4.1.2) and to the comparison with the human intestinal absorption (Section 4.2), the membrane permeability is largely influenced by the chemical composition of molecules, such as ionisable functional groups. The QSAR models give information about which structural parameters influence the membrane permeability for chemical classes, i.e. neutral, acidic, basic, and amphoteric compounds. For all four chemical classes the QSAR models were developed using three sets of descriptors in order to find out which of them gives the best descriptive and predictive models for the membrane permeability. The sets of descriptors are:

- The logarithm of the octanol-water partition coefficient ($\log P_{ow}$), which describes the partition properties for unionised compounds,
- The logarithm of the octanol-water distribution coefficient at a certain pH ($\log D$), which describes the distribution properties and takes account ionisation,
- A set of molecular descriptors from the CODESSA PRO, which includes the constitutional, topological, geometrical, charge-distribution related, and quantum chemical molecular descriptors, including the $\log P_{ow}$ and the $\log D$.

All developed models are externally validated with one new compound.

4.3.1. QSAR models for neutral drug substances

The membrane permeability for 15 neutral drug substances (Paper **III**) was measured and analysed. The compounds were divided to the training (12 compounds) and validation (3 compounds) sets (Paper **III**: Table 1). The comparison of the correlations with the sets of descriptors (Figure 13: a) indicates that the logP_{ow} or the log$D$ are not suitable descriptors for describing the membrane permeability of neutral compounds ($R^2 \sim 0.2$). Best correlations for all pH-s ($R^2 \sim 0.95$) were obtained with the hydrogen bond donor descriptor: the area-weighted surface charge of hydrogen bonding donor atoms (HA-dependent *HDCA2* (Zefirov PC)) [159].

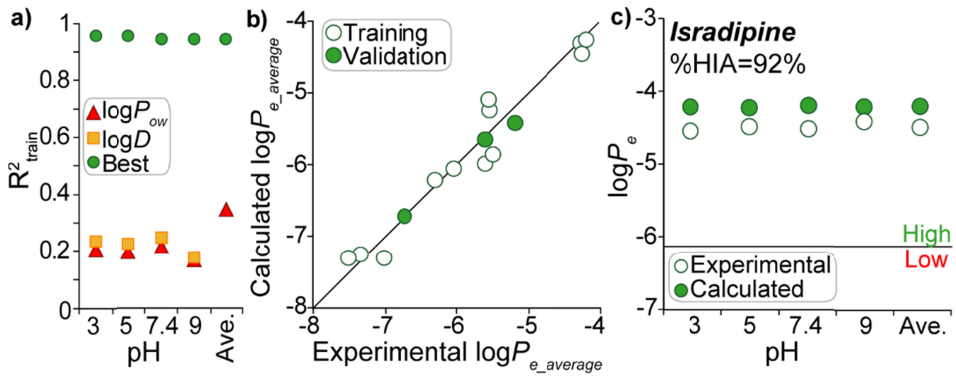


Figure 13. Summary of the QSAR models for neutral compounds: (a) the comparison of the models between the sets of descriptors, (b) the relationship between experimental and calculated membrane permeability for the log$P_{e_average}$ (Ave.), and (c) experimental and predicted membrane permeability data for *Isradipine*.

The *HDCA2* is negatively correlated with the membrane permeability (Table 7), meaning that the molecule, which has a surface area mainly covered by the hydrogen bond donors, has a lower membrane permeability. This occurs because hydrogen bonds can be formed with the acceptor groups of lecithin and water, restricting the movement of a molecule through the membrane. The models for all pH-s have a good descriptive quality for the training set ($R^2_{train} \sim 0.95$) and predictive quality for the validation set ($R^2_{val} > 0.95$). Comparable results for all pH-s based on the training and validation sets indicate that in all cases the mechanism of interaction is similar. This implicates that the average of membrane permeability over pH-s (12 data points, log$P_{e_average}$,) can be considered more accurate value for neutral compounds than that for specific pH-s. The model for the log$P_{e_average}$ has similar description and prediction capability compared to the pH-specific models (Table 7) and it can be considered also as the most accurate model. For the log$P_{e_average}$, the comparison between experimental and predicted values shows excellent match for the training and validation sets (Figure 13: b). The applicability domain was

analysed for all models using the Williams plot (Paper **III**: Figure 3, Figure S1), which reveals no outliers for the training and validation sets.

Table 7. The QSAR models of the membrane permeability for neutral compounds (n_{train}=12, n_{val}=3) with the performance characteristics (Paper **III**: Figure 3, Figure S1).

pH	Prediction model	R^2_{train}	R^2_{cv}	s^2	R^2_{val}
3	$\log P_{e\_pH3} = -3.71(\pm0.15) - 2.63(\pm0.17)\cdot HDCA2$	0.96	0.94	0.06	0.96
5	$\log P_{e\_pH5} = -3.72(\pm0.16) - 2.61(\pm0.18)\cdot HDCA2$	0.96	0.94	0.07	0.95
7.4	$\log P_{e\_pH7.4} = -3.70(\pm0.16) - 2.55(\pm0.18)\cdot HDCA2$	0.95	0.93	0.07	0.97
9	$\log P_{e\_pH9} = -3.73(\pm0.16) - 2.50(\pm0.18)\cdot HDCA2$	0.95	0.93	0.07	0.98
Average	$\log P_{e\_average} = -3.71(\pm0.17) - 2.56(\pm0.19)\cdot HDCA2$	0.95	0.93	0.07	0.99

Isradipine (Figure 13: c) was selected as a control compound to externally validate the developed models for neutral drug substances (Table 7). The predicted membrane permeability (around –4.2) is slightly higher compared to the experimental membrane permeability (around –4.5) for all data series. Experimental membrane permeability can be influenced by the ABL, which decreases the membrane permeability. *Isradipine* is classified as a high absorbed compound based on the human intestinal absorption data (%HIA = 92% [134]), which is coherent with the experimental and predicted membrane permeability for all data series.

4.3.2. QSAR models for acidic drug substances

The membrane permeability values for 36 acidic drug substances have been measured and divided to the training (28 compounds) and validation (8 compounds) sets (Paper **II:** Table 1). The comparison of correlations for the logP_{ow} and the log$D$ shows that the log$D$ has slightly higher correlation with the membrane permeability for all data series in comparison with the logP_{ow} (Figure 14: a). This is because the log$D$ takes into account the ionisation state of the molecule, which in turn influences the membrane permeability. An alternative and more significant descriptor than the log$D$ was selected from a set of molecular descriptors, which relates the membrane permeability of acidic compounds with the hydrogen bond donor properties: the area-weighted surface charge of hydrogen bonding donor atoms divided by the total molecular surface area (HA-dependent *HDCA-2/TMSA* (MOPAC PC) (all)). For all sets of descriptors (Figure 14: a), the determination coefficients for the models at pH 3 and 5 are higher compared to pH 7.4 and 9. This is connected to the distribution of high- and low-permeable compounds in the data series, where the number of high-permeable compounds decreases when pH increases.

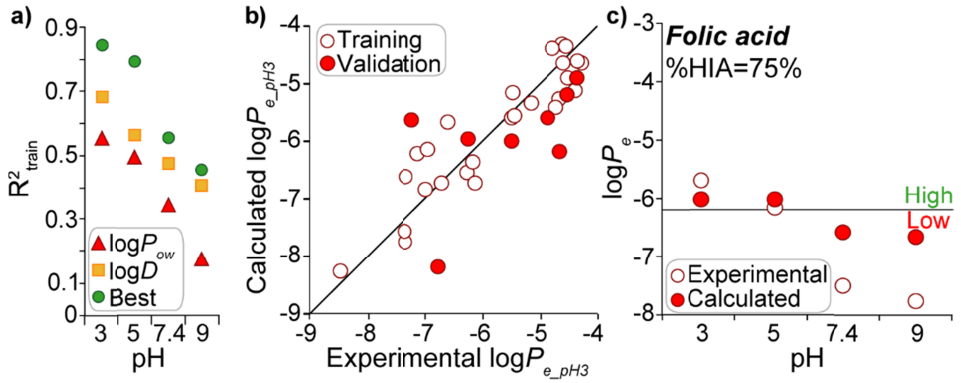


Figure 14. Summary of the QSAR models for acidic compounds: (a) the comparison of the models based on the logP_{ow}, the log$D$ and the set of theoretical molecular descriptors, (b) the correlation between experimental and calculated value for the model at pH 3, and (c) experimental and predicted membrane permeability data for *Folic acid*.

The *HDCA-2/TMSA* [159] has a negative correlation with the membrane permeability (Table 8), which means that compound has a lower membrane permeability when the molecular surface area is mainly covered with positively charged hydrogen bond donors. Consequently, such molecules form hydrogen bonds with the membrane, and this in turn restricts the movement of the molecule through the membrane. The model at pH 3 (Table 8) has the highest descriptive properties ($R^2_{train} = 0.85$), while the outliers have a remarkable influence on the prediction ability ($R^2_{val} = 0.31$). The correlation with experimental and predicted membrane permeability at pH 3 (Figure 14: b) shows three significantly outliers from the validation set that is also detected from the Williams plot (Paper **II**: Figure 2). The model at pH 5 has good descriptive ($R^2_{train} = 0.8$) and slightly lower predictive properties ($R^2_{val} = 0.64$), which is influenced by one outlier. The reason of outliers could be caused by the experimental error and the limited solubility at pH-s 3 and 5. One compound was detected with the high leverage value on the Williams plot (Paper **II**: Figure 2), which has an extreme *HDCA-2/TMSA* value. These aspects show that the models at pH-s 3 and 5 are suitable for predicting the membrane permeability of acidic compounds, but the models at pH 7.4 and 9 have low descriptive quality ($R^2_{train} < 0.6$) and are not suitable for the prediction purposes.

Table 8. The QSAR models of the membrane permeability for acidic compounds (n_{train}=28, n_{val}=8) with the performance characteristics (Paper **II**: Figure 2, Figure S1).

pH	Prediction model	R^2_{train}	R^2_{cv}	s^2	R^2_{val}
3	$\log P_{e\_pH3} = -3.1815 - 365.40 \cdot HDCA\text{-}2/TMSA$	0.85	0.83	0.25	0.31
5	$\log P_{e\_pH5} = -3.4838 - 326.23 \cdot HDCA\text{-}2/TMSA$	0.80	0.77	0.30	0.64
7.4	$\log P_{e\_pH7.4} = -4.5841 - 256.97 \cdot HDCA\text{-}2/TMSA$	0.56	0.5	0.54	*
9	$\log P_{e\_pH9} = -4.8501 - 233.18 \cdot HDCA\text{-}2/TMSA$	0.46	0.38	0.63	*

* Not suitable for the prediction, because R^2 for the training set is less than 0.6.

The models for acidic compounds (Table 8) were tested with *Folic acid*. The predicted and experimental values for *Folic acid* (Figure 14: c) are similar at pH-s 3 and 5, whereas remarkable differences can be observed at pH 7.4 and 9. This allows to conclude that the prediction models at pH 3 and 5 are more accurate and usable than at pH 7.4 and 9, which have low descriptive properties ($R^2_{train} < 0.6$). *Folic acid* is classified as a medium absorbed compound in the GIT (%HIA = 75% [175]). Based on the membrane permeability data, *Folic acid* is high-permeable compound, however the membrane permeability values for acidic pH-s ($\log P_{e\_pH3} = -6$) are very near to the cutoff ($\log P_e = -6.2$). This may indicate overestimation of the PAMPA membrane permeability that can be caused by the low solubility of the compound at acidic pH-s.

4.3.3. QSAR models for basic drug substances

The set of 61 basic compounds with the membrane permeability values at four pH-s (Paper **II**: Table 1) was divided into the training (46 compounds) and validation (15 compounds) sets. The comparison of the correlations between the $\log P_{ow}$ and the $\log D$ (Figure 15: a) shows that the $\log D$ has significantly higher correlation for pH-s 5, 7.4 and 9. For pH 3, no improvement for the correlation is observed with the $\log D$ ($R^2_{train} < 0.05$). The difference of the determination coefficients for the data series is related to the distribution of high- and low-permeable compounds in the training set, which is influenced by the fraction of unionised species. A larger number of high-permeable compounds are present at basic pH-s, where also correlations are stronger ($R^2_{train} > 0.7$). Reversely to basic pH-s, a large number of low-permeable compounds exist at acidic pH-s where obtained correlations are insignificant ($R^2_{train} < 0.3$). After the analysis of the large set of molecular descriptors, the $\log D$ remained as the best descriptor for all pH values and also the second descriptor did not improve the models remarkably. This means that the $\log D$ is the most significant descriptor for the modelling of membrane permeability for basic compounds.

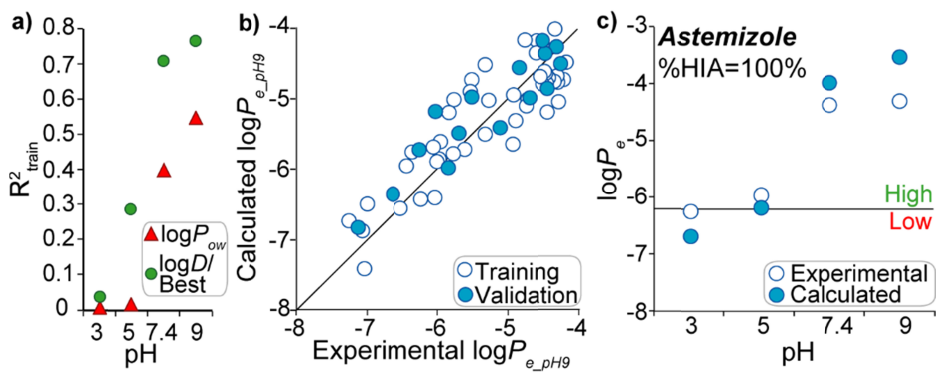


Figure 15. Summary of the QSAR models for basic compounds: (a) the comparison of the models with different sets of descriptors, (b) the relationship between experimental and calculated membrane permeability at pH 9, and (c) experimental and predicted membrane permeability data for *Astemizole*.

Indeed, the log*D* is positively correlated with the membrane permeability (Table 9), showing that more hydrophobic basic compounds have also higher membrane permeability. This indicates that the permeation between membrane and buffer for basic compounds follows the same mechanism as the distribution between octanol and buffer. The model at pH 9 has the highest descriptive (R^2_{train} = 0.76) and predictive (R^2_{val} = 0.85) properties. The relationship between experimental and calculated membrane permeability shows good agreement (Figure 15: b) and no outliers were found according to the Williams plot (Paper **II**: Figure 5). The model at pH 7.4 has slightly lower descriptive (R^2_{train} = 0.70) and predictive (R^2_{val} = 0.72) properties. One outlier for the training set of pH 7.4 was identified on the Williams plot (Paper **II**: Figure 5), which can most likely be attributed to the precision of calculated log*D*. One structurally different point was detected in the training set of pH 7.4 and 9 (Paper **II**: Figure 5), which has the lowest log*D* value over the training and validation sets. The developed models at pH 7.4 and 9 allow predicting the membrane permeability values. The models at pH 3 and 5 show very low correlations for the training set (R^2_{train} < 0.3) and therefore expectedly cannot give reliable predictions.

Table 9. The prediction models of the membrane permeability for basic compounds (n_{train}=46, n_{val}=15) with the performance characteristics (Paper **II**: Figure 5, Figure S2).

pH	Prediction model	R^2_{train}	R^2_{cv}	s^2	R^2_{val}
3	$\log P_{e\_pH3} = -6.7235 + 0.0845 \cdot \log D_{pH3}$	0.03	0.01	0.43	*
5	$\log P_{e\_pH5} = -6.4916 + 0.3326 \cdot \log D_{pH5}$	0.28	0.21	0.49	*
7.4	$\log P_{e\_pH7.4} = -6.1327 + 0.5402 \cdot \log D_{pH7.4}$	0.70	0.68	0.26	0.72
9	$\log P_{e\_pH9} = -6.2507 + 0.5232 \cdot \log D_{pH9}$	0.76	0.74	0.21	0.85

* Not suitable for the prediction, because R^2 for the training set is less than 0.6.

Astemizole was selected as an external test compound to assess the predictive capacity of the models for basic compounds (Figure 15: c). The comparison of experimental and calculated values revealed good predictions for pH 3, 5, and 7.4, but significant difference for pH 9. Interestingly, the predictions for *Astemizole* at pH 3 and 5 are acceptable, even though the statistical parameters for the models at pH 3 and 5 were insignificant (R^2_{train} < 0.3). The predicted membrane permeability value at pH 9 is higher than the experimental value, because the experimental value is influenced by the ABL. *Astemizole* is a high absorbed compound in the GIT (%HIA=100% [175]), which is coherent with the membrane permeability values at pH 7.4 and 9, where *Astemizole* has been shown as a high-permeable compound. This indicates that *Astemizole* will be absorbed at the conditions and in the regions of the GIT, where pH is higher.

4.3.4. QSAR models for amphoteric drug substances

The set of 60 amphoteric drug substances (Paper **III**: Table 2) was divided into the training (46 compounds) and validation (14 compounds) sets. Amphoteric compounds are structurally complex, which makes them more difficult to model and predict in comparison with neutral, acidic, or basic compounds. The highest

membrane permeability for amphoteric compounds can be at any pH, therefore the highest membrane permeability over selected pH-s ($\log P_{e\_highest}$) was added to the data series for amphoteric compounds.

The comparison of the correlations with the $\log P_{ow}$ and the $\log D$ (Figure 16: a) shows that the $\log P_{ow}$ has a significantly higher correlation compared to the $\log D$, although the $\log D$ should also take into account pH and the fraction of unionised species. This is very likely a result of the unreliable calculation of the $\log D$ values for amphoteric compounds. The correlation with the $\log P_{ow}$ was improved with a combination of two descriptors, which describe the polarity and/or the size and shape of the molecule. The determination coefficient of the models with different sets of descriptors reduces in the row pH 5 > pH 3 > $\log P_{e\_highest}$ > pH 7.4 > pH 9. One plausible explanation for such behaviour may be a wider range of membrane permeability values at pH 5 and 3 and the $\log P_{e\_highest}$ in comparison with the pH 7.4 and 9.

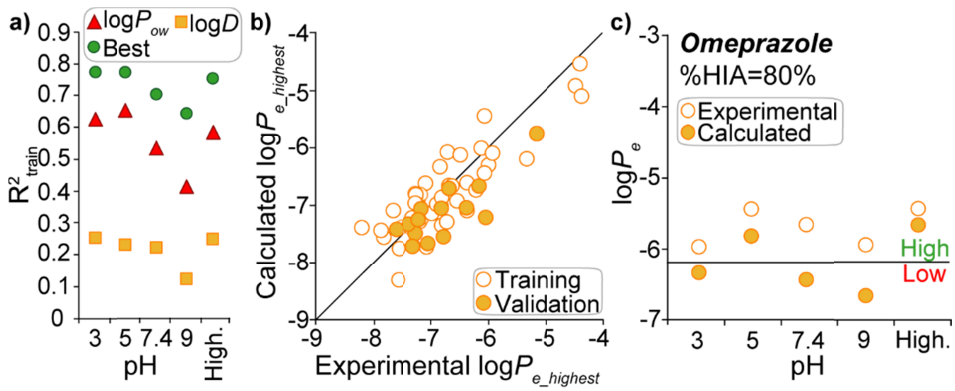


Figure 16. Summary of the QSAR models for amphoteric compounds: (a) the comparison of the models with different sets of descriptors, (b) the relationship between experimental and calculated highest membrane permeability (High. – $\log P_{e\_highest}$), and (c) experimental and predicted membrane permeability data for *Omeprazole*.

The selected models (Table 10) have good descriptive capability of the training set ($R^2_{train} > 0.7$), except the model at pH 9. The validation set has significantly lower statistical parameters compared to the training set of the pH-specific models, which is caused by the outliers (Paper **III**: Figure S2). Excluding strong outliers markedly improves the validation set statistical parameters for all data series (Table 10). The model for $\log P_{e\_highest}$ has most similar statistical parameters for the training and validation sets compared to the models for specific pH. The correlation between experimental and calculated $\log P_{e\_highest}$ (Figure 16: b) reveals that there is no strong outliers, while most of the compounds are low-permeable in the validation set. Thus, low validation set statistics can be related to the fact that the set includes mainly low-permeable compounds. Several structurally different compounds relative to the descriptor space can be observed on the Williams plot (Paper **III**: Figure 7, Figure S2).

