VRIJE UNIVERSITEIT BRUSSEL

COMPARISON OF DIFFERENT STATIONARY PHASES IN SUPERCRITICAL FLUID CHROMATOGRAPHY TO PREDICT THE SKIN PERMEABILITY OF PHARMACEUTICAL COMPOUNDS

ROSEMARIE VERELLEN
2020-2021

Faculty of Medicine and Pharmacy
Department Pharmaceutical and Pharmacological Sciences (FARM)

Toestemmingsformulier
openbaarmaking masterproef

Student
Naam : Rosemarie Verellen
Rolnummer : 0538896
Opleiding : Farmaceutische wetenschappen – master geneesmiddelenontwikkeling
Academiejaar : 2020/2021

Masterproef
Titel : Comparison of different stationary phases in supercritical fluid chromatography to predict the skin permeability of pharmaceutical compounds
Promotor : Yvan Vander Heyden, Debby Mangelings

De masterproef waarvoor de student een credit behaalt, en waaromtrent geen 'non disclosure agreement' (NDA of geheimhoudingsovereenkomst) werd opgesteld, kan kosteloos worden opgenomen in de catalogus van de Universiteitsbibliotheek mits expliciete toestemming van de student.
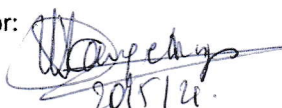
De student kiest in het kader van de mogelijkheid tot kosteloze terbeschikkingstelling van zijn/haar masterproef volgende optie:

- ☒ OPEN ACCESS: wereldwijde toegang tot de full tekst van de masterproef
- ☐ ENKEL VANOP DE CAMPUS: enkel toegang tot de full tekst van de masterproef vanop het VUB-netwerk
- ☐ EMBARGO WAARNA OPEN ACCESS VOLGT: pas wereldwijde toegang tot de full tekst van de masterproef na een opgegeven datum, met name …
- ☐ EMBARGO WAARNA ENKEL TOEGANG VANOP DE CAMPUS VOLGT: enkel vanop de campus toegang tot de full tekst van de masterproef na een opgegeven datum, met name …
- ☐ GEEN TOESTEMMING voor terbeschikkingstelling

De promotor bevestigt de kennisname van het voornemen van de student tot terbeschikkingstelling van de masterproef in de catalogus van de Universiteitsbibliotheek.
Datum:
Handtekening promotor:

Dit document wordt opgenomen in de masterproef. De student die het formulier niet toevoegt aan de masterproef en/of geen keuze heeft aangeduid en/of het formulier niet ondertekend heeft en/of geen kennisgeving aan de promotor heeft gedaan, wordt geacht geen toestemming tot openbaarmaking te verlenen. In dat geval zal de masterproef enkel worden gearchiveerd, maar is deze niet publiek toegankelijk.

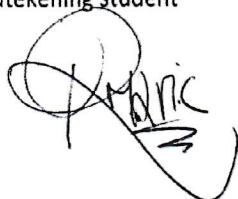Opgesteld te Gent op 20/5/2021.

Handtekening student

Table of contents

1. List of abbreviations

ANN	Artificial neural networks
BEH	Ethylene Bridged Hybrid
CART	Classification and regression trees
CMC	Critical micellar concentration
CTAB	Cetyltrimethylammonium bromide
e.g.	*exempli gratia*
GC	Gas chromatography
HILIC	Hydrophobic interaction liquid chromatography
HPLC	High performance liquid chromatography
IAM	Immobilized artificial membrane
i.e.	*Id est*
LFER	Linear Free Energy Relationship
MARS	Multivariate adaptive regression splines
MEKC	micellar electrokinetic chromatography
MLC	Micellar liquid chromatography
MLR	Multiple linear regression
OECD	Organisation for Economic Co-operation and Development
PAMPA	Parallel Artificial Membrane Permeability Assay
PDA	Photo diode array
PES	Polyethersulfone
PLS	Partial least squares
QSAR	Quantitative structure-activity relationship
RMSEC	Root mean squared error of calibration
RMSECV	Root mean squared error of cross validation
RPLC	Reversed-phase liquid chromatography
SDS	Sodium dodecyl sulfate
SFC	Supercritical fluid chromatography
STC	Sodium taurocholate
USA	United States of America
VIF	Variance inflation factor

2. Acknowledgements

I would like to take the opportunity to thank everyone who helped and supported me with the achievement of this work, my master thesis, directly or indirectly. First and foremost, a special thank you to my co-promotor Yasmine Grooten. I could not have accomplished all of this without her everlasting support, patience, the numerous proofreadings and answers to my infinite number of questions.

I would like to thank my two promotors as well, professor Yvan Vander Heyden and professor Debby Mangelings, for proofreading my final thesis.

Last but definitely not least, I would like to thank people close to me, such as my parents, brother, family, boyfriend and friends for their unconditional support, patience and for having faith in me during this stressful year.

3. Introduction:

3.1. Context

The transdermal administration of drugs has gained interest over the years due to certain advantages over oral and intravenous administration, such as avoiding the first-pass effect, sustained drug delivery and avoiding low pH (e.g. the stomach) which can degrade certain molecules. Good patient compliance is reported, but some patients experience irritated or allergic skin due to the patch or matrix itself of transdermal patches.

To understand the quality and performance of a transdermal drug delivery system, *ex-vivo* studies are required to estimate the permeation of a molecule (or a formulation, in case of ointments). Generally, human skin is used, often obtained from cadavers or plastic surgery and is obviously very scarce and not easily accessible (1). It goes without saying that a lot of (ethical) questions rise in the use of animal skin as an alternative to human skin. Regardless that animal skin in living animals is an alternative than excised human skin, the extrapolation to human skin is not always straightforward. In the last decades, research has been focused on finding new alternatives that do not use animals (and especially not harm them), such as the development of surrogate systems (for example the use of artificial membranes, chromatographic techniques, …)(2).