Table 10. The prediction models of the membrane permeability for amphoteric compounds (n_{train}=46, n_{val}=14) with the performance characteristics (Paper **III**: Figure 7, Figure S2).

pH	Prediction model	R^2_{train}	R^2_{cv}	s^2	R^2_{val}*
3	$\log P_{e\_pH3} = -7.30(\pm0.17) + 0.359(\pm0.040)\cdot\log P_{ow} + 0.0154(\pm0.0035)\cdot HACA - 0.0482(\pm0.0095)\cdot HDCA$	0.77	0.71	0.18	0.17 (0.47)
5	$\log P_{e\_pH5} = -9.27(\pm0.71) + 0.487(\pm0.042)\cdot\log P_{ow} + 4.04(\pm1.14)\cdot S_{XY}/R_{XY} - 0.00222(\pm0.00082)\cdot DPSA1$	0.77	0.72	0.19	0.45 (0.73)
7.4	$\log P_{e\_pH7.4} = -9.81(\pm1.43) + 0.321(\pm0.044)\cdot\log P_{ow} - 7.12(\pm1.48)\cdot N_{N,r} + 3.96(\pm1.74)\cdot {}^2\overline{BIC}$	0.7	0.65	0.17	0.4 (0.69)
9	$\log P_{e\_pH9} = -11.75(\pm1.52) + 0.283(\pm0.043)\cdot\log P_{ow} - 7.43(\pm1.61)\cdot N_{N,r} + 6.24(\pm1.85)\cdot {}^2\overline{BIC}$	0.64	0.57	0.2	0.33 (0.67)
Highest	$\log P_{e\_highest} = -9.25(\pm0.73) + 0.495(\pm0.045)\cdot\log P_{ow} + 4.94(\pm1.19)\cdot S_{XY}/R_{XY} - 0.975(\pm0.272)\cdot {}^2\overline{CIC}$	0.75	0.7	0.2	0.65 (0.75)

* Value without strong outliers is in the brackets.

The common descriptor for all models, the $\log P_{ow}$, has a positive correlation (Table 10), which means that more hydrophobic compounds have a higher membrane permeability. Additional descriptors in the models can be grouped according to the polarity and the size and shape of the molecule (Table 11).

Table 11. Occurrence of descriptors in the models for amphoteric compounds according to the data series.

Descriptor	Data series
Polarity	
H-acceptors charged surface area (*HACA*, MOPAC PC)	pH 3
H-donor charged surface area (*HDCA*, MOPAC PC)	pH 3
First-order difference in charged partial surface area (*DPSA1*, MOPAC PC)	pH 5
Relative number of nitrogen atoms ($N_{N,r}$)	pH 7.4, pH 9
Size and shape of the molecule	
Rate (S_{XY}/R_{XY}) of the surface on the plane (S_{XY}) and XY rectangle (R_{XY})	pH 5, $\log P_{e\_highest}$
Average bonding information content (order 2, ${}^2\overline{BIC}$)	pH 7.4, pH 9
Average complementary information content (order 2, ${}^2\overline{CIC}$)	$\log P_{e\_highest}$

More detailed analysis of the descriptors shows that all models include at least one descriptor connected to the polarity properties of compounds that describe the hydrogen bond donors and acceptors (Table 11). The hydrogen bond donor's property, the *HDCA* [176], has a negative effect to the membrane permeability (Table 10). The descriptor related to the hydrogen bond acceptor, the *HACA* [176], has a positive contribution in the model (Table 10). The $N_{N,r}$ has a negative contribution in the model (Table 10) describing partially hydrogen bond acceptor properties. The *DPSA1* [159], which takes into account both hydrogen bond donors (positive charge) and acceptors (negative charge),

has also a negative contribution to the model (Table 10). Trends of these descriptors show that the ability to form hydrogen bonds plays an important role for the membrane permeability process: hydrogen bond acceptors increase the membrane permeability and hydrogen bond donors decrease the membrane permeability.

Also, descriptors describing the size and shape of the molecule are included in the models of most data series. The S_{XY}/R_{XY} [177] has a positive contribution to the model (Table 10), which means that more compact molecule has a higher membrane permeability. The $^{2}\overline{BIC}$ [178] has also a positive contribution to the model (Table 10), which shows that the less complex molecule has a higher membrane permeability. The $^{2}\overline{CIC}$ [178] has a negative contribution to the model (Table 10), which indicates that the molecules with different functional groups cross the membrane more easily than more homogeneous molecules. Therefore, amphoteric compounds with higher membrane permeability are typically compact with various functional groups.

The models for amphoteric drug substances (Table 10) were tested with *Omeprazole*. Predicted values for all models are slightly lower than experimental values (Figure 16: c). The best agreement between experimental and calculated values is for the $\log P_{e\_highest}$, which also shows the best prediction capabilities for the validation set. The predicted membrane permeability for the selected pH-s varies considerably according to the pH-permeability profile, which is clearly visible and followed. *Omeprazole* is classified as low absorbed compound based on the %HIA value (%HIA = 80% [134]) and it is very near to the cutoff value for the %HIA. However experimental membrane permeability assigns this compound as a high permeable, while predicted data at pH 3, 7.4, and 9 classify it as low permeable.

4.3.5. Comparison of the QSAR models for chemical classes

The QSAR models of chemical classes show differences from two perspectives, the content of molecular descriptors and the quality of predictions (Figure 17).

The membrane permeability for chemical classes is influenced by the different characteristics of the structure (Figure 17). The $\log P_{ow}$ is not suitable descriptor to describe membrane permeability for most of the chemical classes, because it is not taking into account ionisation. The correlation between membrane permeability and $\log P_{ow}$ is non-existent for neutral compounds, whereas for amphoteric compounds, the correlation with the $\log P_{ow}$ can be considered significant. The $\log D$ improves the correlation for acidic and basic compounds, while for amphoteric compounds, the correlation was significantly lower compared to the $\log P_{ow}$. The comparison of descriptors in the models of different data series indicates that the membrane permeability for acidic, basic, and neutral compounds is influenced by the same structural characteristics for all pH-s, but for amphoteric compounds influencing structural characteristics are varying depending on the pH. The membrane permeability of acidic and neutral compounds is determined by the hydrogen bond donor properties. In

case of basic compounds the distribution properties are important for the membrane permeability. Descriptors in the QSAR models for amphoteric compounds are more diverse, because the membrane permeability involves multiple mechanistic interactions described by the partition properties, hydrogen bond donor and acceptor properties, and the size and shape of the molecule, which vary depending on the data series. This shows that the mechanism of membrane permeability and the pattern of interactions of different chemical classes is complex and can be attributed to the different structural characteristics.

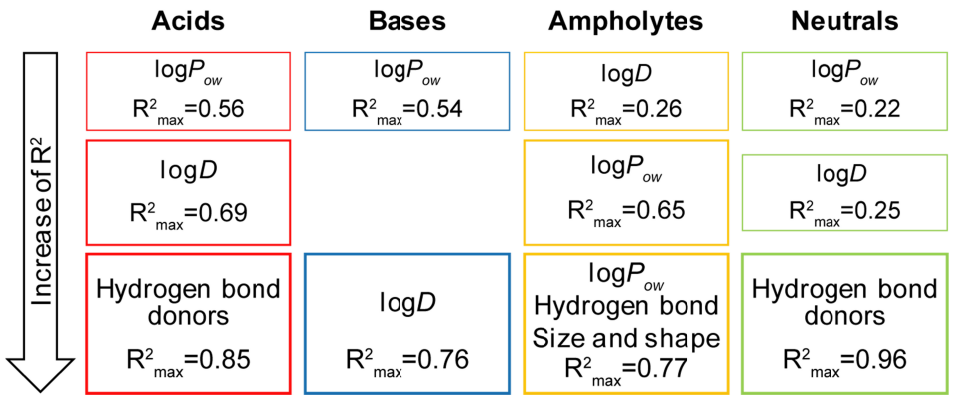


Figure 17. The comparison of important structure parameters for the QSAR models of chemical classes with maximum R^2 value over data series (the size of the box indicates the quality of the model).

The QSAR models for chemical classes show that pH has great influence on the membrane permeability. For example, the quality of the models for acidic and basic compounds significantly depends on pH (Figure 14, Figure 15). In general, the models with lower statistical significance (R^2) were developed for pH-s, where compounds are mainly ionised and do have therefore lower membrane permeability. For example, significant models for acidic compounds (Figure 14) were developed for acidic pH-s (compounds are not ionised), but not for neutral and basic pH-s (compounds are ionised). The quality of the models for amphoteric compounds also reveals the pH dependence (Figure 16), although the quality and the prediction ability of the models for different data series are similar. Only models for neutral compounds are not influenced by pH. All of the above reveals that knowing the pH-permeability profiles of chemical classes is also essential information for the modelling.

It can be concluded that the QSAR models of chemical classes allow predicting the membrane permeability of new compounds. The complete pH-permeability profiles (pH 3, 5, 7.4, and 9) can only be predicted for amphoteric and neutral compounds. For acidic and basic compounds, the membrane permeability can only be predicted at certain pH-s, as the development of models for mainly ionised compounds failed.

4.4. General QSAR models with the cutoff classification

The purpose of deriving the general QSAR models, which consider all chemical classes together, is a multi-faced. The first reason is from the models for chemical classes (Section 4.3), where the QSAR models for acidic and basic compounds at pH-s, where the compounds are mainly ionised, did not give sufficient statistical quality. The general QSAR models for data series with all chemical classes should allow predicting the membrane permeability for acidic and basic compounds for all pH-s. The second reason is related to the more enhanced view to the structural characteristics influencing the membrane permeability that are included in the QSAR models. The third reason is related to the unique opportunity to predict the pH-permeability profiles with the QSAR models derived for each pH.

The general QSAR models were developed for all measured and derived data series (Section 3.2): pH-permeability profile that included four pH-s, highest, and intrinsic membrane permeability. Such QSAR models allow quantitatively predict the membrane permeability values. It is often necessary to categorize the membrane permeability, for example, for the BCS. For this continuous experimental and predicted values can be converted to the classes of high- and low-permeable compounds using the cutoff value of $\log P_e = -6.20$ (Section 3.3). The advantage of such classification is the option to validate the class of high- or low-permeable compounds based on the predicted numerical value, and in this way, to evaluate the appropriateness of the membrane permeability class.

4.4.1. Membrane permeability vs logP_{ow} and log$D$

At first the correlations between membrane permeability and partition or distribution parameter were studied (Figure 18: a, Paper **IV**). The $\log P_{ow}$ has the highest correlation in the case of the $\log P_o$ and the $\log P_{e\_highest}$, because these data series describe membrane permeability for mainly unionised compounds. At a specific pH-s the correlation of $\log P_{ow}$ with membrane permeability becomes negligible due to the compound's ionisation.

In comparison with the $\log P_{ow}$, the $\log D$ improves the relationship with membrane permeability for all data series (Figure 18: a). The improvement is remarkable for the models at acidic pH-s, and slightly lower for the models at neutral and basic pH-s. Analogously, with pH specific data series the correlation with the $\log P_{e\_highest}$ is also better with the $\log D_{highest}$ in comparison to the $\log P_{ow}$.

In conclusion, the general QSAR models with the $\log P_{ow}$ and the $\log D$ alone do not capture the full structural variability needed to describe and predict the membrane permeability. This is because the membrane is a more complex system than octanol. Additional descriptors are therefore needed to improve the coverage of structural variability and describe the interactions in the membrane permeation process. Indeed, the whole-molecule descriptors improve the

statistical characteristics of all models (Figure 18: a, best model), which are discussed in the following sections.

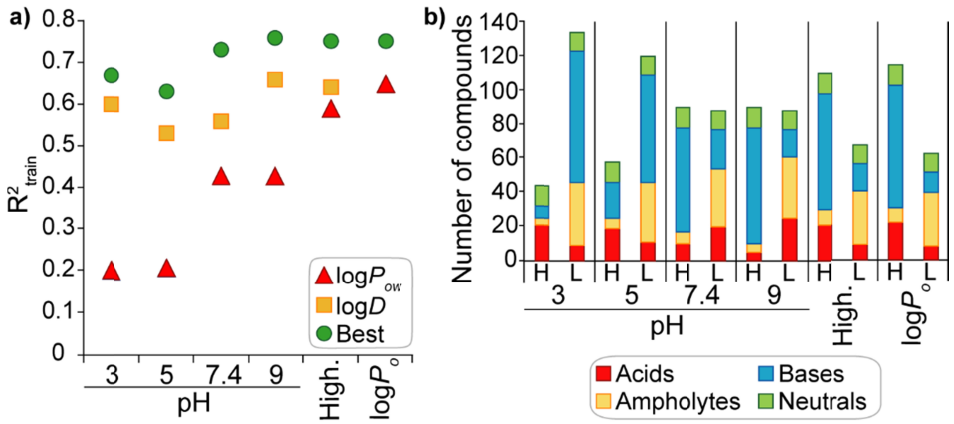


Figure 18. The general QSAR models: (a) the comparison of the models with different sets of descriptors and (b) the distribution of high- (H) and low-permeable (L) compounds in the initial set of compounds for all data series (High. – $\log P_{e\_highest}$).

4.4.2. QSAR models for the pH-permeability profile

Four QSAR models at different pH-s allow predicting the pH-permeability profiles of drug substances (Paper **IV**). The membrane permeability in these models is influenced by the different structure properties of drug substances. In all models the log*D* has a positive contribution (Table 12) meaning that more hydrophobic compounds have a higher membrane permeability.

Table 12. The general QSAR models for the membrane permeability at a certain pH (n_{train}=134, n_{val}=44, n_{ext}=60) with the performance characteristics (Paper **IV**: Figure 4, Figure S6).

pH	Prediction model	R^2_{train}	R^2_{cv}	s^2	R^2_{val}	R^2_{ext}
3	$\log P_{e\_pH3} = -6.85(\pm0.15) + 0.408(\pm0.030)\cdot logD_{pH3} + 0.112(\pm0.026)\cdot RNCS - 1.92(\pm0.76)\cdot HDCA\text{-}2/SQRT(TMSA)$	0.67	0.64	0.35	0.64	0.64
5	$\log P_{e\_pH5} = -6.40(\pm0.16) + 0.381(\pm0.031)\cdot logD_{pH5} + 0.096(\pm0.025)\cdot RNCS - 3.19(\pm0.78)\cdot HDCA\text{-}2/SQRT(TMSA)$	0.63	0.6	0.38	0.61	0.56
7.4	$\log P_{e\_pH7.4} = -5.51(\pm0.13) + 0.305(\pm0.041)\cdot logD_{pH7.4} - 0.00735(\pm0.00081)\cdot HBSA$	0.73	0.72	0.36	0.63	0.7
9	$\log P_{e\_pH9} = -5.58(\pm0.15) + 0.297(\pm0.038)\cdot logD_{pH9} - 0.00700(\pm0.00093)\cdot HBSA$	0.76	0.75	0.36	0.72	0.71

The models for pH 3 and 5 (Table 12) include two additional descriptors: the relative negative charged surface area (*RNCS*) and the hydrogen bond donor-charged surface area divided by the square root of the total molecular surface

area (HA-dependent *HDCA-2/SQRT(TMSA)*, MOPAC PC, all). The *RNCS* [179] has a positive contribution to the models, telling that molecule with a higher negative charge (usually hydrogen bond acceptors) has a higher membrane permeability. The *HDCA-2/SQRT(TMSA)* [138] has a negative correlation with the membrane permeability, which suggests that if a compound has fewer hydrogen bond donors, then the compound will have a higher membrane permeability. The models for pH 7.4 and 9 include only one additional descriptor to complement the log*D* (Table 12), which is the hydrogen bonding surface area (*HBSA*, MOPAC PC). The *HBSA* [176] has also negative contribution to the model, meaning that compound with a smaller hydrogen bonding surface area has a higher membrane permeability. This tells that the membrane permeability for a diverse range of drug substances (different chemical classes) at a certain pH is mainly influenced by two types of structure related properties that describe the hydrophobic/hydrophilic properties together with the ionisation of compounds (log*D*) and the ability of hydrogen bond formation.

The models for pH 3 and 5 have lower statistical characteristics for the training set in comparison with the models for pH 7.4 and 9 (Figure 19: a, Table 12). This can be due to a larger number of low- than high-permeable compounds in the data series for pH 3 and 5 and a similar number of high- and low-permeable compounds in the data series for 7.4 and 9 (Figure 18: b). The validation set shows that the models for pH 3, 5, and 7.4 have lower prediction statistics compared to the model for pH 9. The model for pH 9 shows the highest and consistent description and prediction statistics for the training and validation sets compared to other pH-s (Figure 19: a, Table 12) and very good match between experimental and calculated membrane permeabilities without strong outliers (Figure 19: b). The analysis of the applicability domain (Paper **IV**: Figure S6) revels some outliers and high leverage compounds for all data series. The reason of outliers may be the accuracy of the calculated log*D*. High leverage values for compounds are caused by the extreme values for different descriptors.

The cutoff-based classification for all pH-s has accuracy, sensitivity, and specificity over 0.8 for the training sets (Figure 19: c). For the validation set, all quality measures are also over 0.8, with exception of specificity and accuracy for the model at pH 5, which can be attributed to the lowest determination coefficient for the model in comparison of the models for other pH-s. Comparing statistical parameters for the regression models and the cutoff-based classification for acidic pH-s shows that low determination coefficient for the regression model does not always mean that model has low accuracy.

The QSAR models were externally validated with 60 compounds. The models for pH 7.4 and 9 show significantly higher prediction capability for the external validation sets compared to the models for pH 3 and 5 (Figure 19: a, Table 12). This is due to the lower descriptive properties on the training set for acidic pH-s compared to neutral and basic pH-s. An analysis of the applicability domain for the external validation set of every data series revealed few moderate outliers and high-leverage compounds (Paper **IV**: Figure S6). The cutoff-based classification of the external validation set (Figure 19: c) shows

that most classification parameters for the models are greater than 0.7, which indicates good classification capability for drug substances.

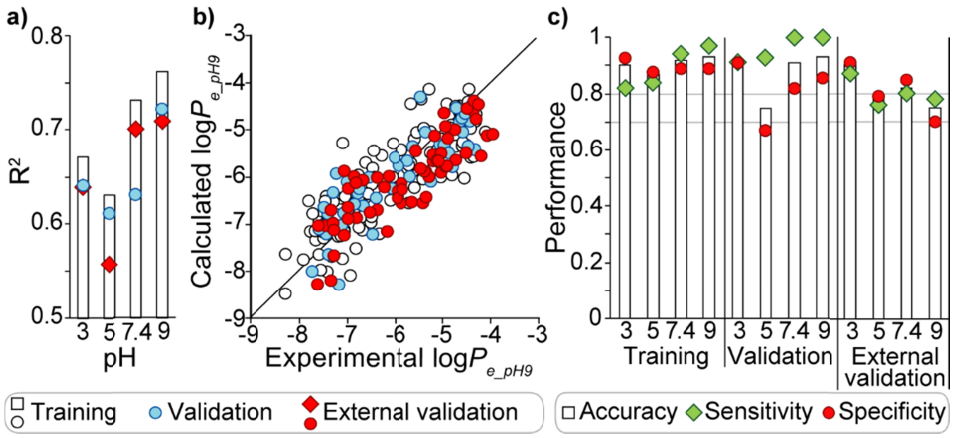


Figure 19. Summary of the QSAR models for the pH-permeability profile: (a) the comparison of determination coefficients (R^2) for the training, validation, and external validation sets, (b) the match between experimental and calculated membrane permeability at pH 9, and (c) the comparison of performance characteristics for the cutoff based classification.

4.4.3. QSAR models for the highest membrane permeability

The highest membrane permeability over selected pH-s allows to describe the maximum absorption in the GIT. In the current thesis, two such QSAR models have been developed for a set of compounds with different sizes.

The first QSAR model for the highest membrane permeability (Paper **I**), a so-called proof of concept model, was developed using the initial set of 58 drug substances (44 in the training and 14 in the validation set). This model consists of two descriptors (Eq. (26)): the logarithm of the octanol-water partition coefficient (logP_{ow}) and the hydrogen bonding surface area (*HBSA*). The logP_{ow} has a positive contribution in the model, which reveals that more hydrophobic compounds have a higher membrane permeability. The *HBSA* [159] has a negative correlation with the membrane permeability, which means that when the *HBSA* values increase, then the membrane permeability decreases due to the formed hydrogen bonds between drug substances and membrane.

$$\begin{gathered}\log P_{e\_highest} = -5.6261(\pm 0.2322) + 0.5182(\pm 0.0519)\cdot logP_{ow} - \\ 0.0075(\pm 0.0010)\cdot HBSA_{DA} \\ n_{train} = 44,\ R^2_{train} = 0.83,\ R^2_{cv} = 0.80,\ s^2 = 0.21,\ n_{val} = 14,\ R^2_{val} = 0.62,\ n_{ext} = 15, \\ R^2_{ext} = 0.66\end{gathered} \tag{26}$$

The model has exceptionally good statistical characteristic for the training set (R^2_{train}=0.83), although for the validation set (14 compounds) statistical characteristic is remarkably lower (R^2_{val}=0.62). This is most likely caused by the

distribution of data points, i.e. more high- than low-permeable compounds are in the validation set (Figure 20: a). The applicability domain analysis using the Williams plot (Paper **I**: Figure 2) revels outliers that deviate due to the accuracy of the logP_{ow} and the limited solubility at corresponding pH and structurally different compounds, with extreme descriptor values. The external validation set (15 compounds, R^2_{ext}=0.66) shows slightly higher statistical characteristics compared to the validation set. Both validation sets show that this model is reliable and suitable for predicting the membrane permeability. This model also confirms that the hypothesis to use the highest membrane permeability values over a wide pH range considerably improves the quality of the QSAR models (see overview of previous models in Paper **I**) and their prediction performance and can be developed further with a larger set of compounds.

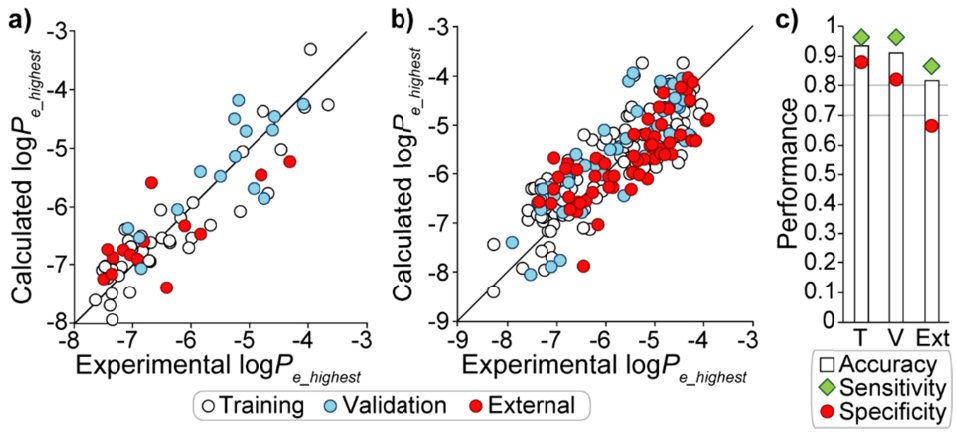


Figure 20. The correlation with experimental and predicted highest membrane permeability over four pH-s (log$P_{e_highest}$): (a) for a set of 58 drug substances (Paper **I**), (b) for a set of 178 drug substances (Paper **IV**), and (c) the comparison of performance characteristics for the cutoff based classification (Paper **IV**).

The second model for the log$P_{e_highest}$ (Eq. (27), Paper **IV**) extends the previous model (Eq. (26)) and was developed using 178 drug substances (134 in the training and 44 in the validation set). The model (Eq. (27)) includes two descriptors: the log$D_{highest}$ and the *HBSA*, which are similar descriptors to the previous model (Eq. (26)) and describe the same structural characteristics. The model shows good agreement between experimental and calculated log$P_{e_highest}$ for the training (R^2_{train}=0.75) and the validation (R^2_{val}=0.70) sets (Figure 20: b). Good statistical characteristics are also visible for the cutoff based classification, where more than 80% of the compounds are correctly classified (Figure 20: c) in the training and validation sets. The applicability domain analysis of the model (Paper **IV**: Figure S6) reveals few outliers; the reason can be accuracy of the log$D$ values, and high leverage compounds with extreme descriptor values. The model was externally validated with a blind set of 60 compounds, which has a lower determination coefficient (R^2_{ext}=0.60) compared

to the validation set (R^2_{val}=0.70). The cutoff based classification has also lower accuracy for the external validation set compared to the validation set, which is due to the misclassification of low-permeable compounds, i.e. the model predicts values higher than they should be.

$$\log P_{e\_highest} = -5.59(\pm 0.17) + 0.341(\pm 0.039)\cdot \log D_{highest} - 0.00657(\pm 0.00086)\cdot HBSA$$
$$n_{train} = 134,\ R^2_{train} = 0.75,\ R^2_{cv} = 0.74,\ s^2 = 0.34,\ n_{val} = 44,\ R^2_{val} = 0.70,\ n_{ext} = 60,\ R^2_{ext} = 0.60 \quad (27)$$

The comparison of two models for the highest membrane permeability reveals that both models take into account a structural variability related to the hydrophobicity and hydrogen bond properties of molecular structure. It is interesting that for the smaller set of compounds, the logP_{ow} is more significant than the log$D_{highest}$, but for the larger set of compounds, the situation is opposite. This most likely can be caused by the precision of calculated log$D$ values, where the correlation with the membrane permeability for the smaller set of compounds is highly influenced by the calculation quality of the log$D$, but the model for the larger set of compounds is less influenced by deviating points. Notably the hydrogen bonding properties are described with the same descriptor and both models are independently derived with the BMLR procedure (Section 3.6.4). Comparing the statistical parameters of two models shows that the statistical characteristics for the training set and for the external validation set are slightly better for the model developed on the smaller set of compounds, but the validation set performs better for the larger set of compounds. In summary, these two models for the highest membrane permeability are composed of similar molecular descriptors and do have comparable statistical quality and stability.

4.4.4. QSAR model for the intrinsic membrane permeability

The intrinsic membrane permeability describes the membrane permeability when the compound is unionised. This makes it easier to develop a QSAR model compared to the pH specific models, because there is no need for the molecular descriptor that takes into account ionisation. As a consequence, the model for the intrinsic membrane permeability (Eq. (28)) includes two descriptors, the logP_{ow} and the hydrogen bonding surface area (*HBSA*). The model shows that the membrane permeability is higher for hydrophobic compounds (positive correlation) with smaller hydrogen bonding surface areas (negative correlation), i.e. compounds having less hydrogen bond acceptors and donors.

$$\log P_o = -5.13(\pm 0.26) + 0.572(\pm 0.060)\cdot \log P_{ow} - 0.00941(\pm 0.00014)\cdot HBSA$$
$$n_{train} = 134,\ R^2_{train} = 0.75,\ R^2_{cv} = 0.73,\ s^2 = 0.93,\ n_{val} = 44,\ R^2_{val} = 0.76,\ n_{ext} = 60,\ R^2_{ext} = 0.65 \quad (28)$$

The model has coincided statistical parameters for the training (R^2_{train}=0.75) and validation (R^2_{val}=0.76) sets. The cutoff based classification (Figure 21: b)

reveals that accuracy and sensitivity (prediction of high-permeable compounds) are over 0.9, while specificity drops below 0.8 indicating that the model for the logP_o predicts values higher than they should be. The comparison between experimental and calculated values (Figure 21: a) show that the training set has two strong outliers, which are also detected by the Williams plot (Paper **IV**: Figure S6). The reason of the outliers can be accuracy of the logP_{ow}, but could also be due to an inaccurate pK_a, which greatly influences the derived logP_o values and can introduce deviations about 1–2 logP_o units depending on the correctness of experimental or calculated pK_a values. The applicability domain (Paper **IV**: Figure S6, Table 2) analysis reveals also some high leverage compounds, which have extreme descriptor values.

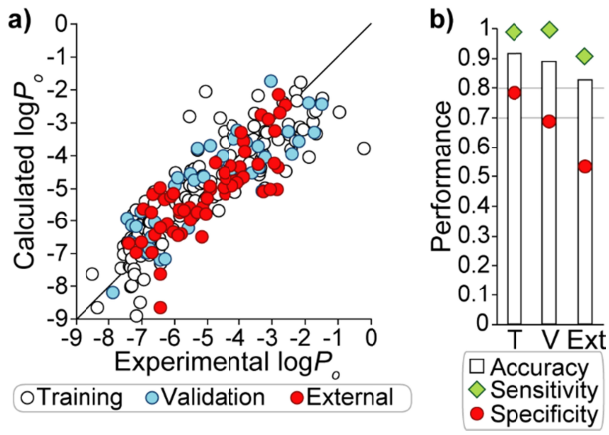


Figure 21. Summary of the QSAR model for the intrinsic membrane permeability: (a) the comparison between experimental and calculated logP_o and (b) the comparison of performance characteristics for the cutoff based classification.

The model for the logP_o was blindly validated with 60 compounds from the external validation set, which showed a lower statistical parameters compared to the training and validation sets based on the QSAR model (Eq. (28)) and related cutoff based classification (Figure 21). The cutoff based classification has very low prediction capability for low-permeable compounds (specificity), which is caused by the small number of low-permeable compounds compared to high-permeable compounds in the external validation set.

4.4.5. Prediction of the pH-permeability profiles

One compound from each chemical class of the external validation set (Paper **IV**) was selected to illustrate the prediction of a pH-permeability profile with the developed QSAR models (Figure 22): acidic *Tolmetin*, basic *Granisetron*, amphoteric *Sulpiride*, and neutral *Coumarin*. For acidic, basic, and amphoteric example compounds, experimental and predicted pH-permeability matches well and only a slight deviation is presented at lower membrane permeability values, where experimental values have lower accuracy. For the

neutral compound the predicted values give a pH-permeability profile systematically lower compared to the experimental values. Interestingly, the prediction for the logP_o for the neutral compound is very near to the experimental value compared to the other models, which can be due to the more precise logP_{ow} value than the log$D$ value.

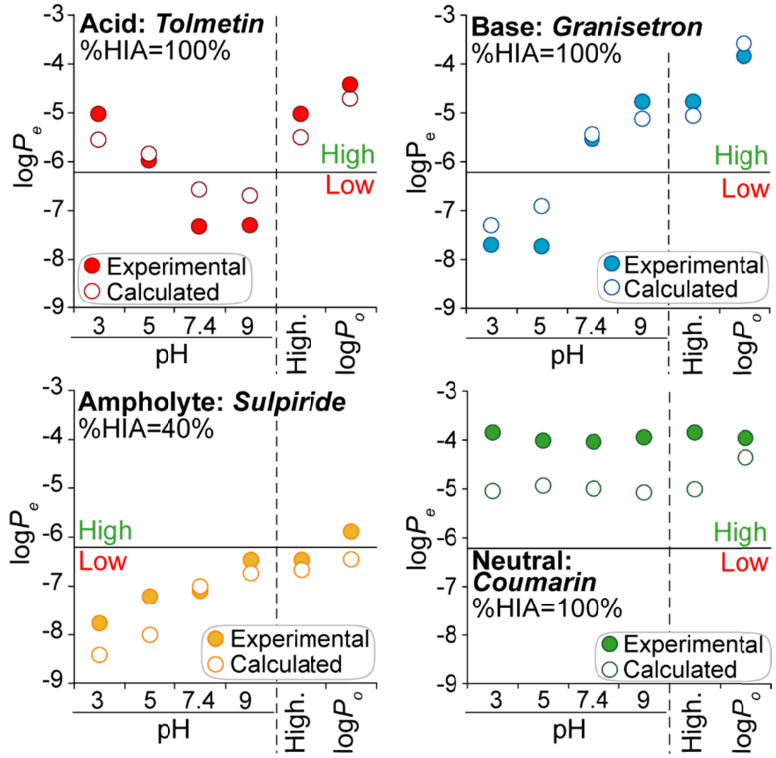


Figure 22. Experimental and predicted pH-permeability profiles for four compounds from the external validation set, one from each chemical class (High. – log$P_{e_highest}$).

The classification results according to the cutoff value are in good agreement (Figure 22), and all predicted membrane permeabilities belong to the same class as experimental membrane permeabilities. The exception is the logP_o for the amphoteric compound, where according to the experimental logP_o the compound is a high permeable, but the calculated value classifies it as a low permeable. The misclassification can be a coincidence, because the experimental and the predicted value are near to the cutoff value for the membrane permeability, but on the different sides of the cutoff value.

A comparison of the membrane permeability with the %HIA [134] shows that high absorbed acidic, basic, and neutral compounds (%HIA ≥ 85%) are classified correctly if membrane permeability at a correct pH is considered (Figure 22), i.e. acidic compound at acidic pH-s, basic compound at basic pH-s, and neutral compound at any pH. The comparison of the membrane perme-

ability with the %HIA for low absorbed *sulpiride* (ampholyte) provides a different perspective, where all predicted and experimental values classify the compound as a low permeable, except the experimental logP_o, which indicates that the compound is a high permeable.

These examples provide evidence that the developed general QSAR models at different pH values are reliable to evaluate the pH-permeability profile and allow the classification of high- or low-permeable compounds. In addition, the predicted pH-permeability profile allows to estimate at which pH the membrane permeability is highest, indicating regions where the compound is absorbed in the GIT and under what conditions.

4.4.6. Comparison of the general QSAR models

The derived general QSAR models have systematic variation in the molecular descriptors. The models at acidic pH-s include three descriptors, while for other data series two descriptors. For all models, the main descriptor is related to the hydrophobicity of molecules (logP_{ow} and log$D$). Interesting is that for the models of chemical classes, only basic and acidic compounds show significant correlation with the log$D$ (Figure 17). This suggests that acidic and basic compounds do have a major effect to the general models. Descriptors logP_{ow} or log$D$ alone are not enough to capture the structural variability in the given data series while modelling the membrane permeability (Figure 18). This is due to octanol being a neutral molecule, while lecithin is ionised, and the membrane has therefore more hydrogen bonding sites than octanol. Therefore, the developed models show clearly that the membrane permeability is also affected by the ability to form hydrogen bonds. Molecules with high hydrogen bonding capability can accumulate into the membrane, but also can form bound structures with the buffer, which prevents their easy permeation into the membrane. For the acidic pH-s, two descriptors were added to the model describing the hydrogen bond donor properties (*HDCA-2/SQRT(TMSA)* and the negative charge of the molecule (*RNCS*), which is indirectly related to the hydrogen bond acceptors. Whereas, the models for pH 7.4, pH 9, log$P_{e_highest}$, and logP_o include one additional descriptor (*HBSA*) to describe both hydrogen bond acceptors and donors. The hydrogen bond donor properties (*HDCA-2/SQRT(TMSA))* are described with similar descriptor as in the models for acidic and neutral compounds, which indicates that the general models at acidic pH-s are mainly influenced by neutral and acidic compounds.

The comparison of data series shows that the quality of the model's statistical characteristics is highly influenced by the distribution of the data points. In the data series for pH 3 and 5, low-permeable compounds dominate (Figure 18: b), and their measured data have a lower accuracy compared to high-permeable compounds. This results in the lower statistical parameters compared to other data series where the data points are evenly distributed and include more high-permeable compounds. It is also necessary to realise that the calculated log$D$ values are, in turn, derived from the calculated pK_a and logP_{ow},

and thus the log$D$ is highly influenced by the accuracy of those calculated parameters.

For the cutoff based classification, in most of the cases the prediction capability for high- and low-permeable compounds together (accuracy) and separately (sensitivity and specificity) is greater than 80%. This can be considered a noticeable achievement taking into account both the structural and chemical diversity of the sets of compounds and the complexity of the membrane permeability process. It must be pointed out that the values of specificity, sensitivity, and accuracy are highly influenced by the predictions that are very near to the cutoff value. Because of this, the classification and the QSAR models complement each other, offering a quick insight into high and low membrane permeability of a compound.

All developed QSAR models and related cutoff based classification can be used for predicting the membrane permeability depending on the purpose: the pH-specific models allow to predict the pH-permeability profiles and regional dependent absorption, the log$P_{e_highest}$ to predict the maximum absorption in the GIT, and the logP_o to predict the maximum membrane permeability for a compound.

4.5. Logistic classification models

The cutoff based classification using the regression models (Section 4.4) show significant capability to distinguish high- and low-permeable compounds. This leads to the classification models for the membrane permeability relying purely on the chemical structure using a specific classification method. The logistic regression method (Section 3.6.5) was selected for such task and two sets of descriptors were used:

- Only hydrophobicity descriptor (logP_{ow} or log$D$), which is significant descriptor for the modelling of membrane permeability,
- Freely available and simply computable 1D and 2D molecular descriptors for deriving models that can be easily interpreted and used.

For the logistic classification task, high-permeable compounds were defined as class “0” and low-permeable compounds as class “1”. This means that a positive correlation in the logistic classification model implicates to low permeability and negative correlation to high permeability.

Regression models are significantly influenced by the distribution of high- and low-permeable compounds (Figure 18: b). The results of the classification models are influenced even more by unbalanced classes in the training set. Therefore, the balanced training sets were constructed as described in Section 3.6.3.

4.5.1. Classification models with a hydrophobicity descriptor

The hydrophobicity descriptors (logP_{ow} or log$D$) have significant correlation with the membrane permeability ($R^2 < 0.65$, Section 4.4.1), which is not enough

for the regression models, but can be sufficient for the classification models. Therefore, it was investigated whether sufficient quality classification models can be developed using only one of these hydrophobicity descriptors in the model. Depending on the nature of data series either log$D$ or logP_{ow} was used. The developed classification models (Table 13) show that the hydrophobicity descriptor has a negative correlation in the model, which means that hydrophobic compounds belongs more likely to class 0, i.e. high permeable, than class 1, i.e. low permeable. This is a similar conclusion to the regression models (Section 4.4), where the hydrophobicity descriptor has a positive correlation with the membrane permeability.

Table 13. The classification models (z) with a hydrophobicity descriptor (z) for all data series (Paper **V**: Figure 4, Figures S1–S5).

Data series	Prediction model
pH 3	$z_{M1} = 1.9182(\pm0.5936) - 1.7339(\pm0.4366)\cdot\log D_{pH3}$
pH 5	$z_{M2} = 1.3569(\pm0.4149) - 1.6234(\pm0.3427)\cdot\log D_{pH5}$
pH 7.4	$z_{M3} = 1.3899(\pm0.3298) - 1.6063(\pm0.2686)\cdot\log D_{pH7.4}$
pH 9	$z_{M4} = 2.0555(\pm0.4488) - 1.6828(\pm0.2695)\cdot\log D_{pH9}$
log$P_{e_highest}$	$z_{M5} = 3.0469(\pm0.7128) - 2.0974(\pm0.4200)\cdot\log D_{highest}$
logP_o	$z_{M6} = 2.7082(\pm0.6470) - 1.7840(\pm0.3696)\cdot\log P_{ow}$

The resulting classification models show good statistical characteristics for the training and validation sets (Figure 23). More than 80% of the compounds are correctly classified in the training set and more than 70% of the compounds are correctly predicted in the validation set. For the training set, the classification models can describe equally high- and low-permeable compounds, i.e. have similar sensitivity and specificity. The validation set shows that high-permeable compounds have higher prediction capability compared to low-permeable compounds, i.e. higher sensitivity than specificity. This however can be caused by the unbalanced validation sets for most of the data series (Table 5). Additionally, the classification models were validated with the external validation set (60 compounds), which show that more than 70% of the compounds are correctly predicted for all data series and prediction capability for high- and low-permeable compounds is well balanced. The training, validation, and external validation sets together show that the classification models with only one hydrophobicity descriptor have a good potential in describing and predicting the permeability classes for all data series.

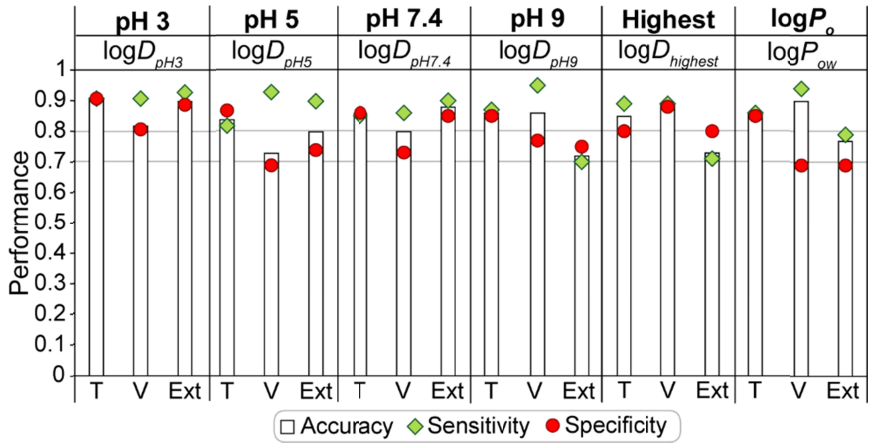


Figure 23. The comparison of the performance characteristics for the classification models with a hydrophobicity descriptor (T – training, V – validation, Ext – external validation)**.**

However, these one-parameter classification models are highly influenced by the precision of the logP_{ow} and the log$D$ values. Three key issues for the log$D$ should be considered while using the developed classification models:

- The precision of the log$D$ values, because they are not systematically evaluated and analysed for various pH-s,
- The prediction of the log$D$ is usually based on the relationship between the logP_{ow} and pK_a, which means that the precision of the log$D$ is highly dependent on the accuracies of the logP_{ow} and pK_a,
- The lack of freely available log$D$ calculators that limits the practical use of the developed classification models.

The log$D$ is an important descriptor for describing the membrane permeability for the different pH-s, because it is able to describe both the partition properties and the ionisation in one parameter. So far, there are no other good solution to take account ionisation for all chemical classes using a single approach. Previously, in modelling the membrane permeability, difference between pK_a and pH has been used (Section 1.3.7), but it can be accurate only for compounds with one ionisable group. Considering the previously mentioned disadvantages it is reasonable to develop classification models without log$D$.

4.5.2. Classification models with theoretical molecular descriptors

Considering the complex nature of a hydrophobicity descriptor and the concerns above, alternatives were sought to replace the log$D$ and the logP_{ow} in the classification models (Paper **V**). The 1D and 2D theoretical molecular descriptors calculated with the PaDEL-Descriptor [160,161] were tested as a source of alternative descriptors. The developed classification models (Table 14) include

two descriptors; the exception is the classification model for pH 7.4, which includes three descriptors.

Table 14. The classification models (z) with theoretical molecular descriptors for all data series (Paper **V**: Figure 4, Figures S1–S5).