3.2. The skin

The skin is the largest organ of the human body and plays an important role in protecting the body from xenophobic agents to enter, on the one hand, and water to leave it, on the other. The skin consists of three main layers (figure 1): the most outer layer is called the epidermis, right underneath lies the dermis, to be by the hypodermis. The barrier function is mainly due to the most outer layer of the epidermis: the *stratum corneum* (10-30µm thick)(2). This layer consists of densely packed corneocytes which contain packed organized keratin filaments and filaggrin. The corneocytes are arranged in layers, varying from 10 up to 25 layers parallel to the skin surface. These cells possess a 'brick and mortar' structure, combined with a lipid matrix composed of ceramides, cholesterol and free fatty acids filling up the space between the corneocytes (1–3). This makes it very hard for foreign molecules to penetrate this layer of alternating hydrophobic and hydrophilic regions. The next layer (figure 1), the dermis, consists of connective tissue (or collagen) that is heavily vasculated with blood- and lymphatic vessels. It has a metabolic and protective function. Throughout the skin, multiple appendages are present, such as hair follicles, sebaceous- and sweat glands (1,3).
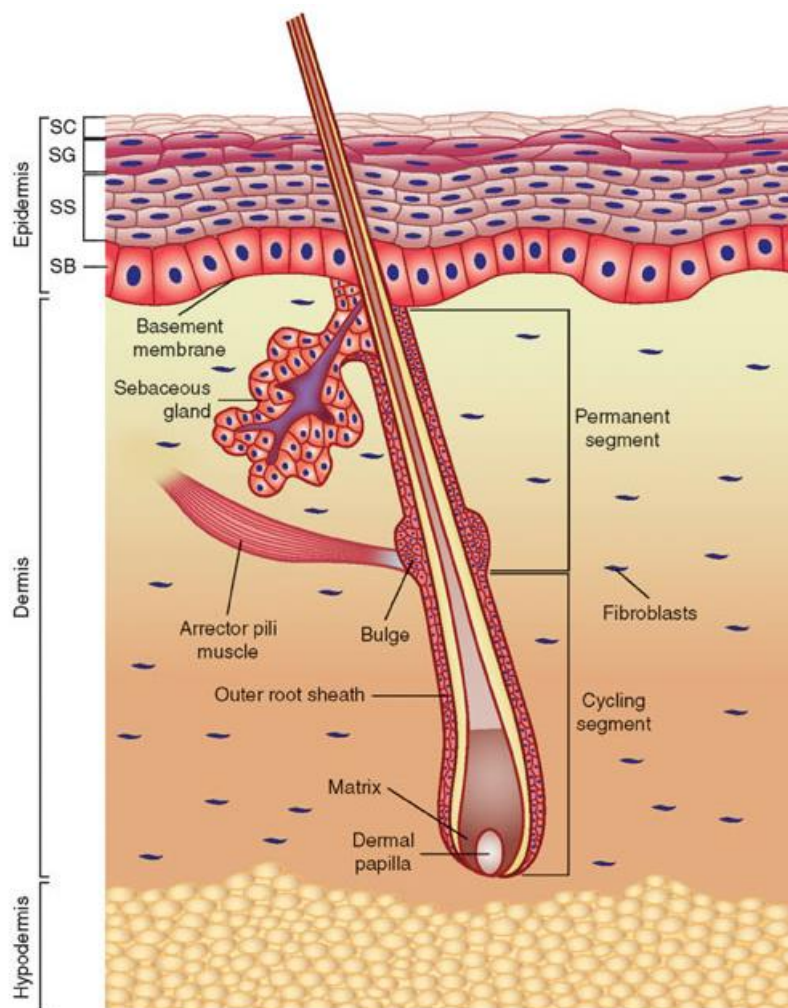
Figure 1: *The anatomy of the skin* (4)

Permeation of compounds through the skin to the vascular region of the dermis can follow three different pathways (figure 2): 1) the transcellular route (C), this pathway uses diffusion across the corneocytes. It is a very resistant pathway for drug permeation due to both the lipophilic (i.e. the lipid matrix) and hydrophilic regions (i.e. cytoplasm of the corneocytes) present, as discussed above. Another way for compounds to reach the dermis is 2) the intercellular route (A), where the mechanism is diffusion across the lipid matrix. The third pathway to reach the blood flow is 3) the appendageal route (B). In this pathway, the foreign substance can travel down the ducts of sweat and sebaceous glands, or down the hair follicle. Because these appendages appear in a small area of the total surface of the skin (approximately 0.1%), it is considered as a minor transdermal permeation route in comparison to the other two routes (1–3,5).
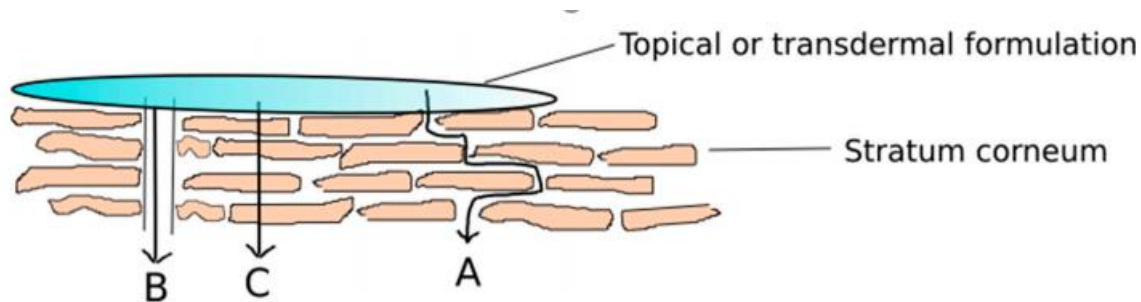
Figure 2: *The different drug permeation routes througt the skin. A: intercellular penetration pathway, B: follicular penetration pathway, C: transcellular penetration pathway (6).*

3.3. Parameters that influence the permeation through the skin

The passage of pharmaceutical substances through the skin depends on their physicochemical properties but also on the interaction with the membrane. Some of the main properties will be discussed below.