Data series	Prediction model
pH 3	$z_{M7} = -2.25 + 3.17 \cdot nBase + 31.77 \cdot ETA\_dEpsilon\_D$
pH 5	$z_{M8} = -93.31 + 12.03 \cdot Mi + 1.26 \cdot nHBDon$
pH 7.4	$z_{M9} = 8.12 + 1.80 \cdot nAcid + 49.38 \cdot ETA\_dEpsilon\_D - 19.44 \cdot ETA\_Psi\_1$
pH 9	$z_{M10} = -4.46 + 3.19 \cdot nAcid + 0.0547 \cdot TopoPSA$
$\log P_{e\_highest}$	$z_{M11} = -19.20 + 1.57 \cdot nHBDon + 10.19 \cdot IC0$
$\log P_o$	$z_{M12} = -13.66 + 0.87 \cdot nHBDon + 7.42 \cdot IC0$

The descriptors in the classification models vary over a data series (Table 14). They describe the properties of the molecular structure associated with the ionisation, the polarity, and the size and complexity of the molecule (Table 15). All classification models for the specific pH include one descriptor for accounting ionisation, which are a number of basic (*nBase*) [160,161] or acidic (*nAcid*) [160,161] groups, or the mean first ionisation potential (scaled on the carbon atom) (*Mi*) [161]. The *nBase*, *nAcid*, and *Mi* have a positive correlation in the classification models, which means that the ionisation of compound reduces the membrane permeability. The classification models for all data series include also descriptors, which are associated with the polarity of molecules that is related to the hydrogen bonds and surface area: a measure of contribution of hydrogen bond donor atoms (*ETA_dEpsilon_D*) [180], a number of hydrogen bond donors (*nHBDon*) [161,181], a measure of the hydrogen bonding propensity of the molecules and/or polar surface area (ETA_Psi_1) [180], and the topological polar surface area (*TopoPSA*) [182]. The *ETA_dEpsilon_D*, *nHBDon*, and *TopoPSA* have a positive correlation in the classification models (Table 14), which means that low-permeable compounds have more hydrogen bond donor groups and higher polar surface area. The *ETA_Psi_1* has a negative correlation in the classification model (Table 14: pH 7.4), which shows that a higher membrane permeability is for the compound with less electronegative atoms, which are parts of the hydrogen bond forming and polar surface area. All these descriptors describe the interactions with buffer and membrane, which in turn causes difficulties to move across the membrane. Only one descriptor is associated with the size and complexity of the molecules, a 0^{th} order information content index (*IC0*) [176,178], which is included in the classification models for the $\log P_{e\_highest}$ and $\log P_o$. The *IC0* has a positive correlation in the classification models (Table 14), telling that low-permeable compounds are more complex molecules with different atom types. All descriptors have a different range and spread of values for high- and low-permeable compounds (Paper **V**: Figure 4, Figures S1-S5), indicating that all these descriptors have a good discriminative ability to classify high- and low-permeable compounds.

Table 15. Theoretical molecular descriptors in the classification models.

Properties of molecular structure	Descriptors	Models
Ionisation	*nBase*	pH 3
	Mi	pH 5
	nAcid	pH 7.4 and 9
Polarity	*ETA_dEpsilon_D*	pH 3 and 7.4
	nHBDon	pH 5, $\log P_{e\_highest}$, $\log P_o$
	ETA_Psi_1	pH 7.4
	TopoPSA	pH 9
Size and complexity of the molecule	*IC0*	$\log P_{e\_highest}$, $\log P_o$

All classification models with theoretical molecular descriptors (Figure 24) show good accuracy (over 0.7) for the training and validation set. Sensitivity and specificity are similarly good for all training and validation sets, except the validation sets for the models at pH 3 and 5. Lower prediction capability for low-permeable compounds (specificity) at pH 3 and 5 indicates that predicting permeability classes for acidic pH-s are more complicated compared to the other data series, which can be due to the composition of the validation sets that includes more low-permeable compounds (100) compared to high-permeable compounds (11) (Table 5). Good prediction capability is confirmed with the external validation set, which has satisfactory performance characteristics, like accuracy over 0.7 and similarly good prediction capability for low- and high-permeable compounds

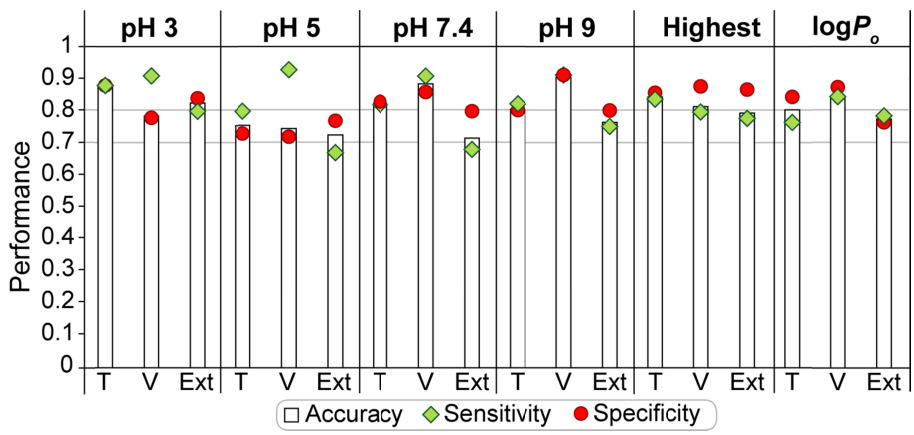


Figure 24. The comparison of the performance characteristics for the classification models with theoretical molecular descriptors (T – training, V – validation, Ext – external validation).

Overall the results show that theoretical molecular descriptors allow to develop classification models for all data series with good performance characteristics in all levels. The performance characteristics comparison of the models with a hydrophobicity descriptor and theoretical molecular descriptors shows that accuracy is typically slightly higher for the models with a hydrophobicity

descriptor, while both types of models can equally well predict high- and low-permeable compounds. This indicates that the hydrophobicity descriptors can be replaced with the theoretical molecular descriptors, which makes the models more robust compared to the models with a hydrophobicity descriptor, because they depend on multiple descriptors with different range of values.

4.5.3. Predicted class-based pH-permeability profiles

The classification models with both hydrophobicity and theoretical molecular descriptors show very good performance for all data series. To see how well the derived classification models can evaluate the class-based pH-permeability profiles (Paper **V**), four compounds from each chemical class were selected (Table 16). The high coherence between experimental and predicted data is evident for both types of classification models. The classification models with a hydrophobicity descriptor can correctly predict class-based pH-permeability profiles for three compounds. The pH-permeability profile for *Propranolol* includes one misclassification, where the classification model with a hydrophobicity descriptor at pH 7.4 incorrectly predicted it as a low-permeable instead of a high-permeable compound. This misclassification can be caused by the precision of the calculated log*D* at pH 7.4. The classification models with theoretical molecular descriptors predict correctly class-based pH-permeability profiles for all four compounds. Both types of classification models predict correctly classes for the $\log P_{e\_highest}$ and the $\log P_o$ for all four compounds.

Table 16. The comparison of the predictions from the classification models with a hydrophobicity descriptor ($\log D/\log P_{ow}$) and theoretical molecular descriptors (Mol. desc.) for four example compounds (H – high, L – low, High. – $\log P_{e\_highest}$).

a) *Ketoprofen*					BCS: high		**b)** *Propranolol*					BCS: high	
Acid	**3**	**5**	**7.4**	**9**	**High.**	**logP_o**	**Base**	**3**	**5**	**7.4**	**9**	**High.**	**logP_o**
Exp	H	H	L	L	H	H	Exp	L	L	H	H	H	H
log$D$/logP_{ow}	H	H	L	L	H	H	log$D$/logP_{ow}	L	L	L	H	H	H
Mol. desc.	H	H	L	L	H	H	Mol. desc.	L	L	H	H	H	H
c) *Famotidine*					BCS: low		**d)** *Carbamazepine*					BCS: high	
Ampholyte	**3**	**5**	**7.4**	**9**	**High.**	**logP_o**	**Neutral**	**3**	**5**	**7.4**	**9**	**High.**	**logP_o**
Exp	L	L	L	L	L	L	Exp	H	H	H	H	H	H
log$D$/logP_{ow}	L	L	L	L	L	L	log$D$/logP_{ow}	H	H	H	H	H	H
Mol. desc.	L	L	L	L	L	L	Mol. desc.	H	H	H	H	H	H

Consequently, according to the four example compounds, the classification models with theoretical molecular descriptors are able to correctly predict the class-based pH-permeability profiles for all four compounds, while the classification models with a hydrophobicity parameter misclassify one compound for one data series, showing a better robustness of theoretical molecular descriptors. Similarly to the general QSAR models (Section 4.4), the use of the classification models depends on the purpose: the pH-specific classification

models allow predicting the class-based pH-permeability profiles and the regional dependent absorption; the classification model for the log$P_{e_highest}$ predicts the highest permeability class in the GIT; and the classification model for the logP_o predicts permeability class for the unionised form of the compound.

4.6. Classifying the U.S. FDA reference drug substances

A set of the U.S. FDA reference drug substances for the BCS permeability class (Section 3.5) was used as the third validation and also a practical use case for the developed logistic classification models (Section 4.5). Experimental membrane permeability data and classification models (Sections 4.5.1 and 4.5.2) were applied to predict the permeability class for the BCS.

The comparison of classes for the experimental membrane permeability with the permeability classes in the BCS shows interesting trends (Figure 25: a, previously unpublished comparison). Experimentally measured high-permeable compounds (10) show a particularly good match at pH 9 and for the log$P_{e_highest}$ and the logP_o. Only half of high-permeable compounds are correctly classified at pH 3, 5, and 7.4. More than 80% of low-permeable compounds (6) are correctly classified with almost all data series, except for the logP_o. The plausible explanation is related to the trend that the logP_o overestimates the permeability class for low-permeable compounds, because all compounds are not fully unionised in the GIT. This indicates that the membrane permeability measured by the PAMPA allows to estimate the permeability class for the BCS with high accuracy.

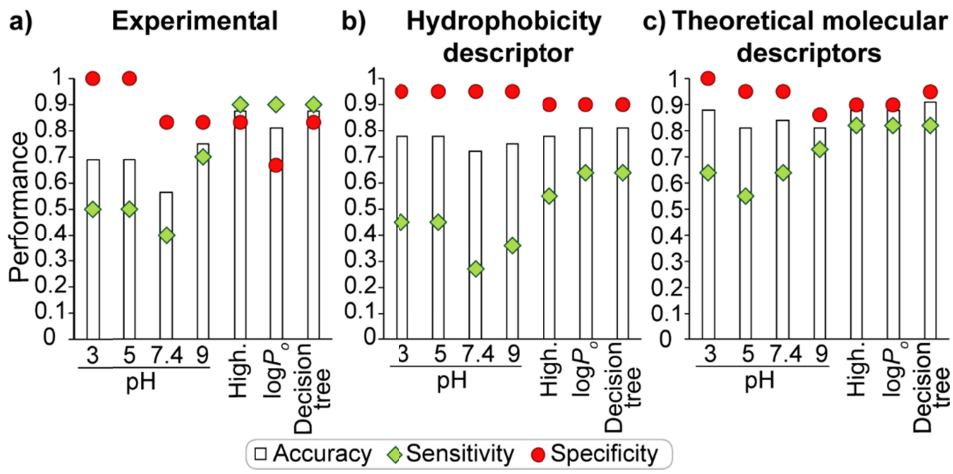


Figure 25. The comparison with the permeability class in the BCS for the U.S. FDA reference drug substances: (a) the experimental membrane permeability class, (b) the predicted membrane permeability class using the classification models with a hydrophobicity descriptor, and (c) the predicted membrane permeability class using the classification model with theoretical molecular descriptors (High. – log$P_{e_highest}$).

The comparison between the permeability class in the BCS and the predicted membrane permeability classes using the classification models with a hydrophobicity descriptor and theoretical molecular descriptors (Figure 25: b,c) shows that the variation of the accuracy over the data series is very small, but sensitivity and specificity vary in a wide range. The trends in the variation of sensitivity and specificity are similar to the experimental data (Figure 25: a). Prediction accuracies for the U.S. FDA reference drug substances across all data series show that the classification models with theoretical molecular descriptors perform better than the classification models with a hydrophobicity descriptor (Figure 25: b,c). This difference is due to the prediction quality for high-permeable compounds, i.e. the classification models with a hydrophobicity descriptor have lower sensitivity compared to the classification models with theoretical molecular descriptors. The prediction quality for low-permeable compounds is similarly good for the classification models with a hydrophobicity descriptor and theoretical molecular descriptors. The best performance characteristics were obtained using the classification models with theoretical molecular descriptors for the $\log P_{e\_highest}$ and the $\log P_o$. These classification models have higher performance because they took into account the full pH range in the GIT and descriptor content in the equations is more diverse.

In order to take the advantage of the four experimental and two derived membrane permeability data series and corresponding classification models, a simple decision trees were developed to improve the prediction confidence of the BCS permeability classes. Three decision trees have been developed: one for the experimental membrane permeability classes, and two for the predicted permeability classes using the classification models with a hydrophobicity descriptor and theoretical molecular descriptors. All decision trees (Figure 26) use the number of the predictions into the high permeability class over six models (#HighPrediction) as a criterion. This value must be at least three to classify a compound as high permeable. The decision tree with experimental membrane permeability classes show comparable results with the $\log P_{e\_highest}$ data (Figure 25: a). The decision tree with a hydrophobicity descriptor has the same performance characteristics as the model for the $\log P_o$ (Figure 25: b). The decision tree with theoretical molecular descriptors has higher accuracy (0.91) and specificity (0.9) than any individual models (Figure 25: c). Comparing the decision trees based on predictions from the classification models, the one with theoretical molecular descriptors has significantly higher accuracy (0.91 vs 0.81), sensitivity (0.82 vs 0.64), and specificity (0.95 vs 0.90) in contrast to the models with a hydrophobicity descriptor. This shows that the classification models relying on a hydrophobicity descriptor capture less structural features than theoretical molecular descriptors. Low-permeable compounds in the BCS are most likely misclassified, because they have a low solubility at pH-s, where the membrane permeability is high, which is decreasing absorption in the GIT, but the solubility is not explicitly considered in the developed classification models for the membrane permeability.

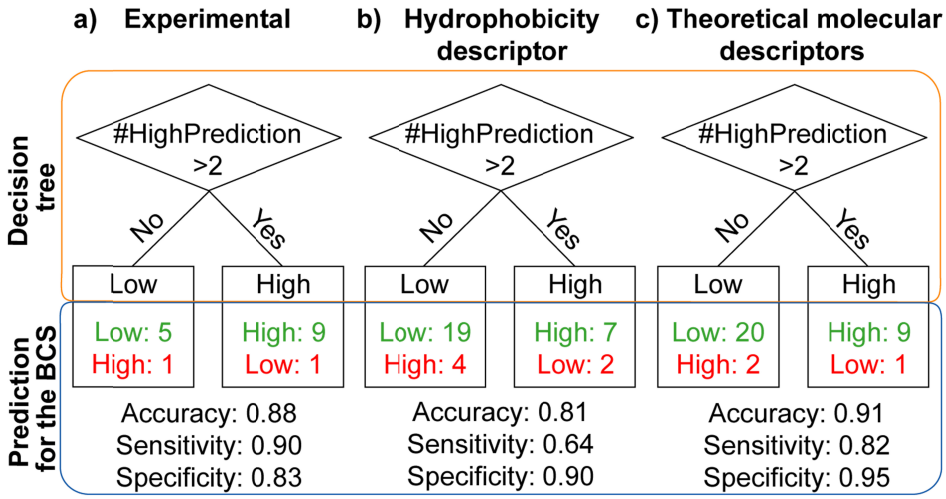


Figure 26. Decision trees for the experimental membrane permeability classes and the predictions from the classification models with a hydrophobicity descriptor or theoretical molecular descriptors.

Conclusively, although some misclassifications are present for both high- and low-permeable compounds, the decision trees predict the BCS permeability classes with equal or higher confidence in comparison to the individual classification models (Figure 25). Especially improvement is achieved when using the classification models with theoretical molecular descriptors. Consequently, the decision trees evaluated the consensus of six classification models and attributed more precisely the BCS permeability classes for the U.S. FDA reference drug substances. From this can be concluded that such an approach will help to apply these classification models and decision trees to predict the passive transport of new drug candidates with high reliability.

SUMMARY

Permeability of drug substances is an important component of the absorption in the gastrointestinal tract alongside solubility and is therefore an essential property of drug discovery and development studies. The permeability describes the movement of drug substances through the intestinal epithelium, which occurs mainly when drug substances are in unionised form. At the same time, it is known that about 80% of drug substances are ionisable in the fluctuating pH range of the GIT (pH ~2 to pH ~8). Therefore, it is important that the pH-permeability profiles of drug substances and drug substance candidates can be analysed using experimental and cheminformatics methods.

During the study, a systematic database of experimental pH-permeability profiles for 274 drug substances has been assembled. The pH-permeability profiles were measured with the parallel artificial membrane permeability assay (PAMPA) at four pH-s (3, 5, 7.4 and 9) introducing technological innovation to measure the time dependence in semi-high-throughput manner. The pH-permeability profiles show that different chemical classes (acids, bases, ampholytes and neutrals) do have characteristic pH-permeability profiles and they are related to the fraction of unionised species at concrete pH conditions. The experimental membrane permeability taking account the full pH range in the GIT is in good agreement with the human intestinal absorption, showing the ability of the artificial membrane method to describe the human intestinal absorption and a clear influence of pH on the efficiency of intestinal absorption.

The developed and externally validated QSAR models for the membrane permeability allow to predict the pH-permeability profiles, the highest membrane permeability over the pH range in the GIT, and the membrane permeability for the unionised form of the compound. The descriptors in the general QSAR models explain that the membrane permeability is mechanistically determined by the hydrophobicity and hydrogen bonding properties. The QSAR models for separate chemical classes provide more detailed view to the structural characteristics determining the membrane permeability. The membrane permeability for acidic and neutral drug substances is described by the hydrogen bond donor properties, for the basic drug substances by the distribution characteristics, and for the amphoteric drug substances by the hydrophobicity, hydrogen bonding, and the size and shape of the molecules. The developed QSAR models allow predicting the membrane permeability for drug substance candidates and provide an estimation for human intestinal absorption.

The developed and triple-validated classification models allow to evaluate permeability classes for specific pH (i.e. class-based pH-permeability profile), over the pH range in the GIT, and for the unionised drug substances. The classification models show that the permeability class can be determined using hydrophobicity descriptor or theoretical molecular descriptors, which describe the ionisation, polarity, and the size and complexity of the molecule. The

classification models were tested with the U.S. FDA reference drug substances for the high and low permeability classes in the BCS. This test gives good coherence between predicted membrane permeability class and the permeability classes in the BCS for the full pH range in the GIT. Further improvement in the identification of the BCS permeability classes was achieved using the decision tree that consolidated predictions from six classification models. As the result, 91% of the U.S. FDA reference drug substances were correctly classified for high and low permeability classes in the BCS.

In the conclusion, the membrane permeability is highly influenced by pH, which convincingly suggests that the pH range in the GIT is an important parameter to consider during the drug discovery process. Within this thesis the developed qualitative and quantitative prediction models for a wide pH range are explicitly validated and openly available in the QsarDB repository. Thus, the developed models can be used for predicting and analysing the pH-permeability profiles for drug candidates during the early stages of drug discovery. In addition, the novel decision trees allow predicting the permeability classes in the BCS with high confidence, and it can be used to improve the selection of the biowaivers for the bioavailability and bioequivalence studies. The developed models are not only useful for drug discovery, but can also be applied in other industries, like cosmetics, food, and industrial chemicals, to avoid animal testing, while estimating permeability in the GIT.

REFERENCES

[1] Prentis, R.A.; Lis, Y.; Walker, S.R. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964–1985). *Br. J. Clin. Pharmacol.* **1988**, *25*, 387–396.

[2] Cook, D.; Brown, D.; Alexander, R.; March, R.; Morgan, P.; Satterthwaite, G.; Pangalos, M.N. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **2014,** *13*, 419–431.

[3] Waring, M.J.; Arrowsmith, J.; Leach, A.R.; Leeson, P.D.; Mandrell, S.; Owen, R.M.; Pairaudeau, G.; Pennie, W.D.; Pickett, S.D.; Wang, J.; Wallace, O.; Weir, A. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **2015**, *14*, 475–486.

[4] Gleeson, M.P.; Leeson, P.D.; van de Waterbeemd, H. Physicochemical properties and compound quality. In *The handbook of medicinal chemistry: principles and practice*, Davis, A.M.; Ward, S., Eds.; Royal Society of Chemistry, 2015, pp. 1–31.

[5] Arrowsmith, J. Trial watch: phase II failures: 2008–2010. *Nat. Rev. Drug Discov.* **2011**, *10*, 328–329.

[6] Charman, W.N.; Porter, C.J.; Mithani, S.; Dressman, J.B. Physicochemical and physiological mechanisms for the effects of food on drug absorption: the role of lipids and pH. *J. Pharm. Sci.* **1997**, *86*, 269–282.

[7] Avdeef, A. *Absorption and drug development. Solubility, permeability, and charge state*, 2nd ed.; WILEY: Hoboken, New Jersey, 2012, pp. 1–498.

[8] DeSesso, J.M.; Jacobson, C.F. Anatomical and physiological parameters affecting gastrointestinal absorption in humans and rats. *Food Chem. Toxicol.* **2001**, *39*, 209–228.

[9] Daugherty, A.L.; Mrsny, R.J. Transcellular uptake mechanisms of the intestinal epithelial barrier: part one. *Pharm. Sci. Technol. To.* **1999**, *2*, 144–151.

[10] Waiver of *in vivo* bioavailability and bioequivalence studies for immediate-release solid oral dosage forms based on a biopharmaceutics classification system. Guidance for industry, **2017**, available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070246.pdf (accessed June 3th, 2019)

[11] Volpe, D.A. Variability in Caco-2 and MDCK cell-based intestinal permeability assays. *J. Pharm. Sci.* **2008**, *97*, 712–725.

[12] 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. *Eur. J. Pharm. Sci.* **2018**, *119*, 219–233.

[13] 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*, 1007–1010.

[14] Diukendjieva, A.; Tsakovska, I.; Alov, P.; Pencheva, T.; Pajeva, I.; Worth, A.P.; Madden, J.C.; Cronin, M.T.D. Advances in the prediction of gastrointestinal absorption: quantitative structure-activity relationship (QSAR) modelling of PAMPA permeability. *Computat. Toxicol.* **2019**, *10*, 51–59.

[15] Koziolek, M.; Grimm, M.; Schneider, F.; Jedamzik, P.; Sager, M.; Kühn, J.-P.; Siegmund, W.; Weitschies, W. Navigating the human gastrointestinal tract for oral drug delivery: uncharted waters and new frontiers. *Adv. Drug Deliver. Rev.* **2016**, *101*, 75–88.

[16] Balimane, P.V.; Chong, S.; Morrison, R.A. Current methodologies used for evaluation of intestinal permeability and absorption. *J. Pharmacol. Tox. Met.* **2000**, *44*, 301–312.

[17] Bergström, C.A.S.; Holm, R.; Jørgensen, S.A.; Andersson, S.B.E.; Artursson, P.; Beato, S.; Borde, A.; Box, K.; Brewster, M.; Dressman, J.; Feng, K.-I.; Halbert, G.; Kostewicz, E.; McAllister, M.; Muenster, U.; Thinnes, J.; Taylor, R.; Mullertz, A. Early pharmaceutical profiling to predict oral drug absorption: current status and unmet needs. *Eur. J. Pharm. Sci.* **2014**, *57*, 173–199.

[18] Sjögren, E.; Abrahamsson, B.; Augustijns, P.; Becker, D.; Bolger, M.B.; Brewster, M.; Brouwers, J.; Flanagan, T.; Harwoodh, M.; Heinen, C.; Holm, R.; Juretschke, H.-P.; Kubbinga, M.; Lindahl, A.; Lukacova, V.; Münster, U.; Neuhoff, S.; Nguyen, M. A.; van Peer, A.; Reppas, C.; Hodjegan, A.R.: Tannergren, C.; Weitschies, W.; Wilson, C.; Zane, P.; Lennernäs, H.; Langguth, P. *In vivo* methods for drug absorption – Comparative physiologies, model selection, correlations with *in vitro* methods (IVIVC), and applications for formulation/API/excipient characterization including food effects. *Eur. J. Pharm. Sci.* **2014**, *57*, 99–151.

[19] Riethorst, D.; Mols, R.; Duchateau, G.; Tack, J.; Brouwers, J.; Augustijns, P. Characterization of human duodenal fluids in fasted and fed state conditions. *J. Pharm. Sci.* **2016**, *105*, 673–681.

[20] Dahlgren, D.; Roos, C.; Lundqvist, A.; Abrahamsson, B.; Tannergren, C.; Hellström, P.M.; Sjögren, E.; Lennernäs, H. Regional intestinal permeability of three model drugs in human. *Mol. Pharm.* **2016**, *13*, 3013–3021.

[21] Ingels, F.; Deferme, S.; Destexhe, E.; Oth, M.; Van den Mooter, G.; Augustijns, P. Simulated intestinal fluid as transport medium in the Caco-2 cell culture model. *Int. J. Pharm.* **2002**, *232*, 183–192.

[22] Vertzoni, M.; Fotaki, N.; Nicolaides, E.; Reppas, C.; Kostewicz, E.; Stippler, E.; Leuner, C.; Dressman, J. Dissolution media simulating the intralumenal composition of the small intestine: physiological issues and practical aspects. *J. Pharm. Pharmacol.* **2004**, *56*, 453–462.

[23] Riethorst, D.; Brouwers, J.; Motmans, J.; Augustijns, P. Human intestinal fluid factors affecting intestinal drug permeation *in vitro*. *Eur. J. Pharm. Sci.* **2018**, *121*, 338–346.

[24] Lennernäs, H. Regional intestinal drug permeation: biopharmaceutics and drug development. *Eur. J. Pharm. Sci.* **2014**, *57*, 333–341.

[25] 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. *Pharm. Res.* **1995**, *12*, 413–420.

[26] Pade, V.; Stavchansky, S. Link between drug absorption solubility and permeability measurements in Caco-2 cells. *J. Pharm. Sci.* **1998**, *87*, 1604–1607.

[27] Yu, L.X.; Amidon, G.L.; Polli, J.E.; Zhao, H.; Mehta, M.U.; Conner, D.P.; Shah, V.P.; Lesko, L.J.; Chen, M.-L.; Lee, V.H.; Hussain, A.S. Biopharmaceutics classification system: the scientific basis for biowaiver extensions. *Pharm. Res.* **2002**, *19*, 921–925.

[28] Avdeef, A. Physicochemical profiling (solubility, permeability and charge state). *Curr. Top. Med. Chem.* **2001**, *1*, 277–351.
[29] Shore, P.A.; Brodie, B.B.; Hogben, C.A. The gastric secretion of drugs: a pH partition hypothesis. *J. Pharmacol. Exp. Ther.* **1957**, *119*, 361–369.
[30] Zimmermann, I. Determination of pK_a values from solubility data. *Int. J. Pharm.* **1982**, *13*, 57–65.
[31] Sieger, P.; Cui, Y.; Scheuerer, S. pH-dependent solubility and permeability profiles: a useful tool for prediction of oral bioavailability. *Eur. J. Pharm. Sci.* **2017**, *105*, 82–89.
[32] Varma, M.V.; Gardner, I.; Steyn, S.J.; Nkansah, P.; Rotter, C.J.; Whitney-Pickett, C.; Zhang, H.; Di, L.; Cram, M.; Fenner, K.S.; El-Kattan, A.F. pH-dependent solubility and permeability criteria for provisional biopharmaceutics classification (BCS and BDDCS) in early drug discovery. *Mol. Pharm.* **2012**, *9*, 1199–1212.
[33] Tsume, Y.; Mudie, D.M.; Langguth, P.; Amidon, G.E.; Amidon, G.L. The biopharmaceutics classification system: subclasses for *in vivo* predictive dissolution (IPD) methodology and IVIVC. *Eur. J. Pharm. Sci.* **2014**, *57*, 152–163.
[34] Tsume, Y.; Langguth, P.; Garcia-Arieta, A.; Amidon, G.L. *In silico* prediction of drug dissolution and absorption with variation in intestinal pH for BCS class II weak acid drugs: ibuprofen and ketoprofen. *Biopharm. Drug Dispos.* **2012**, *33*, 366–377.
[35] Fagerberg, J.H.; Tsinman, O.; Sun, N.; Tsinman, K.; Avdeef, A.; Bergström, C.A.S. Dissolution rate and apparent solubility of poorly soluble drugs in biorelevant dissolution media. *Mol. Pharm.* **2010**, *7*, 1419–1430.
[36] Augustijns, P.; Wuyts, B.; Hens, B.; Annaert, P.; Butler, J.; Brouwers, J. A review of drug solubility in human intestinal fluids: implications for the prediction of oral absorption. *Eur. J. Pharm. Sci.* **2014**, *57*, 322–332
[37] Dahan, A.; Miller, J.M.; Hilfinger, J.M.; Yamashita, S.; Yu, L.X.; Lennernäs, H.; Amidon, G.L. High-permeability criterion for BCS classification: segmental/pH dependent permeability considerations. *Mol. Pharm.* **2010**, *7*, 1827–1834.
[38] 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*, 597–614.
[39] Li, A.P. Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov. Today* **2001**, *6*, 357–366.
[40] Di, L.; Artursson, P.; Avdeef, A.; Ecker, G.F.; Faller, B.; Fischer, H.; Houston, J.B.; Kansy, M.; Kerns, E.H.; Krämer, S.D.; Lennernäs, H.; Sugano, K. Evidence-based approach to assess passive diffusion and carrier-mediated drug transport. *Drug Discov. Today* **2012**, *17*, 905–912.
[41] Brennan, M.B. Drug discovery – Filtering out failures early in the game. *Chem. Eng. News* **2000**, *78*, 63–73.
[42] Adson, A.; Raub, T.J.; Burton, P.S.; Barsuhn, C.L.; Hilgers, A.R.; Audus, K.L.; Ho, N.F. Quantitative approaches to delineate paracellular diffusion in cultured epithelial cell monolayers. *J. Pharm. Sci.* **1994**, *83*, 1529–1536.
[43] Wang, B.; Siahaan, T.; Soltero, R. *Drug delivery: principles and application,* WILEY: Hoboken, New Jersey, 2005, pp. 15–26.

[44] Palm, K.; Luthman, K.; Ros, J.; Gråsjö, J.; Artursson, P. Effect of molecular charge on intestinal epithelial drug transport: pH-dependent transport of cationic drugs. *J. Pharmacol. Exp. Ther.* **1999**, *291*, 435–443.
[45] 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*, 241–251.
[46] Thomae, A.V.; Wunderli-Allenspach, H.; Krämer, S.D. Permeation of aromatic carboxylic acids across lipid bilayers: the pH-partition hypothesis revisited. *Biophys. J.* **2005**, *89*, 1802–1811.
[47] Velický, M.; Tam, K.Y.; Dryfe, R.A.W. Permeation of a fully ionized species across a polarized supported liquid membrane. *Anal. Chem.* **2012**, *84*, 2541−2547.
[48] Fischer, H.; Kansy, M.; Avdeef, A.; Senner, F. Permeation of permanently positive charged molecules through artificial membranes – influence of physico-chemical properties. *Eur. J. Pharm. Sci.* **2007**, *31*, 32–42.
[49] Giacomini, K.M.; Huang, S.M.; Tweedie, D.J.; Benet, L.Z.; Brouwer, K.L.R.; Chu, X.; Dahlin, A.; Evers, R.; Fischer, V.; Hillgren, K.M.; Hoffmaster, K.A.; Ishikawa, T.; Keppler, D.; Kim, R.B.; Lee, C.A.; Niemi, M.; Polli, J.W.; Sugiyama, Y.; Swaan, P.W.; Ware, J.A.; Wright, S.H.; Yee, S.W.; Zamek-Gliszczynski, M.J.; Zhang, L. Membrane transporters in drug development. *Nat. Rev. Drug Discov.* **2010**, *9*, 215–236.
[50] Beumer, J.H.; Beijnen, J.H.; Schellens, J.H.M. Mass balance studies, with a focus on anticancer drugs. *Clin. Pharmacokinet.* **2006**, *45*, 33–58.
[51] Roig, T; Vinardell, M.P. Intestinal perfusion *in vivo* for the study of absorptive processes. *Comp. Biochem. Phys. A* **1991**, *98*, 3–7.
[52] Stappaerts, J.; Brouwers, J.; Annaert, P.; Augustijns, P. *In situ* perfusion in rodents to explore intestinal drug absorption: challenges and opportunities. *Int. J. Pharm.* **2015**, *478*, 665–681.
[53] Ungell, A.-L.; Nylander, S.; Bergstrand, S.; Sjöberg, Å.; Lennernäs, H. Membrane transport of drugs in different regions of the intestinal tract of the rat. *J. Pharm. Sci.* **1998**, *87*, 360–366.
[54] Luo, Z.; Liu, Y.; Zhao, B.; Tang, M.; Dong, H.; Zhang, L.; Lv, B.; Wei, L. *Ex vivo* and *in situ* approaches used to study intestinal absorption. *J. Pharmacol. Tox. Met.* **2013**, *68*, 208–216.
[55] Doke, S.K.; Dhawale, S.C. Alternatives to animal testing: a review. *Saudi Pharm. J.* **2015**, *23*, 223–229.
[56] Mols, R.; Deferme, S.; Augustijns, P. Sulfasalazine transport in *in-vitro*, *ex-vivo* and *in-vivo* absorption models: contribution of efflux carriers and their modulation by co-administration of synthetic nature-identical fruit extracts. *J. Pharm. Pharmacol.* **2005**, *57*, 1565–1573.
[57] Rogers, S.M.; Back, D.J.; Orme, M.L. Intestinal metabolism of ethinyloestradiol and paracetamol *in vitro*: studies using Ussing chambers. *Brit. J. Clin. Pharmaco.* **1987**, *23*, 727–734.
[58] Hilgers, A.R.; Conradi, R.A.; Burton, P.S. Caco-2 cell monolayers as a model for drug transport across the intestinal mucosa. *Pharm. Res.* **1990**, *7*, 902–910.
[59] Sambuy, Y.; De Angelis, I.; Ranaldi, G.; Scarino, M.L.; Stammati, A.; Zucco, F. The Caco-2 cell line as a model of the intestinal barrier: Influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biol. Toxicol.* **2005**, *21*, 1–26.

[60] Artursson, P. Epithelial transport of drugs in cell culture. I: a model for studying the passive diffusion of drugs over intestinal absorptive (Caco-2) cells. *J. Pharm. Sci.* **1990**, *79*, 476–482.
[61] Cho, M.J.; Thompson, D.P.; Cramer, C.T.; Vidmar, T.J.; Scieszka, J.F. The Madin Darby canine kidney (MDCK) epithelial cell monolayer as a model-cellular transport barrier. *Pharm. Res.* **1989**,*6*, 71–77.
[62] Avdeef, A.; Tam, K.Y. How well can the Caco-2/Madin−Darby Canine Kidney models predict effective human jejunal permeability? *J. Med. Chem.* **2010**, *53*, 3566–3584.
[63] Walter, E.; Janich, S.; Roessler, B.J.; Hilfinger, J.M.; Amidon, G.L. HT29-MTX/Caco-2 cocultures as an *in vitro* model for the intestinal epithelium: *in vitro-in vivo* correlation with permeability data from rats and humans. *J. Pharm. Sci.* **1996**, *85*, 1070–1076.
[64] Sun, D.; Lennernas, H.; Welage, L.S.; Barnett, J.L.; Landowski, C.P.; Foster, D.; Fleisher, D.; Lee, K.-D.; Amidon, G.L. Comparison of human duodenum and Caco-2 gene expression profiles for 12,000 gene sequences tags and correlation with permeability of 26 drugs. *Pharm. Res.* **2002**, *19*, 1400–1416.
[65] Ruell, J. Membrane-based drug assays: Permeability assays and oral absorption modelling can make the difference between drugs and dregs. *Mod. Drug Discov.* **2003**, 28–30.
[66] 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. *Eur. J. Med. Chem.* **2002**, *37*, 399–407.
[67] Kansy, M.; Avdeef, A.; Fischer, H. Advances in screening for membrane permeability: high-resolution PAMPA for medicinal chemists. *Drug Discov. Today Technol.* **2004**, *1*, 349–355.
[68] Faller, B. Artificial membrane assays to assess permeability. *Curr. Drug Metab.* **2008**, *9*, 886–892.
[69] Flaten, G.E.; Dhanikula, A.B.; Luthman, K.; Brandl, M. Drug permeability across a phospholipid vesicle based barrier: a novel approach for studying passive diffusion. *Eur. J. Pharm. Sci.* **2006**, *27*, 80–90.
[70] di Cagno, M.; Bibi, H.A.; Bauer-Brandl, A. New biomimetic barrier Permeapad™ for efficient investigation of passive permeability of drugs. *Eur. J. Pharm. Sci.* **2015**, *73*, 29–34.
[71] Bibi, H.A.; di Cagno, M.; Holm, R.; Bauer-Brandl, A. Permeapad™ for investigation of passive drug permeability: the effect of surfactants, co-solvents and simulated intestinal fluids (FaSSIF and FeSSIF). *Int. J. Pharm.* **2015**, *493*, 192–197.
[72] Berben, P.; Brouwers, J.; Augustijns, P. Assessment of passive intestinal permeability using an artificial membrane insert system. *J. Pharm. Sci.* **2018**, *107*, 250–256.
[73] Montal, M.; Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 3561–3566.
[74] Thompson, M.; Lennox, R.B.; McClelland, R.A. Structure and electrochemical properties of microfiltration filter-lipid membrane systems. *Anal. Chem.* **1982**, *54*, 76–81.

[75] Cools, A.A.; Janssen, L.H.M. Influence of sodium ion-pair formation on transport kinetics of warfarin through octanol-impregnated membranes. *J. Pharm. Pharmacol.* **1983**, *35*, 689–691.
[76] Camenisch, G.; Folkers, G.; van de Waterbeemd, H. Comparison of passive drug transport through Caco-2 cells and artificial membranes. *Int. J. Pharm.* **1997**, *147*, 61–70.
[77] Wohnsland, F.; Faller, B. High-throughput permeability pH profile and high-throughput alkane/water logP with artificial membranes. *J. Med. Chem.* **2001**, *44*, 923–930.
[78] Avdeef, A.; Strafford, M.; Block, E.; Balogh, M.P.; Chambliss, W.; Khan, I. Drug absorption *in vitro* model: filter-immobilized artificial membranes: 2. Studies of the permeability properties of lactones in *Piper methysticum* Forst. *Eur. J. Pharm. Sci*. **2001**, *14*, 271–280.
[79] Sugano, K.; Hamada, H.; Machida, M.; Ushio, H.; Saitoh, K.; Terada, K. Optimized conditions of bio-mimetic artificial membrane permeation assay. *Int. J. Pharm.* **2001**, *228*, 181–188.
[80] 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: 7. Comparing rat *in situ*, Caco-2, and PAMPA permeability of fluoroquinolones. *Eur. J. Pharm. Sci.* **2004**, *21*, 429–441.
[81] Chen, X.; Murawski, A.; Patel, K.; Crespi, C.L.; Balimane, P.V. A novel design of artificial membrane for improving the PAMPA model. *Pharm. Res.* **2008**, *25*, 1511–1520.
[82] Di, L.; Kerns, E.H.; Fan, K.; McConnell, O.J.; Carter, G.T. High throughput artificial membrane permeability assay for blood-brain barrier. *Eur. J. Med. Chem.* **2003**, *38*, 223–232.
[83] 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*, 3948–3954.
[84] Sinkó, B.; Garrigues, T.M.; Balogh, G.T.; Nagy, Z.K.; Tsinman, O.; Avdeef, A.; Takács-Novák, K. Skin-PAMPA: a new method for fast prediction of skin penetration. *Eur. J. Pharm. Sci.* **2012**, *45*, 698–707.
[85] Dargó, G.; Vincze, A.; Müller, J.; Kiss, H.J.; Nagy, Z.Z.; Balogh, G.T. Corneal-PAMPA: a novel, non-cell-based assay for prediction of corneal drug permeability. *Eur. J. Pharm. Sci.* **2019**, *128*, 232–239.
[86] 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*, 189–196.
[87] Avdeef, A.; Tsinman, O. PAMPA – a drug absorption in vitro model: 13. Chemical selectivity due to membrane hydrogen bonding: in combo comparison of HDM-, DOPC-, and DS-PAMPA models. *Eur. J. Pharm. Sci.* **2006**, *28*, 43–50.
[88] Seo, P.R.; Teksin, Z.S.; Kao, J.P.Y.; Polli, J.E. Lipid composition effect on permeability across PAMPA. *Eur. J. Pharm. Sci.* **2006**, *29*, 259–268.
[89] Abraham M.H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **1993,** *22*, 73–83.
[90] Manallack, D.T. The acid-base profile of a contemporary set of drugs: implications for drug discovery. *SAR QSAR Environ. Res.* **2009**, *20*, 611–655.