3.3.1. The partition coefficient

Log P, or P (the partition coefficient), is the ratio of the concentrations of a solute or chemical component in two immiscible solvents (a hydrophilic and lipophilic phase, usually water and 1-octanol, respectively). It can be determined, for example, by the 'shake-flask method'(7). This partition coefficient tells something about the lipophilicity of the substance and about the expected passive diffusion through the phospholipid bilayer of the biological membranes of human cells. The Lipinski rule of five indicates that molecules with a log P greater than 5 answered to poor oral absorption. An ideal value for oral absorption is a log P between 1.35 – 1.8 (8).

3.3.2. Molecular weight

Molecular weight is defined as the average mass of a molecule of a compound compared to $1/12$ of the mass of carbon 12 and calculated as the sum of the atomic weights of the constituent atoms. It is often expressed in the units g/mol or Dalton. According to the Lipinski rule of five for oral absorption, a molecular weight above 500 Dalton translates to poor absorption (8). Of course this is for oral absorption, not permeation through the skin, but the substance also has to pass membranes in the stomach/intestine just as through the skin. Molecular weight plays a fairly large role in skin permeability and the larger the molecule, the worse the absorption (9). Generally, molecules larger than 500 Dalton are considered impermeable through the skin. Arguments for this are: 1) all molecules that cause contact allergy are smaller than 500 Dalton, 2) all molecules used in topical formulations or transdermal drug delivery systems are smaller than 500 Dalton (9).

3.3.3. Melting point

The melting point of a substance is the temperature (or temperature-range) where the phase of the substance changes from solid to liquid. For skin permeability, a melting point below 200 ºC is necessary, which is related to an appropriate solubility (10).

3.4. Relation between skin permeability and steady state flux

As mentioned above, drug permeation through the skin happens mainly by passive diffusion through biological membranes. Adolf Fick developed two laws to describe and quantify this process: the First Law assumes steady state diffusion, while Fick's Second Law predicts the concentration gradient changes through diffusion, with time. In other words, the second law describes diffusion through the irregular intercellular regions of the *stratum corneum*. Only the steady-state diffusion will be discussed further. In this model, the *stratum corneum* lipid matrix is considered as a homogeneous medium surrounding corneocytes, through which the chemical product diffuses from high to low concentration according to the first law of Fick (11):

$$Q = D.a/h.t.\Delta Cs \qquad \text{(eq.1)}$$

where Q stands for the amount of substance that diffuses in a small period of time over a small surface, D is the diffusion constant, a describes the surface area of the membrane (skin) (in cm²), t is a time period (in s), h represents the distance (in mm) the substance will travel (in this case, the thickness of the *stratum corneum*) and ΔCs the concentration gradient (in millimoles) across the surface of the skin (11).

The above equation (eq. 1) can be rearranged as follows to define steady state flux (Jss) (11):

$$J_{ss} = Q/(a.t) = D.\Delta Cs/h \qquad \text{(eq. 2)}$$

The flux is defined as the amount of solute per unit area and per unit time, often mass/cm² per hour.

Because it is easier and more convenient to use concentrations in terms of the dissolved substance in the vehicle (Cv) and a partition coefficient K (which describes the partition between the vehicle and the skin, no units) instead of using Cs (solute concentration in the skin), eq. 2 can also be written as (11):

$$J_{ss} = (K.D.\Delta Cv)/h \quad \text{(where K = Cs/Cv)} \qquad \text{(eq. 3)}$$

The relation between K_p (the permeability coefficient, expressed in centimeter per second) and the steady state flux across the skin is described in equation 4. This equation indicates

that the skin permeability (K_p) is actually the steady state flux across the skin normalized by the concentration gradient, ΔCv (11):

$$K_p = J_{ss} / \Delta Cv \qquad\qquad\qquad (eq.\ 4)$$

The skin permeability coefficient (K_p) of a substance is a parameter that is determined by several techniques.

3.5. *In-vivo* skin permeation studies

In-vivo testing on human skin is still considered the golden standard for skin permeation testing, although it is not always possible because of ethical questions and other considerations. Most *in-vivo* studies are organized in (pre)clinical studies and are performed by taking several blood and/or urine samples from the participants at specific time intervals. This is, of course, very invasive, time-consuming and labor-intensive and not applicable for high-throughput screening. *In-vivo* skin testing is often limited to hypersensitivity and irritation tests (1).

Besides on human skin, these *in-vivo* experiments are also often executed on animals such as rats or dogs. The Organisation for Economic Co-operation and Development (OECD) describes a guideline (427) in which recommendations are defined depending on the used method. These recommendations cover, for example, the selection of the species, the number and sex of the species, the housing and feeding conditions, the application on the skin, the duration of exposure and the sampling. This is necessary to create standardized methods and to take animal rights into account (12).

Concerning animal testing, there is always a need for reduction, refinement and replacement (the 3R principle)(13). Luckily there are already some techniques that reduce the use of animals, or even replace them. Unfortunately, it is still very difficult to reproduce the human skin due to for instance the complexity of the different layers, the presence of living cells and the blood flow. Some of the alternatives to *in-vivo* animal testing that can be used to determine the skin permeability of a substance are listed in Section 1.6..

3.6. *In-vitro* skin permeation studies
3.6.1. Diffusion cells

There are several *in-vitro* methods. One type of testing uses the so-called **diffusion cells**. Two types can be distinguished: the static diffusion cell (*Franz cell*) and the flow-through diffusion cell (*Bronaugh cell*).
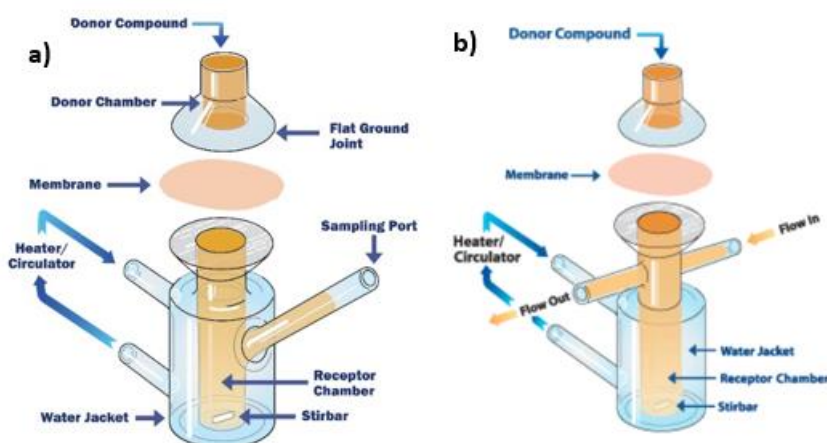
Figure 3: *a) static diffusion cell* (14) *b) flow-through diffusion cell* (15)

The static diffusion cell (figure 3a) consists of two chambers, the donor chamber and the receptor chamber, which are separated by a metal grid. Skin from humans or animals can be placed on the metal grid. Both chambers contain a solvent and the receptor chamber is placed in a water bath at 37°C to mimic real-life skin conditions. The receptor fluid is kept constant in temperature and homogenous in concentration with a stirring device. When a substance is dissolved and added to the donor chamber, several measurements are taken from the receptor chamber over time to test skin permeation (1,2,16,17).

The flow-through diffusion cell (figure 3b) is similar to the static diffusion cell, except that it has a continuous flow, the fluid in the receptor chamber is renewed to mimic the blood flow underneath real-life skin. This also provides continuous automatic sampling (16,17).

Diffusion cells are a simpler, faster and cheaper technique than *in-vivo* methods. A disadvantage of diffusion cells are the interlaboratory differences. Guidelines to minimize these differences are postulated by OECD guideline 428 (18). Obviously, only a small area of the skin is used, causing the physiological conditions (such as blood flow) to disappear. This is a drawback compared to *in-vivo* testing. The use of animal skin should also be taken into account, since this makes it harder to extrapolate to human skin.

3.6.2. Artificial membranes

An alternative for human skin is highly demanded. Although animal skin is much less expensive than human skin, it also gives a less good correlation with skin permeability in comparison to human skin and much more variation due to differences in race, age and sex of the animals which could potentially lead to differences in skin permeation. These considerations indicate the need for the development of synthetically produced membranes.

3.6.2.1.   Parallel Artificial Membranes Permeability Assay

One of the techniques using **artificial membranes** for *in-vitro* skin permeation measurements is the Parallel Artificial Membrane Permeability Assay (PAMPA) (figure 4).
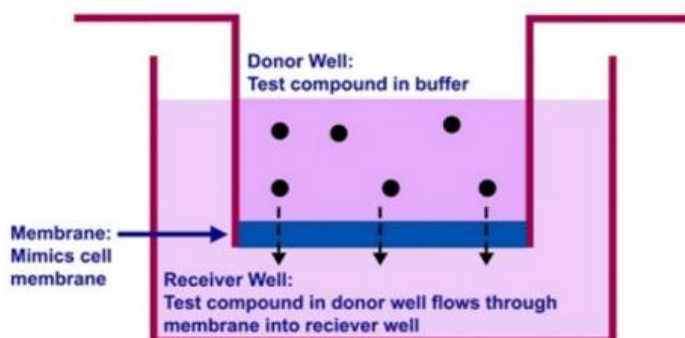
Figure 4: *Parallel Artificial Membrane Permeability Assay (PAMPA)* (19)

It consists of a 96-well plate system containing an artificial membrane which separates the donor and receptor compartments. The receptor compartment contains a buffer solution while the donor compartment contains the dissolved substance of which the permeation will be determined. By adding cetramides, cholesterol and stearic acid to the already existing membrane (which is made of phospholipid/dodecane), a solid alternative for the skin membrane is obtained (2). Because of PAMPA's short time period required for a high-throughput screening it is preferred above diffusion cells. It can be considered as a low-cost and time-effective alternative method for *in-vivo* testing. However certain structural elements of the real life human skin are not present (such as corneocytes and follicles) (2).

3.6.2.2.   Artificial membranes

Other *in-vitro* methods that use artificial membranes are 3D tissue models that contain living cells (meaning they can metabolize substances and are able to perform mitosis). These cells are created by culturing certain components of the human skin, such as keratinocytes and fibroblasts. Several of these *in-vitro* models are commercialized, for example EpiDerm™, EpiSkin™, and Labskin™. However, studies have shown that these membranes containing living cells are more permeable than the human skin itself (20). Therefore, these 3D models are mostly applied in the context of irritation and toxicity testing (1,2).

Artificial membranes based on silicone (such as Silatos™, Silastic®, Strat-M™ and polydimethylsiloxane) can also be used and give acceptable results according to different studies for both lipophilic and hydrophilic substances (21,22). These synthetic membranes are a cheap alternative in comparison to *in-vivo* studies. For Strat-M™, the membrane consists of multiple layers: the outer layer consists of two layers of polyethersulfone (PES,

more resistant to diffusion), while the bottom layer is a more diffusive polyolefin layer. Although this is a good attempt to imitate the human skin, the outer layer is not much more resistant in comparison to the bottom layer and does not function as a rate-determining barrier as does the *stratum corneum* in human skin (1,2).

3.7. *In-silico* approaches for predicting skin permeability

Another approach to predict the skin permeability is with *in-silico* methods. These models are made based on the molecular representation of the molecule. They try to describe the relationship between (theoretical) molecular descriptors and an experimental activity value (such as skin permeability). These molecular descriptors may include physiochemical, topological and geometrical properties of the molecule. Descriptors can be measured experimentally (for example, log P) or are theoretical and thus calculated *in-silico* (for example, theoretical descriptors). The relationship built is called a quantitative structure-activity relationship (QSAR) and allows to predict the activity of a new compound (without actually measuring it). Obviously this is a valuable technique because it can be used early in development, even before the molecule is synthesized (23).