[91] Fairstein, M.; Swissa, R.; Dahan, A. Regional-dependent intestinal permeability and BCS classification: elucidation of pH-related complexity in rats using pseudoephedrine. *AAPS J.* **2013**, *15*, 589–597.
[92] Velický, M.; Bradley, D.F.; Tam, K.Y.; Dryfe, R.A.W. *In situ* artificial membrane assay under hydrodynamic control: permeability-pH profiles of warfarin and verapamil. *Pharm. Res.* **2010**, *27*, 1644–1658.
[93] Liu, H.; Sabus, C.; Carter, G.T.; Du, C.; Avdeef, A.; Tichler, M. *In vitro* permeability of poorly aqueous soluble compounds using different solubilizers in the PAMPA assay with liquid chromatography/mass spectrometry detection. *Pharm. Res.* **2003**, *20*, 1820–1826.
[94] Bendels, S.; Tsinman, O.; Wagner, B.; Lipp, D.; Parrilla, I.; Kansy, M.; Avdeef, A. PAMPA – excipient classification gradient map. *Pharm. Res.* **2006**, *23*, 2525–2535.
[95] Ruell, J.A.; Tsinman, O.; Avdeef, A. Acid-base cosolvent method for determining aqueous permeability of amiodarone, itraconazole, tamoxifen, terfenadine and other very insoluble molecules. *Chem. Pharm. Bull.* **2004**, *52*, 561–565.
[96] Avdeef, A.; Nielsen, P.; Du, C. Method and apparatus for improving in vitro measurement of membrane permeability of chemical compounds. Patent US20030219716A1, 2003.
[97] Avdeef, A.; Bendels, S.; Di, L.; Faller, B.; Kansy, M.; Sugano, K.; Yamauchi, Y. PAMPA – Critical factors for better predictions of absorption. *J. Pharm. Sci.* **2007**, *96*, 2893–2909.
[98] 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. *Eur. J. Pharm. Sci.* **2004**, *22*, 365–374.
[99] Ruell, J.A.; Tsinman, K.L.; Avdeef, A. PAMPA—a drug absorption *in vitro* model: 5. Unstirred water layer in iso-pH mapping assays and pK_a^{flux}—optimized design (pOD-PAMPA). *Eur. J. Pharm. Sci.* **2003**, *20*, 393–402.
[100] Avdeef, A. The rise of PAMPA. *Expert Opin. Drug Metab. Toxicol.* **2005**, *1*, 325–342.
[101] Galinis-Luciani, D.; Nguyen, L.; Yazdanian, M. Is PAMPA a useful tool for discovery? *J. Pharm. Sci.* **2007**, *96*, 2886–2892.
[102] Miret, S.; Abrahamse, L.; de Groene, E.M. Comparison of *in vitro* models for the prediction of compound absorption across the human intestinal mucosa. *J. Biomol. Screen.* **2004**, *9*, 598–606.
[103] Hwang, K.-K.; Martin, N.E.; Jiang, L. Permeation prediction of M100240 using the parallel artificial membrane permeability assay. *J. Pharm. Pharmaceut. Sci.* **2003**, *6*, 315–320.
[104] Balimane, P.V.; Han, Y.-H.; Chong, S. Current industrial practices of assessing permeability and P-glycoprotein interaction. *AAPS J.* **2006**, *8*, E1–E13.
[105] Obata, K.; Sugano, K.; Machida, M.; Aso, Y. Biopharmaceutics classification by high throughput solubility assay and PAMPA. *Drug Dev. Ind. Pharm.* **2004**, *30*, 181–185.
[106] Ano, R.; Kimura, Y.; Shima, M.; Matsuno, R.; Ueno, T.; Akamatsu, M. Relationships between structure and high-throughput screening permeability of peptide derivatives and related compounds with artificial membranes: application to prediction of Caco-2 cell permeability. *Bioorg. Med. Chem.* **2004**, *12*, 257–264.

[107] Wu, Y.; Hu, M.; Yang, L.; Li, X.; Bian, J.; Jiang, F.; Sun, H.; You, Q.; Zhang, X. Novel natural-product-like caged xanthones with improved druglike properties and *in vivo* antitumor potency. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 2584–2588.

[108] Kawada, H.; Ebiike, H.; Tsukazaki, M.; Yamamoto, S.; Koyama, K.; Nakamura, M.; Morikami, K.; Yoshinari, K.; Yoshida, M.; Ogawa, K.; Shinma, N.; Tsukuda, T.; Ohwada, J. Modification of a dihydropyrrolopyrimidine phosphoinositide 3-kinase (PI3K) inhibitor to improve oral bioavailability. *Bioorg. Med. Chem.* **2015**, *23*, 7650–7660.

[109] Flanders, Y.; Dumas, S.; Caserta, J.; Nicewonger, R.; Baldino, M.; Lee, C.-S.; Baldino, C.M. A versatile synthesis of novel pan-PIM kinase inhibitors with initial SAR study. *Tetrahedron Lett.* **2015**, *56*, 3186–3190.

[110] Kim, J.; Chin, J.; Im, C.Y.; Yoo, E.K.; Woo, S.; Hwang, H.J.; Cho, J.-H.; Seo, K.-A.; Song, J.; Hwang, H.; Kim, K.-H.; Kim, N.D.; Yoon, S.K.; Jeon, J.-H.; Yoon, S.-Y.; Jeon, Y.H.; Choi, H.-S.; Lee, I.-K.; Kim, S.H.; Cho, S.J. Synthesis and biological evaluation of novel 4-hydroxytamoxifen analogs as estrogen-related receptor gamma inverse agonists. *Eur. J. Med. Chem.* **2016**, *120*, 338–352.

[111] Ikuma, Y.; Hochigai, H.; Kimura, H.; Nunami, N.; Kobayashi, T.; Uchiyama, K.; Umezome, T.; Sakurai, Y.; Sawada, N.; Tadano, J.; Sugaru, E.; Ono, M.; Hirose, Y.; Nakahira, H. Discovery of 3H-imidazo[4,5-c]quinolin-4(5H)-ones as potent and selective dipeptidyl peptidase IV (DPP-4) inhibitors: use of a carboxylate prodrug to improve bioavailability. *Bioorg. Med. Chem.* **2015**, *23*, 779–790.

[112] Grosche, P.; Sirockin, F.; Sweeney, A.M.; Ramage, P.; Erbel, P.; Melkko, S.; Bernardi, A.; Hughes, N.; Ellis, D.; Combrink, K.D.; Jarousse, N.; Altmann, E. Structure-based design and optimization of potent inhibitors of the adenoviral protease. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 438–443.

[113] Savić, J.; Dobričić, V.; Nikolic, K.; Vladimirov, S.; Dilber, S.; Brborić, J. *In vitro* prediction of gastrointestinal absorption of novel β-hydroxy-β-arylalkanoic acids using PAMPA technique. *Eur. J. Pharm. Sci.* **2017**, *100*, 36–41.

[114] Yamaki, S.; Suzuki, D.; Fujiyasu, J.; Neya, M.; Nagashima, A.; Kondo, M.; Akabane, T.; Kadono, K.; Moritomo, A.; Yoshihara, K. Synthesis and structure activity relationships of glycine amide derivatives as novel Vascular Adhesion Protein-1 inhibitors. *Bioorg. Med. Chem.* **2017**, *25*, 187–201.

[115] Zhang, X.; Glunz, P.W.; Johnson, J.A.; Jiang, W.; Jacutin-Porte, S.; Ladziata, V.; Zou, Y.; Phillips, M.S.; Wurtz, N.R.; Parkhurst, B.; Rendina, A.R.; Harper, T.M.; Cheney, D.L.; Luettgen, J.M.; Wong, P.C.; Seiffert, D.; Wexler, R.R.; Priestley, E.S. Discovery of a highly potent, selective, and orally bioavailable macrocyclic inhibitor of blood coagulation factor VIIa–tissue factor complex. *J. Med. Chem.* **2016**, *59*, 7125–7137.

[116] Tzvetkov, N.T.; Stammler, H.-G.; Neumann, B.; Hristova, S.; Antonov, L.; Gastreich, M. Crystal structures, binding interactions, and ADME evaluation of brain penetrant N-substituted indazole-5-carboxamides as subnanomolar, selective monoamine oxidase B and dual MAO-A/B inhibitors. *Eur. J. Med. Chem.* **2017**, *127*, 470–492.

[117] Furukawa, A.; Townsend, C.E.; Schwochert, J.; Pye, C.R.; Bednarek, M.A.; Lokey, R.S. Passive membrane permeability in cyclic peptomer scaffolds is robust to extensive variation in side chain functionality and backbone geometry. *J. Med. Chem.* **2016**, *59*, 9503–9512.

[118] Frings, M.; Bolm, C.; Blum, A.; Gnamm, C. Sulfoximines from a medicinal chemist's perspective: physicochemical and *in vitro* parameters relevant for drug discovery. *Eur. J. Med. Chem.* **2017**, *126*, 225–245.
[119] Diukendjieva, A.; Alov, P.; Tsakovska, I.; Pencheva, T.; Richarz, A.; Kren, V.; Cronin, M.T.D.; Pajeva, I. *In vitro* and *in silico* studies of the membrane permeability of natural flavonoids from *Silybum marianum* (L.) Gaertn. and their derivatives. *Phytomedicine* **2019**, *53*, 79–85.
[120] Huque, F.T.T.; Box, K.; Platts, J.A.; Comer, J. Permeability through DOPC/ dodecane membranes: measurement and LFER modelling. *Eur. J. Pharm. Sci.* **2004**, *23*, 223–232.
[121] Fujikawa, M.; Ano, R.; Nakao, K.; Shimizu, R.; Akamatsu, M. Relationships between structure and high-throughput screening permeability of diverse drugs with artificial membranes: application to prediction of Caco-2 cell permeability. *Bioorg. Med. Chem.* **2005**, *13*, 4721–4732.
[122] Fujikawa, M.; Nakao, K.; Shimizu, R.; Akamatsu, M. QSAR study on permeability of hydrophobic compounds with artificial membranes. *Bioorg. Med. Chem.* **2007**, *15*, 3756–3767.
[123] Verma, R.P.; Hansch, C.; Selassie, C.D. Comparative QSAR studies on PAMPA/modified PAMPA for high throughput profiling of drug absorption potential with respect to Caco-2 cells and human intestinal absorption. *J. Comput. Aid. Mol. Des.* **2007**, *21*, 3–22.
[124] Acharya, C.; Seo, P.R.; Polli, J.E.; MacKerell, A.D. Computational model for predicting chemical substituent effects on passive drug permeability across parallel artificial membranes. *Mol. Pharm.* **2008**, *5*, 818–828.
[125] 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. Aid. Mol. Des.* **2009**, *23*, 309–319.
[126] Tulp, I.; Sild, S.; Maran, U. Relationship between structure and permeability in artificial membranes: theoretical whole molecule descriptors in development of QSAR Models. *QSAR Comb. Sci.* **2009**, *8*, 811–814.
[127] Karelson, M.; Karelson, G.; Tamm, T.; Tulp, I.; Jänes, J.; Tämm, K.; Lomaka, A.; Savchenko, D.; Dochev, D. QSAR study of pharmacological permeabilities. *ARKIVOC* **2009**, *ii*, 218–238.
[128] Wang, C.K.; Northfield, S.E.; Swedberg, J.E.; Colless, B.; Chaousis, S.; Price, D.A.; Liras, S.; Craik, D.J. Exploring experimental and computational markers of cyclic peptides: charting islands of permeability. *Eur. J. Med. Chem.* **2015**, *97*, 202–213.
[129] 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. *Bioorg. Med. Chem.* **2017**, *25*, 1266–1276.
[130] Gordon, S.; Daneshian, M.; Bouwstra, J.; Caloni, F.; Constant, S.; Davies, D.E.; Dandekar, G.; Guzman, C.A.; Fabian, E.; Haltner, E.; Hartung, T.; Hasiwa, N.; Hayden, P.; Kandarova, H.; Khare, S.; Krug, H.F.; Kneuer, C.; Leist, M.; Lian, G.; Marx, U.; Metzger, M.; Ott, K.; Prieto, P.; Roberts, M.S.; Roggen, E.L.; Tralau, T.; van den Braak, C.; Walles, H.; Lehr, C.M. Non-animal models of epithelial barriers (skin, intestine and lung) in research, industrial applications and regulatory toxicology, ALTEX **2015,** *32*, 327–378.
[131] Wiki-pKa^{TM} database; available at http://www.in-adme.com/wiki_pka.php. (accessed June 3th, 2019)

[132] pCEL-X™ software, version 4.03, in-ADME Research Inc.; available at http://www.in-adme.com/pcel_x.html (accessed June 3th, 2019)
[133] JChem for Excel, ChemAxon; available at http://www.chemaxon.com (accessed June 3th, 2019)
[134] Newby, D.; Freitas, A.A.; Ghafourian, T. Decision trees to characterise the roles of permeability and solubility on the prediction of oral absorption. *Eur. J. Med. Chem.* **2015**, *90*, 751–765.
[135] SciFinder®, American Chemical Society, available at http://scifinder.cas.org/ (accessed June 3th, 2019)
[136] Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672.
[137] DrugBank database, available at https://www.drugbank.ca/ (downloaded: January 27th, 2016, accessed June 3th, 2019)
[138] Karelson, M. *Molecular descriptors in QSAR/QSPR*, WILEY: New York, 2000, pp. 141–383.
[139] Ruusmann, V.; Sild, S.; Maran, U. QSAR DataBank - an approach for the digital organization and archiving of QSAR model information. *J. Cheminf.* **2014**, *6*, 25.
[140] Ruusmann, V.; Sild, S.; Maran, U. QSAR DataBank repository: open and linked qualitative and quantitative structure–activity relationship models. *J. Cheminf.* **2015**, *7*, 32.
[141] QsarDB repository, available at http://qsardb.org/ (accessed June 3th, 2019)
[142] Bolton, E.E.; Wang, Y.; Thiessen, P.A.; Bryant, S.H. PubChem: integrated platform of small molecules and biological activities, In: *Annual reports in computational chemistry*, Volume 4, American Chemical Society, Washington, DC, 2008, pp. 217–241.
[143] PubChem database, available at http://pubchem.ncbi.nlm.nih.gov (accessed June 3th, 2019)
[144] MacroModel version 9, Schrödinger, Inc., Portland, 2000, available at www.schrodinger.com (accessed June 3th, 2019)
[145] Halgren, T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem*. **1996**, *17*, 490–519.
[146] Halgren, T.A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem*. **1996**, *17*, 520–552.
[147] Halgren, T.A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem*. **1996**, *17*, 553–586
[148] Halgren, T.A.; Nachbar, R.B. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem*. **1996**, *17*, 587–615
[149] Halgren, T.A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem*. **1996**, *17*, 616–641.
[150] Halgren, T.A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, *20*, 730–748.
[151] Chang, G.; Guida, W.C.; Still, W.C. An internal-coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.

[152] Saunders, M.; Houk, K.N.; Wu, Y.D.; Still, W.C.; Lipton, M.; Chang, G.; Guida, W.C. Conformations of cycloheptadecane. A comparison of methods for conformational searching. *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.
[153] Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
[154] Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*. **1985**, *107*, 3902–3909.
[155] Baker, J. An algorithm for the location of transition-states. *J. Comput. Chem.* **1986**, *7*, 385–395.
[156] Stewart, J. J. P. *MOPAC Program Package 7.0*, QCPE Bloomiagton, IN. USA, 1993.
[157] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
[158] CODESSA PRO, version 1.0, available at http://www.codessa-pro.com/ (accessed June 3th, 2019)
[159] CODESSA PRO User's manual, University of Florida, U.S.A 2005.
[160] PaDEL-Descriptor, version 2.21, available at http://www.yapcwsoft.com/dd/padeldescriptor/ (accessed June 3th, 2019)
[161] Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474.
[162] Syracuse research corporation. *Physical/chemical Property Database–(PHYSPROP)*, SRC Environmental Science Centre: Syracuse, NY, USA, 1994.
[163] Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* **2007**, *47*, 2140–2148.
[164] XlogP3, version 3.2.2, available at http://www.sioc-ccbg.ac.cn/?p=42&software=xlogp3 (accessed June 3th, 2019)
[165] Piir, G.; Sild, S.; Maran, U. Classifying bio-concentration factor with random forest algorithm, influence of the bio-accumulative vs. non-bio-accumulative compound ratio to modelling result, and applicability domain for random forest model. *SAR QSAR Environ. Res.* **2014**, *25*, 967–981.
[166] Newby, D.; Freitas, A.A.; Ghafourian, T. Coping with unbalanced class data sets in oral absorption models. *J. Chem. Inf. Model.* **2013**, *53*, 461–474.
[167] Chen, J.J.; Tsai, C.A.; Young, J.F.; Kodell, R.L. Classification ensembles for unbalanced class sizes in predictive toxicology. *SAR QSAR Environ. Res.* **2005**, *16*, 517–529.
[168] Atkinson, A.C. Plots, transformation, regression: an introduction to graphical methods of diagnostic regression analysis, 1985, Oxford University Press, Oxford, UK.
[169] Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR applicability domain estimation by projection of the training set in descriptor space: a review. *Altern. Lab. Anim.* **2005**, *33*, 445–459.
[170] Tropsha, A.; Gramatica, P.; Gombar, V.K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.

[171] Netzeva, T.I.; Worth, A.P.; Aldenberg, T.; Benigni, R.; Cronin, M.T.D.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G.Y.; Perkins, R.; Roberts, D.W.; Schultz, T.W.; Stanton, D.T.; van de Sandt, J.J.M.; Tong, W.; Veith, G.; Yang, C. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Altern. Lab. Anim.* **2005**, *33*, 155–173.

[172] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, available at https://www.R-project.org/ (accessed December 13th, 2016)

[173] Cox, D. R. The regression analysis of binary sequences. *J. Royal Statistical Society* 1958, *20*, 215–242.

[174] Pearson, K. *Mathematical contributions to the theory of evolution. XIII. On the theory of contingency and its relation to association and normal correlation*, London, Dulau and Co., 1904.

[175] Suenderhauf, C.; Hammann, F.; Maunz, A.; Helma, C.; Huwyler, J. Combinatorial QSAR modeling of human intestinal absorption. *Mol. Pharm.* **2011**, *8*, 213–224.

[176] Todeschini, R.; Consonni, V. *Molecular descriptors for chemoinformatics*; 2nd ed.; WILEY-VCH, Weinheim, Germany, 2009.

[177] Rohrbaugh, R.H.; Jurs, P.C. Molecular shape and the prediction of high-performance liquid chromatographic retention indexes of polycyclic aromatic hydrocarbons. *Anal. Chem.* **1987**, *59*, 1048–1054.

[178] Maran, U.; Sild, S.; Tulp, I.; Takkis, K.; Moosus, M. Molecular descriptors from two-dimensional chemical structure. In *In silico toxicology: principles and applications*; Cronin, M.T.D.; Madden, J.C., Eds.; Royal Society of Chemistry, 2010; pp. 148–192.

[179] Stanton, D.T.; Jurs, P.C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. *Anal. Chem.* **1990**, *62*, 2323–2329.

[180] Roy, K.; Das, R.N. QSTR with extended topochemical atom (ETA) indices. 15. Development of predictive models for toxicity of organic chemicals against fathead minnow using second-generation ETA indices. *SAR QSAR Environ. Res.* **2012**, *23*, 125–140.

[181] CDK. https://cdk.github.io/cdk/1.5/docs/api/org/openscience/cdk/qsar/descriptors/molecular/HBondDonorCountDescriptor.html (accessed March 15th, 2017)

[182] Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.

SUMMARY IN ESTONIAN

Raviainete efektiivse membraaniläbitavuse pH profiilide eksperimentaalne uurimine ja modelleerimine

Raviaine läbitavus koos lahustuvusega on seedetrakti absorptsiooni oluline komponent ja on seetõttu ravimite väljatöötamise uuringute tähtis osa. Läbitavus kirjeldab ühendite liikumist läbi seedetrakti epiteelkihi, mis peamiselt toimub kui raviaine on mitteioniseerunud kujul. Samal ajal on teada, et ligikaudu 80% raviainetest ioniseeruvad seedetrakti muutuvas pH-vahemikus, mis varieerub alates väga happelisest pH-st (~2), kuni aluselise pH-ni (~8). Seetõttu on oluline uurida eksperimentaalselt ja keemiainformaatika meetodeid kasutades raviainete membraaniläbitavuse sõltuvust pH-st.

Töö käigus koostati süsteemne andmebaas, mis sisaldab 274 raviaine eksperimentaalseid membraaniläbitavuse pH profiile. Membraaniläbitavuse pH profiilid mõõdeti tehislikul membraanil põhineva läbitavuse määramise katsega (PAMPA) neljal pH-l (pH 3, 5, 7.4 ja 9), mida täiendati tehnoloogiliselt innovaatilise ajast sõltuvuse määramise lähenemisega. Mõõtmiste tulemused näitavad, et erinevatel aineklassidel (happed, alused, amfolüüdid ja neutraalid) on iseloomulikud pH-membraaniläbitavuse profiilid, mis on seotud mitteioniseeritud osakeste hulgaga lahuses kindlal pH-l. Eksperimentaalsed membraaniläbitavuse väärtused laias pH-vahemikus on heas kooskõlas inimese seedetrakti absorptsiooniga, mis näitab, et tehisliku membraani läbitavus on rakendatav seedetrakti absorptsiooni kirjeldamiseks ja seedetrakti absorptsiooni efektiivsus on selgelt mõjutatud pH-st.