When making a selection of the required set of molecular descriptors, there are two options. Thousands of molecular descriptors are described in the literature. To to make a selection of those ones are important, variable selection techniques can be applied, while (multiple) chemometric regression techniques can be used to build the QSAR model. Some examples of chemometric regression techniques are multiple linear regression (MLR) (24), principal component regression (PCR) (25) and partial least squares (PLS) regression (25). More complex non-linear modelling techniques can also be used, such as artificial neural networks (ANN) (26), classification and regression trees (CART) and multivariate adaptive regression splines (MARS). These techniques narrow thousands of potential descriptors down to just a few applied. Another approach to narrow down descriptors, is to apply a set of molecular descriptors as in popular QSARs, such as the Linear Free Energy Relationship (LFER) models. The Lipinski's rule of five (rule of thumb) is a qualitative approach. Lipinki's rule of five describes some values of descriptors that, if answered positively, relate to poor oral absorption. These rules include: a molecular weight over 500 Dalton, more than five hydrogen bond donors, more than ten hydrogen bond acceptors, and a log P above five. LFER describes relationships between some free-energy related property, such as solubility, absorption, partitioning between hydrophilic and lipophilic phases and adsorption to five molecular descriptors (23). The following equation is used:

$$SP = c + eE + sS + aA + bB + vV \qquad \qquad (eq. 5)$$

where SP describes the molecular activity, in our case log K_p. The capital letters describe five different solute (molecular) descriptors and the lower case letters describe the

corresponding coefficients of these molecular descriptors. E is the solute excess molar refractivity in $(cm^3/mol)/10$, S the solute dipolarity/polarizability, A the solute's hydrogen bond acidity, B the solute's hydrogen bond basicity and V the McGowan characteristic volume in $(cm^3/mol)/100$ (27).

3.8. Chromatography

Other methods to estimate the permeation of different active pharmaceutical and cosmetic substances involve chromatographic techniques. Skin permeability data can be predicted by combining the chromatographic retention with molecular descriptors (28).

3.8.1. Collagen columns

Collagen is a protein in the epidermis. Stationary phases that contain collagen show retention caused by the combination of specific (polar) and hydrophobic solute–stationary phase interactions. Although research (29) showed that the interaction of substances with collagen is less important in comparison to their hydrophobicity and binding to keratin, these columns may be an addition to improve chromatographic models for skin permeability.

3.8.2. Micellar liquid chromatography

Micellar liquid chromatography (MLC) is a mode of reversed-phase liquid chromatography (RPLC) in which the mobile phase contains a surfactant at a concentration above the critical micellar concentration (CMC). At this concentration, micelles begin to form which form a kind of pseudo stationary phase in the mobile phase with both hydrophilic and hydrophobic areas. This phenomenon shows a good resemblance to the human skin phospholipids and leads to accurate predictions of drug partitioning over membranes. Surfactants that are often used are cationic cetyltrimethylammonium bromide (CTAB), anionic sodium dodecyl sulfate (SDS) or non-ionic polyoxyethylene 23 lauryl ether (Brij-35). The mobile phase is kept at conditions close to the human skin, such as a buffer pH of 5.5 and a temperature of 37°C. The analyzed compound partitions over three different phases until a dynamic equilibrium is established (figure 5): i.e. from the bulk solvent phase into micelles (pseudo-phase) and into the stationary phase, but also via direct transfer from micelles into the stationary phase. The partitioning of polar compounds

goes from the mobile phase into the stationary phase, while highly hydrophobic compounds go directly from micelles into the stationary phase (30,31).
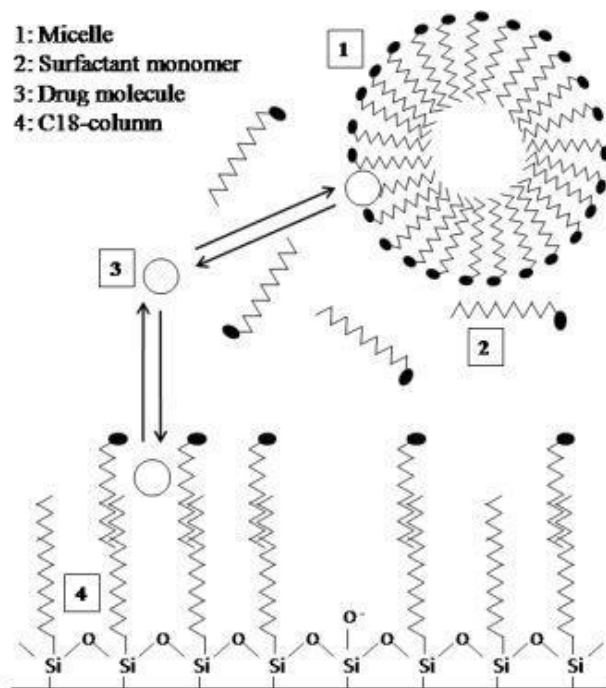
Figure 5: Interaction of a drug molecule in an MLC system (32)

Because the surfactant monomers also cover a number of free silanol groups of the stationary phase, less tailing occurs. Degradation of the column occurs also less frequently in comparison to the IAM columns (see further) since the tenside molecules are renewed permanently and protect the nonpolar chains of the stationary phase . This has as a disadvantage that mass transfer happens rather slow due to making the stationary phase more polar (31). Furthermore, this technique has been shown to be simple, cheap and fast to obtain micelle-water partition coefficient values (logPmw) which allow the prediction of the skin permeability (log K_p) via QSAR (30).

Another technique used for predicting skin permeability that uses surfactants above the critical micellar concentration is micellar electrokinetic chromatography (MEKC). This is a modification of capillary electrophoresis. Electrophoretic methods are faster and cheaper than chromatographic systems. This technique differs most from the biopartitioning process of the skin in comparison to other techniques further described (33).

3.8.3. Immobilized artificial membrane chromatography

The use of **Immobilized artificial membrane** columns (**IAM** columns) is an option to measure and predict skin permeability, which can also be classified as an RP-LC system. These columns are 'packed' with covalently bound phospholipids such as phosphatidylcholine, which functions as an artificial membrane (figure 6). The zwitterionic stationary phase includes an anionic phosphate group and a cationic choline group. Both are located into the direction of the mobile phase, making sure the mobile phase makes contact with the hydrophilic regions first (as is the case in real life human skin). Theoretically, this would be a good technique to determine skin permeability, yet a C_{18} stationary phase performs better in terms of predicting log K_p (34).
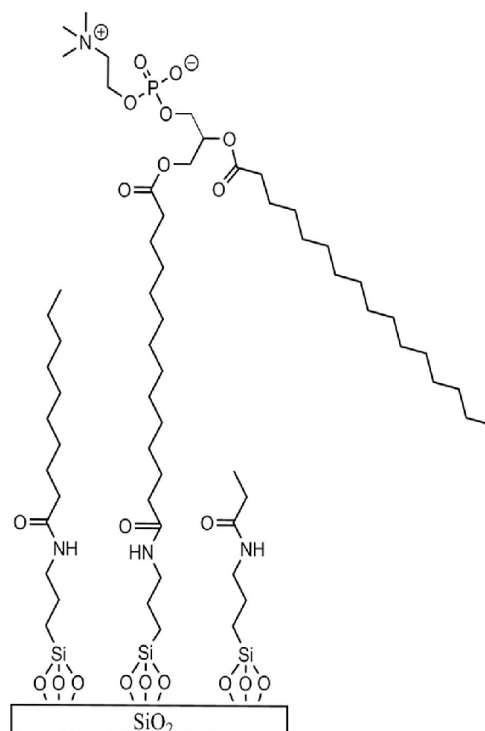
Figure 6: *Structure of an IAM column* (35)

3.8.4. C₁₈ columns

A C_{18} column has octadecyl chains bound to the silica surface, which are usually hydrophobic and highly retain nonpolar compounds. This column is used a comparative study between RP-LC (C_{18} column), IAM and micellar electrokinetic chromatography (MEKC) states that the C_{18} column comes closest to modelling the human skin permeability, the SDS MEKC and the octanol-water partition systems are decent. The IAM column and the sodium taurocholate (STC) MEKC gave the less fitted but still acceptable model for skin permeability, compared to the other studied methods. (33) The best studied alternative method, an HPLC system with a C_{18} column, was selected. Based on data from chromatographic measurements obtained by this system, a method was developed for predicting skin permeability of new compounds (33).

3.8.5. Cholesterol columns

Cholesterol is an essential building block in biological cell membranes and has a major influence on the permeability of these membranes. The use of cholesterol bonded stationary phases is therefore obvious for modelling (skin) permeability. These types of columns are excellent to predict lipophilicity and bioactivity. Studies (36) have shown that the mechanism of retention of bonded cholesterol stationary phases mainly depends on the size and shape of substances. Another quality is the great shape selectivity, a property

that depends on temperature. This is likely caused by the liquid crystal properties of cholesterol in its native state (37). Another advantage is the possibility to use highly aqueous mobile phases without the collapse of the column itself or major influence on the retention. This type of column can be used in both reversed-phase and normal-phase liquid chromatography. A correlation study investigating the retention from a cholesterol column and biological descriptors (such as skin permeability) was performed and concluded that significant similarities between chromatographic, partition and biological parameters were present (38).

3.8.6. Keratin columns

Because the *stratum corneum* contains a lot of keratin, it is only logical that skin permeability of compounds also depends on possible interactions with keratin. Since a good model for skin permeability can be accomplished using IAM columns, which is based solely on the lipophilicity of the substance, keratin columns can be a valid addition to improve these models by including interactions between drugs and keratin. These stationary phases, in which keratin is immobilized on silica support, withhold acidic substances in specific. A study concludes that this type of column can be used to quantify and measure different drug interactions with the skin protein keratin (39).

3.8.7. Supercritical fluid chromatography

Supercritical fluid chromatography (SFC) is a chromatographic technique in which the mobile phase is kept under supercritical conditions. It was originally invented as a manner to analyze thermally labile compounds that require high temperatures in gas chromatography (GC), but decompose at those temperatures. More pressure was added as a result to compensate for the high temperatures that are required. This resulted in supercritical conditions. One may describe SFC as a bridge between GC and HPLC, as the mobile phase is a supercritical fluid with properties in between a gas and a liquid. Another advantage of a supercritical fluid is the low viscosity (and high diffusivity), which creates high columns efficiencies. This technique has never been studied in the context of skin permeability (40).

3.8.7.1.   Mobile phase

The most commonly used mobile phase is carbon dioxide, which behaves as a supercritical fluid above its critical temperature of 31.0 °C and critical pressure of 73.8 bar (figure 11). Carbon dioxide is nonpolar and miscible with polar solvents such as methanol, ethanol and acetonitrile. Solvent gradients with a high percentage of organic modifier (up to 60%) (41) can be used which lead to higher viscosity resulting in (unwittingly working with) subcritical conditions. This also depends on the instrument design or the operational conditions, such

as working at a lower temperature or pressure. Slight deviations from supercritical conditions do not make significant differences in chromatography (40).
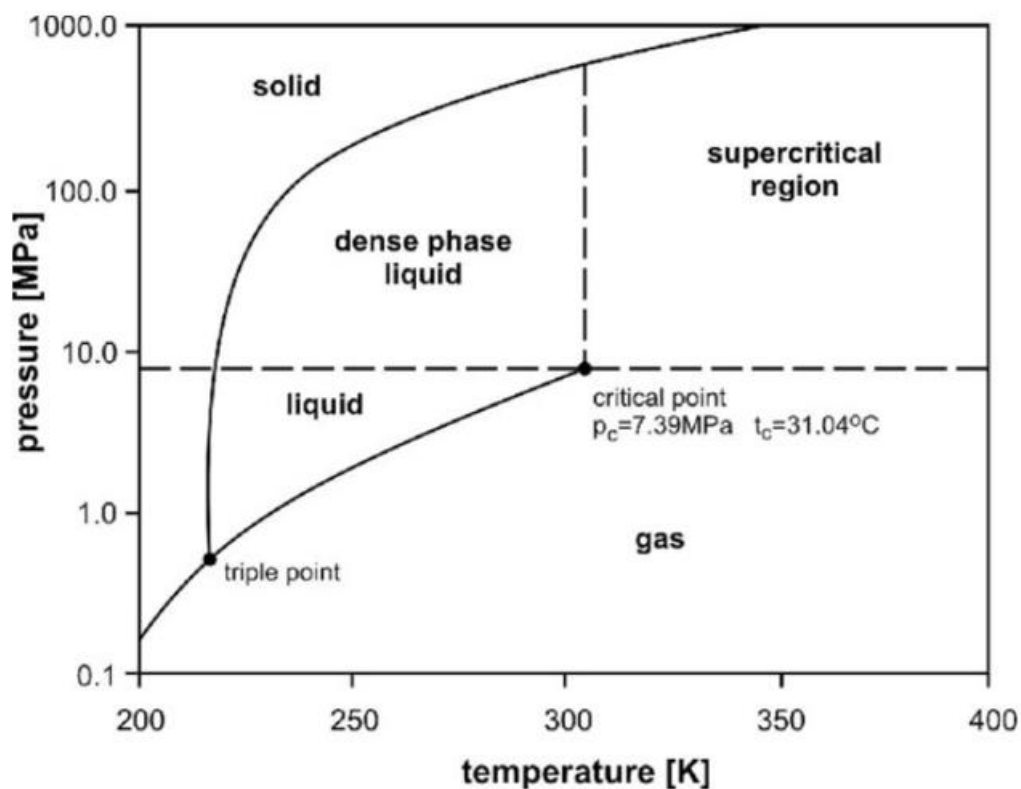
Figure 11: *Pressure-temperature phase diagram of carbon dioxide* (42)