Uurimise käigus arendati ja sõltumatult valideeriti kvantitatiivseid struktuur-aktiivsus sõltuvuse (QSAR) prognoosmudelid, mida on võimalik kasutada membraaniläbitavuse pH profiilide, kõrgeima membraaniläbitavuse ja mitteioniseerunud ühendi membraaniläbitavuse prognoosimiseks. Arendatud QSAR mudelid näitavad, et membraaniläbitavus on mõjutatud raviaine hüdrofoobsusest ja nende võimest anda vesiniksidemeid. Erinevate aineklasside QSAR mudelid viitasid, et membraaniläbitavus hapete ja neutraalsete ühendite korral sõltub vesiniksideme doonori omadustest, aluste puhul hüdrofoobsusest ja amfolüütide korral hüdrofoobsusest, vesiniksideme andmise võimest ja molekuli suurusest ning kujust. Arendatud prognoosmudelid on kasutatavad uute raviainete membraaniläbitavuse prognoosimiseks ja seedetrakti absorptsiooni hindamiseks.

Arendatud ja kolmekordselt valideeritud klassifikatsioonimudeleid saab kasutada kõrge ja madala membraaniläbitavuse hindamiseks erinevatel pH-del, üle kogu pH vahemiku seedetraktis ja mitteioniseerunud ühendile. Klassifikatsioonimudelid näitavad, et membraaniläbitavuse klassi saab määrata kasutades struktuuri parameetreid, mis kirjeldavad hüdrofoobsust või kombineerides teoreetilisi molekulaar-deskriptoreid, mis kirjeldavad ionisatsiooni, polaarsust ja molekuli suurust ja kuju. Praktikas rakendati klassifikatsiooni mudeleid referentsühenditele, mis on määratud USA Toidu- ja Ravimiameti poolt bio-

farmatseutilise klassifikatsiooni läbitavuse klassi hindamiseks. Prognoositud membraaniläbitavuse klassid olid heas vastavuses biofarmatseutilise klassifikatsiooni läbitavuse klassiga kui võeti arvesse kogu pH vahemikku seedetraktis. Kõrgema prognoosimise täpsuse saavutamiseks kasutati kuue mudeli prognoositud väärtusi otsustuspuus. Selle tulemusena klassifitseeriti 91% referentsühenditest korrektselt biofarmatseutilise klassifikatsiooni läbitavuse klassi.

Kokkuvõtteks, membraaniläbitavus on märkimisväärselt mõjutatud pH-st, mis selgelt näitab, et oluline on võtta arvesse pH mõju läbitavusele varajastes ravimiarenduse etappides. Arendatud kvalitatiivsed ja kvantitatiivsed prognoosmudeleid laias pH-vahemikus on põhjalikult valideeritud ja on avalikult kättesaadavad QsarDB andmebaasis. Seega saab neid mudeleid rakendada ravimikandidaatide membraaniläbitavuse pH profiilide prognoosimiseks varajases ravimiarenduse etappides. Erinevaid klassifitseerimise mudeleid koondav otsustuspuu võimaldab hinnata biofarmatseutilise klassifikatsiooni läbitavuse klasse ja see võimaldab efektiivsemalt valida ühendeid biokättesaadavuse uuringutesse. Arendatud prognoosmudelid ei leia rakendust mitte ainult ravimiarenduses, vaid neid mudeleid saab kasutada ka toidu-, kosmeetika- kui ka keemiatööstuses, et hinnata ühendite läbitavust seedetraktis ilma loomkatseteta.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Uko Maran for his guidance during my studies. I am also thankful to Dr. Sulev Sild for his excellent collaboration as co-author of the publication. I am grateful to Dr. Alex Avdeef for his support and helpful discussions. I would like to acknowledge colleagues from my research group for the kind help and great atmosphere. Finally, special thanks go to my parents, brother and his family, and friends for supporting me during the study years.

I acknowledge financial support from the Estonian Ministry for Education and Research [grant numbers SF0140031Bs09 and IUT34-14], the Estonian Research Council [grant number 7709], and the European Union Regional Development Fund [grant numbers 3.2.1201.13-0021 and TK143].

This work has been partially supported by the Graduate School of Functional Materials and Technologies (FMTDK) receiving funding from the European Social Fund [grant number 1.2.0401.09-0079] and the European Regional Development Fund [grant number 2014-2020.4.01.16-0027].

Scholarships for the international conferences and training schools have had a significant impact on my work. Travel scholarships have been received from following grant programs: Kristjan Jaak Scholarship for short study visit managed by Archimedes Foundation in collaboration with the Ministry of Education and Research; Dora Plus scholarships for short study visits funded by the European Regional Development Fund and the Republic of Estonia; COST Action “European Network on Understanding Gastrointestinal Absorption-related Processes” (CA16205); COST Action “Targeted chemotherapy towards diseases caused by endoparasites” (CM1307); International Pharmaceutical Federation (FIP); and the American Society for Cellular and Computational Toxicology (ASCCT).

PUBLICATIONS

CURRICULUM VITAE

Name	Mare Oja
Date of birth	July 6th, 1989, Tartu, Estonia
Citizenship	Estonian
Address	Institute of Chemistry, University of Tartu, Ravila 14A, 50411, Tartu, Estonia
E-mail	mare.oja@ut.ee

Education

2013–…	University of Tartu, Molecular Engineering, PhD student
2011–2013	University of Tartu, MSc in Chemistry
2008–2011	University of Tartu, BSc in Chemistry
2005–2008	Hugo Treffner Gymnasium

Professional employment

09.2017–…	University of Tartu, Institute of Chemistry, Junior Research Fellow (1.0)
01.–08.2017	University of Tartu, Institute of Chemistry, Chemist (0.50)
2013–2015	University of Tartu, Institute of Chemistry, Chemist (0.50)

R&D related managerial and administrative work

2013−…	member, International Association of Physical Chemists (IAPC)

Professional self-improvement

2017	3rd International Summer School on Drug Development (ISSDD-3), Zagreb, Croatia, September 1–3, 2017.
2016	Short Introduction in Drug Development, Zhuhai, China, August 26, 2016.
2015	Targeted chemotherapy towards diseases caused by endoparasites, Fraunhofer-IME ScreeningPort (COST Action CM1307), Hamburg, Germany, September 28 – October 1, 2015.
2013	1st European Summer School on Formulation Development (ESSFD-1), Dubrovnik, Croatia, September 20–22, 2013.

Awards and scholarships

2016	Best young researcher oral presentation, 17th International Conference on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2016), Miami Beach, USA.
2015	Best young researcher poster presentation, 8th International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (CMTPI2015), Chios, Greece.

2011 Prize to Bachelor's thesis "Applying molecular descriptors in the analysis of structural space of drugs and non-drugs and in setting cut-off values for the disease categories", Estonian Academy of Sciences students' research competition.

2011 Diploma from governmental competition of students research works to Bachelor's thesis "Applying molecular descriptors in the analysis of structural space of drugs and non-drugs and in setting cut-off values for the disease categories".

Scientific publications

Oja, M.; Sild, S.; Maran, U. Logistic classification models for pH-permeability profile: predicting permeability classes for the biopharmaceutical classification system. *J. Chem. Inf. Model.* **2019**, *59*, 2442–2455.

Oja, M.; Maran, U. pH-permeability profiles for drug substances: experimental detection, comparison with human intestinal absorption and modelling. *Eur. J. Pharm. Sci.* **2018**, *123*, 429–440.

Oja, M.; Maran, U. Quantitative structure–permeability relationships at various pH values for neutral and amphoteric drugs and drug-like compounds. *SAR QSAR Environ. Res.* **2016**, *27*, 813–832.

Oja, M.; Maran, U. Quantitative structure–permeability relationships at various pH values for acidic and basic drugs and drug-like compounds. *SAR QSAR Environ. Res.* **2015**, *26*, 701–719.

Oja, M.; Maran, U. The permeability of an artificial membrane for wide range of pH in human gastrointestinal tract: experimental measurements and quantitative structure-activity relationship. *Mol. Inf.* **2015**, *34*, 493–506.

García-Sosa, A.T.; **Oja, M.**; Hetényi, C.; Maran, U. Disease-specific differentiation between drugs and non-drugs using principal component analysis of their molecular descriptor space. *Mol. Inf.* **2012**, *31*, 369–383.

García-Sosa, A.T.; **Oja, M.**; Hetényi, C.; Maran, U. DrugLogit: logistic discrimination between drugs and non-drugs including disease-specificity by assigning probabilities based on molecular properties. *J. Chem. Inf. Model.* **2012**, *52*, 2165–2180.

Oral presentations at international conferences

Oja, M.; Sild, S.; Maran, U. Logistic classification models for pH-permeability profile to improve prediction of absorption in gastrointestinal tract. UNGAP Spring Meeting, February 12–13, 2019, Sofia, Bulgaria.

Oja, M.; Sild, S.; Maran, U. Classification models for pH-permeability profile to describe regional based absorption in gastrointestinal tract. 18th International Conference on (Q)SAR in Environmental and Health Sciences (QSAR2018), June 11–15, 2018, Bled, Slovenia.

Oja, M.; Maran, U. Measuring, modelling and predicting pH-profiles of membrane permeability. IAPC-6: 6th World Conference on Physico Chemical Methods in Drug Discovery and Development (PCMDDD-6) & 3rd World Conference on ADMET and DMPK (ADMET-3), September 4–6, 2017, Zagreb, Croatia.

Oja, M.; Maran, U. Measuring, modelling and predicting pH-profiles of membrane permeability. 6th FIP Pharmaceutical Sciences World Congress 2017: Young Scientist Satellite Conference, May 19–21, 2017, Uppsala, Sweden.

Oja, M.; Maran, U. Comparison and modelling of experimental and calculated distribution coefficient (logD) values at various pHs. 5th IAPC Meeting: Emerging Technologies in Drug Discovery and Development, August 23–26, 2016, Zhuhai, China.

Oja, M., Maran, U. Quantitative relationship between skin sensitization and skin extract membrane penetration. 17th International Conference on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2016), June 13–17, 2016, Miami Beach, USA.

ELULOOKIRJELDUS

Nimi Mare Oja
Sünniaeg 6. juuli 1989, Tartu, Eesti
Kodakondsus Eesti
Aadress Keemia instituut, Tartu Ülikool, Ravila 14A, 50411, Tartu, Eesti
E-mail mare.oja@ut.ee

Haridus

2013–… Tartu Ülikool, doktoriõpe, molekulaartehnoloogia
2011–2013 Tartu Ülikool, loodusteaduste magister (MSc), keemia
2008–2011 Tartu Ülikool, loodusteaduste bakalaureuse kraad (BSc), keemia
2005–2008 Hugo Treffneri Gümnaasium

Teenistuskäik

09.2017–… Tartu Ülikool, Keemia instituut, nooremteadur (1,0)
01.–08.2017 Tartu Ülikool, Keemia instituut, keemik (0,50)
2013–2015 Tartu Ülikool, Keemia instituut, keemik (0,50)

Teadusorganisatsiooniline ja -administratiivne tegevus

2013−... liige, *International Association of Physical Chemists (IAPC)*

Erialane enesetäiendus

2017 *3rd International Summer School on Drug Development (ISSDD-3)*, 1.–3. september 2017, Zagreb, Horvaatia.
2016 *Short Introduction in Drug Development*, 26. august 2016, Zhuhai, Hiina.
2015 *Targeted chemotherapy towards diseases caused by endoparasites*, Fraunhofer-IME ScreeningPort (COST Action CM1307), 28. september – 1. oktoober 2015, Hamburg, Saksamaa.
2013 *1st European Summer School on Formulation Development (ESSFD-1)*, 20.–22. september 2013, Dubrovnik, Horvaatia.

Tunnustused

2016 Parim noorteadlase suuline ettekanne, *17th International Conference on Quantitative Structure-Activity Relationships in Environmental and Health Sciences* (QSAR2016), Miami Beach, USA.
2015 Parim noorteadlase stendiettekanne, *8th International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources* (CMTPI2015), Chios, Kreeka.
2011 Eesti Teaduste Akadeemia üliõpilastööde konkursil auhind bakalaureusetöö eest: "Molekulaardeskriptorite kasutamine ravimite ja mitteravimite struktuuriruumi kirjeldamiseks ja haiguspõhiste ravimikategooriate piiritlemiseks".

2011 Üliõpilaste teadustööde riiklikul konkursil diplom bakalaureusetöö eest: "Molekulaardeskriptorite kasutamine ravimite ja mitteravimite struktuuriruumi kirjeldamiseks ja haiguspõhiste ravimikategooriate piiritlemiseks".

Teaduspublikatsioonid

Oja, M.; Sild, S.; Maran, U. Logistic classification models for pH-permeability profile: predicting permeability classes for the biopharmaceutical classification system. *J. Chem. Inf. Model.* **2019**, *59*, 2442–2455.

Oja, M.; Maran, U. pH-permeability profiles for drug substances: experimental detection, comparison with human intestinal absorption and modelling. *Eur. J. Pharm. Sci.* **2018**, *123*, 429–440.

Oja, M.; Maran, U. Quantitative structure–permeability relationships at various pH values for neutral and amphoteric drugs and drug-like compounds. *SAR QSAR Environ. Res.* **2016**, *27*, 813–832.

Oja, M.; Maran, U. Quantitative structure–permeability relationships at various pH values for acidic and basic drugs and drug-like compounds. *SAR QSAR Environ. Res.* **2015**, *26*, 701–719.

Oja, M.; Maran, U. The permeability of an artificial membrane for wide range of pH in human gastrointestinal tract: experimental measurements and quantitative structure-activity relationship. *Mol. Inf.* **2015**, *34*, 493–506.

García-Sosa, A.T.; **Oja, M.**; Hetényi, C.; Maran, U. Disease-specific differentiation between drugs and non-drugs using principal component analysis of their molecular descriptor space. *Mol. Inf.* **2012**, *31*, 369–383.

García-Sosa, A.T.; **Oja, M.**; Hetényi, C.; Maran, U. DrugLogit: logistic discrimination between drugs and non-drugs including disease-specificity by assigning probabilities based on molecular properties. *J. Chem. Inf. Model.* **2012**, *52*, 2165–2180.

Suulised ettekanded rahvusvahelistel konverentsidel

Oja, M.; Sild, S.; Maran, U. Logistic classification models for pH-permeability profile to improve prediction of absorption in gastrointestinal tract. *UNGAP Spring Meeting*, 12.–13. veebruar, 2019, Sofia, Bulgaaria.

Oja, M.; Sild, S.; Maran, U. Classification models for pH-permeability profile to describe regional based absorption in gastrointestinal tract. *18th International Conference on (Q)SAR in Environmental and Health Sciences (QSAR2018)*, 11.–15. juuni, 2018, Bled, Sloveenia.

Oja, M.; Maran, U. Measuring, modelling and predicting pH-profiles of membrane permeability. *IAPC-6: 6th World Conference on Physico Chemical Methods in Drug Discovery and Development (PCMDDD-6) & 3rd World Conference on ADMET and DMPK (ADMET-3)*, 4.–6. september 2017, Zagreb, Horvaatia.

Oja, M.; Maran, U. Measuring, modelling and predicting pH-profiles of membrane permeability. *6th FIP Pharmaceutical Sciences World Congress 2017: Young Scientist Satellite Conference*, 19.–21. mai 2017, Uppsala, Rootsi.

Oja, M.; Maran, U. Comparison and modelling of experimental and calculated distribution coefficient (logD) values at various pHs. *5th IAPC Meeting: Emerging Technologies in Drug Discovery and Development*, 23.–26. august 2016, Zhuhai, Hiina.

Oja M., Maran U. Quantitative relationship between skin sensitization and skin extract membrane penetration. *17th International Conference on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2016)*, 13.–17. juuni 2016, Miami Beach, USA.

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS

1. **Toomas Tamm.** Quantum-chemical simulation of solvent effects. Tartu, 1993, 110 p.
2. **Peeter Burk.** Theoretical study of gas-phase acid-base equilibria. Tartu, 1994, 96 p.
3. **Victor Lobanov.** Quantitative structure-property relationships in large descriptor spaces. Tartu, 1995, 135 p.
4. **Vahur Mäemets.** The ^{17}O and ^{1}H nuclear magnetic resonance study of H_2O in individual solvents and its charged clusters in aqueous solutions of electrolytes. Tartu, 1997, 140 p.
5. **Andrus Metsala.** Microcanonical rate constant in nonequilibrium distribution of vibrational energy and in restricted intramolecular vibrational energy redistribution on the basis of slater's theory of unimolecular reactions. Tartu, 1997, 150 p.
6. **Uko Maran.** Quantum-mechanical study of potential energy surfaces in different environments. Tartu, 1997, 137 p.
7. **Alar Jänes.** Adsorption of organic compounds on antimony, bismuth and cadmium electrodes. Tartu, 1998, 219 p.
8. **Kaido Tammeveski.** Oxygen electroreduction on thin platinum films and the electrochemical detection of superoxide anion. Tartu, 1998, 139 p.
9. **Ivo Leito.** Studies of Brønsted acid-base equilibria in water and non-aqueous media. Tartu, 1998, 101 p.
10. **Jaan Leis.** Conformational dynamics and equilibria in amides. Tartu, 1998, 131 p.
11. **Toonika Rinken.** The modelling of amperometric biosensors based on oxidoreductases. Tartu, 2000, 108 p.
12. **Dmitri Panov.** Partially solvated Grignard reagents. Tartu, 2000, 64 p.
13. **Kaja Orupõld.** Treatment and analysis of phenolic wastewater with microorganisms. Tartu, 2000, 123 p.
14. **Jüri Ivask.** Ion Chromatographic determination of major anions and cations in polar ice core. Tartu, 2000, 85 p.
15. **Lauri Vares.** Stereoselective Synthesis of Tetrahydrofuran and Tetrahydropyran Derivatives by Use of Asymmetric Horner-Wadsworth-Emmons and Ring Closure Reactions. Tartu, 2000, 184 p.
16. **Martin Lepiku.** Kinetic aspects of dopamine D_2 receptor interactions with specific ligands. Tartu, 2000, 81 p.
17. **Katrin Sak.** Some aspects of ligand specificity of P2Y receptors. Tartu, 2000, 106 p.
18. **Vello Pällin.** The role of solvation in the formation of iotsitch complexes. Tartu, 2001, 95 p.
19. **Katrin Kollist.** Interactions between polycyclic aromatic compounds and humic substances. Tartu, 2001, 93 p.

20. **Ivar Koppel.** Quantum chemical study of acidity of strong and superstrong Brønsted acids. Tartu, 2001, 104 p.
21. **Viljar Pihl.** The study of the substituent and solvent effects on the acidity of OH and CH acids. Tartu, 2001, 132 p.
22. **Natalia Palm.** Specification of the minimum, sufficient and significant set of descriptors for general description of solvent effects. Tartu, 2001, 134 p.
23. **Sulev Sild.** QSPR/QSAR approaches for complex molecular systems. Tartu, 2001, 134 p.
24. **Ruslan Petrukhin.** Industrial applications of the quantitative structure-property relationships. Tartu, 2001, 162 p.
25. **Boris V. Rogovoy.** Synthesis of (benzotriazolyl)carboximidamides and their application in relations with *N*- and *S*-nucleophyles. Tartu, 2002, 84 p.
26. **Koit Herodes.** Solvent effects on UV-vis absorption spectra of some solvatochromic substances in binary solvent mixtures: the preferential solvation model. Tartu, 2002, 102 p.
27. **Anti Perkson.** Synthesis and characterisation of nanostructured carbon. Tartu, 2002, 152 p.
28. **Ivari Kaljurand.** Self-consistent acidity scales of neutral and cationic Brønsted acids in acetonitrile and tetrahydrofuran. Tartu, 2003, 108 p.
29. **Karmen Lust.** Adsorption of anions on bismuth single crystal electrodes. Tartu, 2003, 128 p.
30. **Mare Piirsalu.** Substituent, temperature and solvent effects on the alkaline hydrolysis of substituted phenyl and alkyl esters of benzoic acid. Tartu, 2003, 156 p.
31. **Meeri Sassian.** Reactions of partially solvated Grignard reagents. Tartu, 2003, 78 p.
32. **Tarmo Tamm.** Quantum chemical modelling of polypyrrole. Tartu, 2003. 100 p.
33. **Erik Teinemaa.** The environmental fate of the particulate matter and organic pollutants from an oil shale power plant. Tartu, 2003. 102 p.
34. **Jaana Tammiku-Taul.** Quantum chemical study of the properties of Grignard reagents. Tartu, 2003. 120 p.
35. **Andre Lomaka.** Biomedical applications of predictive computational chemistry. Tartu, 2003. 132 p.
36. **Kostyantyn Kirichenko.** Benzotriazole – Mediated Carbon–Carbon Bond Formation. Tartu, 2003. 132 p.
37. **Gunnar Nurk.** Adsorption kinetics of some organic compounds on bismuth single crystal electrodes. Tartu, 2003, 170 p.
38. **Mati Arulepp.** Electrochemical characteristics of porous carbon materials and electrical double layer capacitors. Tartu, 2003, 196 p.
39. **Dan Cornel Fara.** QSPR modeling of complexation and distribution of organic compounds. Tartu, 2004, 126 p.
40. **Riina Mahlapuu.** Signalling of galanin and amyloid precursor protein through adenylate cyclase. Tartu, 2004, 124 p.