3.8.7.2.   Stationary phase

A lot of different stationary phases can be used because of the wide polarity range of SFC. This is because CO_2 is very miscible with most organic modifiers. As a result, both reversed-phase liquid chromatography and normal-phase liquid chromatography columns can be used. A selection of the most commonly used columns is listed below (figure 7):

Stationary phase choices
Silica/BEH
2-ethylpyridine
2--picolylamine
Aminopropyl
Diol
Amide
PentaFluoro-Phenylpropyl
Phenyl
$C_{18} < C_8$

Normal-phase range

SFC range

Reversed-phase range

Figure 7: *Stationary phase options for reversed-phase, normal-phase, and supercritical fluid chromatography*

The silica base an aminopropyl (NH_2) stationary phase is chemically modified with amino groups, as can be seen in figure 8A. They function as a weak anion exchanger resulting in polar selectivity. This type of column is considered a normal-phase column (43).

Another stationary phase used in SFC is the phenyl column. The silicasurface of a phenyl-column is modified with ether-linked phenyl, as can be seen in figure 8B. This results in an increased retention of highly polar and aromatic substances, the latter is caused by π-π interactions with conjugated compounds (44).

The HILIC column (HILIC stands for *hydrophobic interaction liquid chromatography*) is also a stationary phase used in SFC. The coating of the HILIC-column consist of cross-linked diol groups bound to the silica surface. This results in better polar selectivity. This column is also suitable for use in SFC. A representation of the modified silica surface is given in figure 8C (45).

At last there are stationary phases made with Ethylene Bridged Hybrid (=BEH) technology. BEH particles are hybrid materials which consist of silica and organosiloxanes. These differ from silica because of the BEH groups which are distributed throughout the particle. Columns of this type provide interaction with polar groups such as phospholipids. Modifications are possible with multiple groups, such as C_{18}, C_8 or a phenyl-group (46–48).
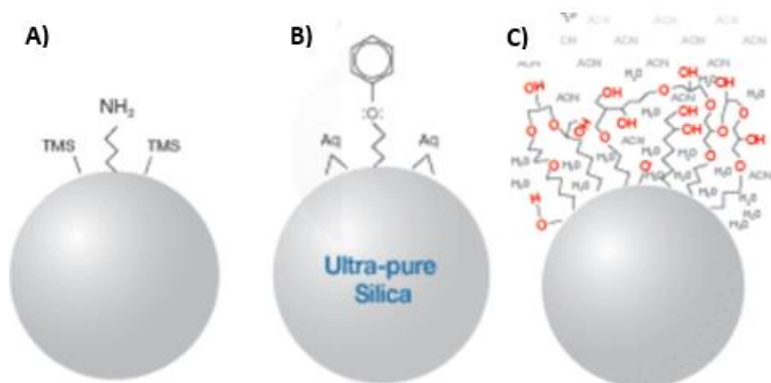
Figure 8: *A) Particle of the NH₂ column* (43) *B) Particle of the phenyl column* (44) *C) Particle of the HILIC column* (45)

3.8.7.3. Advantages/drawbacks of SFC

Advantages of SFC over HPLC are that analysis happens are faster, shorter column lengths can be used (more interactions occur in a shorter time/distance) and it produces a better resolution (up to five times in comparison to HPLC). A great advantage over methods like GC is the possibility to analyze compounds that are thermally labile, also with a high resolution. Even compounds with a high molecular weight can be analyzed due to a more solubilizing power of the supercritical mobile phase (40).

Drawbacks of SFC are mainly mobile phase or equipment related and include the limited choice of mobile phases and reactions with the mobile phase (for example, at supercritical conditions, CO_2 forms carbamic acids with primary and secondary amines). Furthermore is the analysis of extremely polar compounds not possible due to the nonpolar mobile phase, CO_2 (40).

4. Aim

The aim of this thesis is to search for an alternative method to predict skin permeability. Multiple orthogonal stationary phases (based on a previous study (49)) in SFC will be used to relate log k (the logarithm of the retention factor) to log K_p (the logarithm of skin permeability) using quantitative structure-activity relationship (QSAR) models. Fifty-eight compounds are tested to obtain log k data, using a mobile phase composition determined by a gradient screening (based on a limited number of test compounds). To optimize the model between log k and log K_p, theoretical molecular descriptors are added. To obtain proper models, MLR, stepwise MLR and PLS modelling approaches were used. Four stationary phases (HILIC, NH_2, phenyl and BEH) and two sets of theoretical molecular descriptors (Vega ZZ descriptors and E-dragon descriptors respectively) will be used to develop QSAR models for skin permeability. In the end, the obtained models for each stationary phase will be compared to empirical models and to each other, to evaluate whether log k is in fact an added value to the model and to evaluate which stationary phase is best for predicting skin permeability, respectively.