41. **Mihkel Kerikmäe.** Some luminescent materials for dosimetric applications and physical research. Tartu, 2004, 143 p.
42. **Jaanus Kruusma.** Determination of some important trace metal ions in human blood. Tartu, 2004, 115 p.
43. **Urmas Johanson.** Investigations of the electrochemical properties of polypyrrole modified electrodes. Tartu, 2004, 91 p.
44. **Kaido Sillar.** Computational study of the acid sites in zeolite ZSM-5. Tartu, 2004, 80 p.
45. **Aldo Oras.** Kinetic aspects of dATPαS interaction with $P2Y_1$ receptor. Tartu, 2004, 75 p.
46. **Erik Mölder.** Measurement of the oxygen mass transfer through the air-water interface. Tartu, 2005, 73 p.
47. **Thomas Thomberg.** The kinetics of electroreduction of peroxodisulfate anion on cadmium (0001) single crystal electrode. Tartu, 2005, 95 p.
48. **Olavi Loog.** Aspects of condensations of carbonyl compounds and their imine analogues. Tartu, 2005, 83 p.
49. **Siim Salmar.** Effect of ultrasound on ester hydrolysis in aqueous ethanol. Tartu, 2006, 73 p.
50. **Ain Uustare.** Modulation of signal transduction of heptahelical receptors by other receptors and G proteins. Tartu, 2006, 121 p.
51. **Sergei Yurchenko.** Determination of some carcinogenic contaminants in food. Tartu, 2006, 143 p.
52. **Kaido Tämm.** QSPR modeling of some properties of organic compounds. Tartu, 2006, 67 p.
53. **Olga Tšubrik.** New methods in the synthesis of multisubstituted hydrazines. Tartu. 2006, 183 p.
54. **Lilli Sooväli.** Spectrophotometric measurements and their uncertainty in chemical analysis and dissociation constant measurements. Tartu, 2006, 125 p.
55. **Eve Koort.** Uncertainty estimation of potentiometrically measured ph and pK_a values. Tartu, 2006, 139 p.
56. **Sergei Kopanchuk.** Regulation of ligand binding to melanocortin receptor subtypes. Tartu, 2006, 119 p.
57. **Silvar Kallip.** Surface structure of some bismuth and antimony single crystal electrodes. Tartu, 2006, 107 p.
58. **Kristjan Saal.** Surface silanization and its application in biomolecule coupling. Tartu, 2006, 77 p.
59. **Tanel Tätte.** High viscosity $Sn(OBu)_4$ oligomeric concentrates and their applications in technology. Tartu, 2006, 91 p.
60. **Dimitar Atanasov Dobchev**. Robust QSAR methods for the prediction of properties from molecular structure. Tartu, 2006, 118 p.
61. **Hannes Hagu**. Impact of ultrasound on hydrophobic interactions in solutions. Tartu, 2007, 81 p.
62. **Rutha Jäger.** Electroreduction of peroxodisulfate anion on bismuth electrodes. Tartu, 2007, 142 p.

63. **Kaido Viht.** Immobilizable bisubstrate-analogue inhibitors of basophilic protein kinases: development and application in biosensors. Tartu, 2007, 88 p.
64. **Eva-Ingrid Rõõm.** Acid-base equilibria in nonpolar media. Tartu, 2007, 156 p.
65. **Sven Tamp.** DFT study of the cesium cation containing complexes relevant to the cesium cation binding by the humic acids. Tartu, 2007, 102 p.
66. **Jaak Nerut.** Electroreduction of hexacyanoferrate(III) anion on Cadmium (0001) single crystal electrode. Tartu, 2007, 180 p.
67. **Lauri Jalukse.** Measurement uncertainty estimation in amperometric dissolved oxygen concentration measurement. Tartu, 2007, 112 p.
68. **Aime Lust.** Charge state of dopants and ordered clusters formation in CaF_2:Mn and CaF_2:Eu luminophors. Tartu, 2007, 100 p.
69. **Iiris Kahn**. Quantitative Structure-Activity Relationships of environmentally relevant properties. Tartu, 2007, 98 p.
70. **Mari Reinik.** Nitrates, nitrites, N-nitrosamines and polycyclic aromatic hydrocarbons in food: analytical methods, occurrence and dietary intake. Tartu, 2007, 172 p.
71. **Heili Kasuk.** Thermodynamic parameters and adsorption kinetics of organic compounds forming the compact adsorption layer at Bi single crystal electrodes. Tartu, 2007, 212 p.
72. **Erki Enkvist.** Synthesis of adenosine-peptide conjugates for biological applications. Tartu, 2007, 114 p.
73. **Svetoslav Hristov Slavov**. Biomedical applications of the QSAR approach. Tartu, 2007, 146 p.
74. **Eneli Härk.** Electroreduction of complex cations on electrochemically polished Bi(*hkl*) single crystal electrodes. Tartu, 2008, 158 p.
75. **Priit Möller.** Electrochemical characteristics of some cathodes for medium temperature solid oxide fuel cells, synthesized by solid state reaction technique. Tartu, 2008, 90 p.
76. **Signe Viggor.** Impact of biochemical parameters of genetically different pseudomonads at the degradation of phenolic compounds. Tartu, 2008, 122 p.
77. **Ave Sarapuu.** Electrochemical reduction of oxygen on quinone-modified carbon electrodes and on thin films of platinum and gold. Tartu, 2008, 134 p.
78. **Agnes Kütt.** Studies of acid-base equilibria in non-aqueous media. Tartu, 2008, 198 p.
79. **Rouvim Kadis.** Evaluation of measurement uncertainty in analytical chemistry: related concepts and some points of misinterpretation. Tartu, 2008, 118 p.
80. **Valter Reedo.** Elaboration of IVB group metal oxide structures and their possible applications. Tartu, 2008, 98 p.
81. **Aleksei Kuznetsov.** Allosteric effects in reactions catalyzed by the cAMP-dependent protein kinase catalytic subunit. Tartu, 2009, 133 p.

82. **Aleksei Bredihhin.** Use of mono- and polyanions in the synthesis of multisubstituted hydrazine derivatives. Tartu, 2009, 105 p.
83. **Anu Ploom.** Quantitative structure-reactivity analysis in organosilicon chemistry. Tartu, 2009, 99 p.
84. **Argo Vonk.** Determination of adenosine A_{2A}- and dopamine D_1 receptor-specific modulation of adenylate cyclase activity in rat striatum. Tartu, 2009, 129 p.
85. **Indrek Kivi.** Synthesis and electrochemical characterization of porous cathode materials for intermediate temperature solid oxide fuel cells. Tartu, 2009, 177 p.
86. **Jaanus Eskusson.** Synthesis and characterisation of diamond-like carbon thin films prepared by pulsed laser deposition method. Tartu, 2009, 117 p.
87. **Marko Lätt.** Carbide derived microporous carbon and electrical double layer capacitors. Tartu, 2009, 107 p.
88. **Vladimir Stepanov.** Slow conformational changes in dopamine transporter interaction with its ligands. Tartu, 2009, 103 p.
89. **Aleksander Trummal.** Computational Study of Structural and Solvent Effects on Acidities of Some Brønsted Acids. Tartu, 2009, 103 p.
90. **Eerold Vellemäe.** Applications of mischmetal in organic synthesis. Tartu, 2009, 93 p.
91. **Sven Parkel.** Ligand binding to 5-HT_{1A} receptors and its regulation by Mg^{2+} and Mn^{2+}. Tartu, 2010, 99 p.
92. **Signe Vahur.** Expanding the possibilities of ATR-FT-IR spectroscopy in determination of inorganic pigments. Tartu, 2010, 184 p.
93. **Tavo Romann**. Preparation and surface modification of bismuth thin film, porous, and microelectrodes. Tartu, 2010, 155 p.
94. **Nadežda Aleksejeva.** Electrocatalytic reduction of oxygen on carbon nanotube-based nanocomposite materials. Tartu, 2010, 147 p.
95. **Marko Kullapere.** Electrochemical properties of glassy carbon, nickel and gold electrodes modified with aryl groups. Tartu, 2010, 233 p.
96. **Liis Siinor.** Adsorption kinetics of ions at Bi single crystal planes from aqueous electrolyte solutions and room-temperature ionic liquids. Tartu, 2010, 101 p.
97. **Angela Vaasa.** Development of fluorescence-based kinetic and binding assays for characterization of protein kinases and their inhibitors. Tartu 2010, 101 p.
98. **Indrek Tulp.** Multivariate analysis of chemical and biological properties. Tartu 2010, 105 p.
99. **Aare Selberg.** Evaluation of environmental quality in Northern Estonia by the analysis of leachate. Tartu 2010, 117 p.
100. **Darja Lavõgina.** Development of protein kinase inhibitors based on adenosine analogue-oligoarginine conjugates. Tartu 2010, 248 p.
101. **Laura Herm.** Biochemistry of dopamine D_2 receptors and its association with motivated behaviour. Tartu 2010, 156 p.

102. **Terje Raudsepp.** Influence of dopant anions on the electrochemical properties of polypyrrole films. Tartu 2010, 112 p.
103. **Margus Marandi.** Electroformation of Polypyrrole Films: *In-situ* AFM and STM Study. Tartu 2011, 116 p.
104. **Kairi Kivirand.** Diamine oxidase-based biosensors: construction and working principles. Tartu, 2011, 140 p.
105. **Anneli Kruve.** Matrix effects in liquid-chromatography electrospray mass-spectrometry. Tartu, 2011, 156 p.
106. **Gary Urb.** Assessment of environmental impact of oil shale fly ash from PF and CFB combustion. Tartu, 2011, 108 p.
107. **Nikita Oskolkov.** A novel strategy for peptide-mediated cellular delivery and induction of endosomal escape. Tartu, 2011, 106 p.
108. **Dana Martin.** The QSPR/QSAR approach for the prediction of properties of fullerene derivatives. Tartu, 2011, 98 p.
109. **Säde Viirlaid.** Novel glutathione analogues and their antioxidant activity. Tartu, 2011, 106 p.
110. **Ülis Sõukand.** Simultaneous adsorption of Cd^{2+}, Ni^{2+}, and Pb^{2+} on peat. Tartu, 2011, 124 p.
111. **Lauri Lipping.** The acidity of strong and superstrong Brønsted acids, an outreach for the "limits of growth": a quantum chemical study. Tartu, 2011, 124 p.
112. **Heisi Kurig.** Electrical double-layer capacitors based on ionic liquids as electrolytes. Tartu, 2011, 146 p.
113. **Marje Kasari.** Bisubstrate luminescent probes, optical sensors and affinity adsorbents for measurement of active protein kinases in biological samples. Tartu, 2012, 126 p.
114. **Kalev Takkis.** Virtual screening of chemical databases for bioactive molecules. Tartu, 2012, 122 p.
115. **Ksenija Kisseljova**. Synthesis of aza-β^3-amino acid containing peptides and kinetic study of their phosphorylation by protein kinase A. Tartu, 2012, 104 p.
116. **Riin Rebane.** Advanced method development strategy for derivatization LC/ESI/MS. Tartu, 2012, 184 p.
117. **Vladislav Ivaništšev.** Double layer structure and adsorption kinetics of ions at metal electrodes in room temperature ionic liquids. Tartu, 2012, 128 p.
118. **Irja Helm.** High accuracy gravimetric Winkler method for determination of dissolved oxygen. Tartu, 2012, 139 p.
119. **Karin Kipper.** Fluoroalcohols as Components of LC-ESI-MS Eluents: Usage and Applications. Tartu, 2012, 164 p.
120. **Arno Ratas.** Energy storage and transfer in dosimetric luminescent materials. Tartu, 2012, 163 p.
121. **Reet Reinart-Okugbeni**. Assay systems for characterisation of subtype-selective binding and functional activity of ligands on dopamine receptors. Tartu, 2012, 159 p.

122. **Lauri Sikk.** Computational study of the Sonogashira cross-coupling reaction. Tartu, 2012, 81 p.
123. **Karita Raudkivi.** Neurochemical studies on inter-individual differences in affect-related behaviour of the laboratory rat. Tartu, 2012, 161 p.
124. **Indrek Saar.** Design of GalR2 subtype specific ligands: their role in depression-like behavior and feeding regulation. Tartu, 2013, 126 p.
125. **Ann Laheäär.** Electrochemical characterization of alkali metal salt based non-aqueous electrolytes for supercapacitors. Tartu, 2013, 127 p.
126. **Kerli Tõnurist.** Influence of electrospun separator materials properties on electrochemical performance of electrical double-layer capacitors. Tartu, 2013, 147 p.
127. **Kaija Põhako-Esko.** Novel organic and inorganic ionogels: preparation and characterization. Tartu, 2013, 124 p.
128. **Ivar Kruusenberg.** Electroreduction of oxygen on carbon nanomaterial-based catalysts. Tartu, 2013, 191 p.
129. **Sander Piiskop.** Kinetic effects of ultrasound in aqueous acetonitrile solutions. Tartu, 2013, 95 p.
130. **Ilona Faustova**. Regulatory role of L-type pyruvate kinase N-terminal domain. Tartu, 2013, 109 p.
131. **Kadi Tamm.** Synthesis and characterization of the micro-mesoporous anode materials and testing of the medium temperature solid oxide fuel cell single cells. Tartu, 2013, 138 p.
132. **Iva Bozhidarova Stoyanova-Slavova.** Validation of QSAR/QSPR for regulatory purposes. Tartu, 2013, 109 p.
133. **Vitali Grozovski**. Adsorption of organic molecules at single crystal electrodes studied by *in situ* STM method. Tartu, 2014, 146 p.
134. **Santa Veikšina**. Development of assay systems for characterisation of ligand binding properties to melanocortin 4 receptors. Tartu, 2014, 151 p.
135. **Jüri Liiv.** PVDF (polyvinylidene difluoride) as material for active element of twisting-ball displays. Tartu, 2014, 111 p.
136. **Kersti Vaarmets.** Electrochemical and physical characterization of pristine and activated molybdenum carbide-derived carbon electrodes for the oxygen electroreduction reaction. Tartu, 2014, 131 p.
137. **Lauri Tõntson**. Regulation of G-protein subtypes by receptors, guanine nucleotides and Mn^{2+}. Tartu, 2014, 105 p.
138. **Aiko Adamson.** Properties of amine-boranes and phosphorus analogues in the gas phase. Tartu, 2014, 78 p.
139. **Elo Kibena**. Electrochemical grafting of glassy carbon, gold, highly oriented pyrolytic graphite and chemical vapour deposition-grown graphene electrodes by diazonium reduction method. Tartu, 2014, 184 p.
140. **Teemu Näykki.** Novel Tools for Water Quality Monitoring – From Field to Laboratory. Tartu, 2014, 202 p.
141. **Karl Kaupmees.** Acidity and basicity in non-aqueous media: importance of solvent properties and purity. Tartu, 2014, 128 p.

142. **Oleg Lebedev**. Hydrazine polyanions: different strategies in the synthesis of heterocycles. Tartu, 2015, 118 p.
143. **Geven Piir.** Environmental risk assessment of chemicals using QSAR methods. Tartu, 2015, 123 p.
144. **Olga Mazina.** Development and application of the biosensor assay for measurements of cyclic adenosine monophosphate in studies of G protein-coupled receptor signalinga. Tartu, 2015, 116 p.
145. **Sandip Ashokrao Kadam.** Anion receptors: synthesis and accurate binding measurements. Tartu, 2015, 116 p.
146. **Indrek Tallo.** Synthesis and characterization of new micro-mesoporous carbide derived carbon materials for high energy and power density electrical double layer capacitors. Tartu, 2015, 148 p.
147. **Heiki Erikson.** Electrochemical reduction of oxygen on nanostructured palladium and gold catalysts. Tartu, 2015, 204 p.
148. **Erik Anderson.** *In situ* Scanning Tunnelling Microscopy studies of the interfacial structure between Bi(111) electrode and a room temperature ionic liquid. Tartu, 2015, 118 p.
149. **Girinath G. Pillai.** Computational Modelling of Diverse Chemical, Biochemical and Biomedical Properties. Tartu, 2015, 140 p.
150. **Piret Pikma.** Interfacial structure and adsorption of organic compounds at Cd(0001) and Sb(111) electrodes from ionic liquid and aqueous electrolytes: an *in situ* STM study. Tartu, 2015, 126 p.
151. **Ganesh babu Manoharan.** Combining chemical and genetic approaches for photoluminescence assays of protein kinases. Tartu, 2016, 126 p.
152. **Carolin Siimenson**. Electrochemical characterization of halide ion adsorption from liquid mixtures at Bi(111) and pyrolytic graphite electrode surface. Tartu, 2016, 110 p.
153. **Asko Laaniste.** Comparison and optimisation of novel mass spectrometry ionisation sources. Tartu, 2016, 156 p.
154. **Hanno Evard.** Estimating limit of detection for mass spectrometric analysis methods. Tartu, 2016, 224 p.
155. **Kadri Ligi.** Characterization and application of protein kinase-responsive organic probes with triplet-singlet energy transfer. Tartu, 2016, 122 p.
156. **Margarita Kagan.** Biosensing penicillins' residues in milk flows. Tartu, 2016, 130 p.
157. **Marie Kriisa.** Development of protein kinase-responsive photoluminescent probes and cellular regulators of protein phosphorylation. Tartu, 2016, 106 p.
158. **Mihkel Vestli.** Ultrasonic spray pyrolysis deposited electrolyte layers for intermediate temperature solid oxide fuel cells. Tartu, 2016, 156 p.
159. **Silver Sepp**. Influence of porosity of the carbide-derived carbon on the properties of the composite electrocatalysts and characteristics of polymer electrolyte fuel cells. Tartu, 2016, 137 p.
160. **Kristjan Haav**. Quantitative relative equilibrium constant measurements in supramolecular chemistry. Tartu, 2017, 158 p.

161. **Anu Teearu**. Development of MALDI-FT-ICR-MS methodology for the analysis of resinous materials. Tartu, 2017, 205 p.
162. **Taavi Ivan**. Bifunctional inhibitors and photoluminescent probes for studies on protein complexes. Tartu, 2017, 140 p.
163. **Maarja-Liisa Oldekop**. Characterization of amino acid derivatization reagents for LC-MS analysis. Tartu, 2017, 147 p.
164. **Kristel Jukk**. Electrochemical reduction of oxygen on platinum- and palladium-based nanocatalysts. Tartu, 2017, 250 p.
165. **Siim Kukk**. Kinetic aspects of interaction between dopamine transporter and *N*-substituted nortropane derivatives. Tartu, 2017, 107 p.
166. **Birgit Viira**. Design and modelling in early drug development in targeting HIV-1 reverse transcriptase and Malaria. Tartu, 2017, 172 p.
167. **Rait Kivi**. Allostery in cAMP dependent protein kinase catalytic subunit. Tartu, 2017, 115 p.
168. **Agnes Heering**. Experimental realization and applications of the unified acidity scale. Tartu, 2017, 123 p.
169. **Delia Juronen**. Biosensing system for the rapid multiplex detection of mastitis-causing pathogens in milk. Tartu, 2018, 85 p.
170. **Hedi Rahnel.** ARC-inhibitors: from reliable biochemical assays to regulators of physiology of cells. Tartu, 2018, 176 p.
171. **Anton Ruzanov.** Computational investigation of the electrical double layer at metal–aqueous solution and metal–ionic liquid interfaces. Tartu, 2018, 129 p.
172. **Katrin Kestav.** Crystal Structure-Guided Development of Bisubstrate-Analogue Inhibitors of Mitotic Protein Kinase Haspin. Tartu, 2018, 166 p.
173. **Mihkel Ilisson.** Synthesis of novel heterocyclic hydrazine derivatives and their conjugates. Tartu, 2018, 101 p.
174. **Anni Allikalt.** Development of assay systems for studying ligand binding to dopamine receptors. Tartu, 2018, 160 p.
175. **Ove Oll.** Electrical double layer structure and energy storage characteristics of ionic liquid based capacitors. Tartu, 2018, 187 p.
176. **Rasmus Palm.** Carbon materials for energy storage applications. Tartu, 2018, 114 p.
177. **Jörgen Metsik.** Preparation and stability of poly(3,4-ethylenedioxythiophene) thin films for transparent electrode applications. Tartu, 2018, 111 p.
178. **Sofja Tšepelevitš.** Experimental studies and modeling of solute-solvent interactions. Tartu, 2018, 109 p.
179. **Märt Lõkov.** Basicity of some nitrogen, phosphorus and carbon bases in acetonitrile. Tartu, 2018, 104 p.
180. **Anton Mastitski**. Preparation of α-aza-amino acid precursors and related compounds by novel methods of reductive one-pot alkylation and direct alkylation. Tartu, 2018, 155 p.
181. **Jürgen Vahter**. Development of bisubstrate inhibitors for protein kinase CK2. Tartu, 2019, 186 p.

182. **Piia Liigand.** Expanding and improving methodology and applications of ionization efficiency measurements. Tartu, 2019, 189 p.
183. **Sigrid Selberg.** Synthesis and properties of lipophilic phosphazene-based indicator molecules. Tartu, 2019, 74 p.
184. **Jaanus Liigand.** Standard substance free quantification for LC/ESI/MS analysis based on the predicted ionization efficiencies. Tartu, 2019, 254 p.
185. **Marek Mooste.** Surface and electrochemical characterisation of aryl film and nanocomposite material modified carbon and metal-based electrodes. Tartu, 2019, 304 p.