5. Methods and materials

5.1.1. Sample preparation

The test set contained 58 pharmaceutical and cosmetic compounds, covering relevant log K_p (-5.52; -0.24) and log P ranges (-1.08; 4.51). Standards with a concentration of 0.1 mg/mL were prepared from a stock solution containing 1 mg/mL in methanol (VWR international, Radnor, Pennsylvania, USA, HPLC grade). More concentrated standards (0.25 ; 0.5 ; 0.75 and 1 mg/mL in methanol) were prepared for the compounds which gave no signal at a concentration of 0.1 mg/mL.

The pharmaceutical substances used for the test set are given in table 1. The compounds were obtained from Sigma (Saint Louis, Missouri, USA), Aldrich (Milwaukee, Wisconsin, USA), Merck (Darmstadt, Germany), Bios (Brussels, Belgium), Fluka (Neu-Ulm, Switzerland), Sigma-Aldrich (Steinheim, Germany), Bios Coutelier (Anderlecht, Belgium), Diosynth OSS Holland (Oss, The Netherlands), Certa (Braine-l'Alleud, Belgium). All compounds had a purity higher than 95%.

Table 1: *Specifications of the pharmaceuticals in the test set*[1,2]

Compound	Manufacturer	Log Kp[1]	Molecular weight (Dalton)[2]
17α-hydroxyprogesterone	Sigma	-3.22	330.46
2,4,6-Trichlorophenol	Aldrich	-1.23	197.45
2,4-Dichlorophenol	Aldrich	-1.22	163.00
2-amino-4-nitrophenol	Aldrich	-3.18	154.12
2-nitro-p-phenylenediamine	Aldrich	-3.30	153.14
4-amino-2-nitrophenol	Aldrich	-2.55	154.12
Acetylsalicylic acid	Sigma	-2.14	180.16
Aminopyrine	Sigma	-2.99	231.23
Amylobarbital	Bios	-2.64	226.27
Antipyrine	Unknown	-4.18	188.23
Atropine	Sigma	-5.07	289.37
Barbital	Bios Coutelier	-3.96	184.19
Benzoic acid	Merck	-1.52	122.12
Benzyl alcohol	Sigma-Aldrich	-2.22	108.14
Caffeine	Fluka	-2.78	194.19

---

[1] Log K_p data was extracted from a database found in literature (51)
[2] Molecular weight data was extracted from Vega ZZ

Chloroxylenol	Aldrich	-1.23	156.61
Chlorpheniramine maleate	Sigma	-2.66	274.79
Cortexolone	Sigma	-4.12	346.46
Cortexone	Sigma	-3.35	330.46
Corticosterone	Sigma	-3.19	346.46
Cortisone	Sigma	-5.00	360.44
Diclofenac	Sigma	-1.74	296.15
Ephedrine(.HCl)	Aldrich	-2.22	165.23
Estriol	Unknown	-4.40	288.38
Estrone	Diosynth OSS Holland	-2.44	270.37
Flurbiprofen	Sigma	-0.34	244.26
Haloperidol	Unknown	-4.04	375.86
Hydrocortisone	Certa	-5.52	362.46
Ibuprofen	Sigma	-0.24	206.28
Indomethacin	Sigma	-1.30	357.79
Ketoprofen	Sigma	-1.23	254.28
Lidocaine	Sigma	-1.69	234.33
m-Cresol	Sigma-Aldrich	-1.82	108.14
Methyl-4-hydroxybenzoate	Fluka	-2.04	152.15
m-Nitrophenol	Sigma-Aldrich	-2.25	139.11
Naproxen	Sigma	-1.42	230.26
Nicotinate, ethyl	Aldrich	-2.20	151.16
Nicotinate, methyl	Aldrich	-2.49	137.14
o-Chlorophenol	Aldrich	-1.48	128.56
o-Cresol	Sigma-Aldrich	-1.80	108.14
Paracetamol	Sigma	-3.35	151.16
p-Cresol	Aldrich	-0.91	108.14
Phenobarbitone	Unknown	-3.35	232.24
Phenol	Merck	-1.71	94.11
Piroxicam	Sigma-Aldrich	-2.47	331.35
p-Nitrophenol	Sigma-Aldrich	-2.25	139.11
p-Phenylenediamine	Aldrich	-3.62	108.14
Prednisolone	Sigma	-4.35	360.44
Progesterone	Sigma	-2.82	314.46

Resorcinol	Merck	-3.62	110.11
Salicylic acid	Sigma	-2.20	138.12
Testosterone	Sigma	-3.40	288.42
Thiourea	Merck	-4.02	76.12
Thymol	Sigma	-1.28	150.22
Triamcinolone	Sigma	-5.40	394.43
Triamcinolone acetonide	Sigma-Aldrich	-4.69	434.50
β-Estradiol	Sigma	-2.37	272.38
β-Naphthol	Merck	-1.55	144.17

5.2. Materials

5.2.1. Experimental conditions

The analyses were performed on an Acquity Ultra Performance Convergence Chromatography (UPC²) from Waters (Milford, Massachusetts, USA), consisting of a binary solvent delivery pump, autosampler with a loop of 10 µL, a convergence manager, an external Acquity column oven without active pre-heating, a PDA detector and a back-pressure regulator.

The four stationary phases selected for these experiments were Luna NH_2 (100 mm x 4.6 mm i.d., 3 µm, Phenomenex, Utrecht, The Netherlands), Luna HILIC (100 mm x 4.6 mm i.d., 3 µm, Phenomenex), Synergi Polar RP (100 mm x 4.6 mm i.d., 4µm, Phenomenex) and Acquity UPC² BEH (100 mm x 3 mm i.d., 1.7 µm, Waters). This selection was based on a previous study (49) in which a dissimilar set of SFC columns was determined.

An injection volume of 10 µL, a flow rate of 3.0 mL/min and a backpressure of 150 bar were applied. For the BEH column an injection volume of 2 µL and a flow rate of 1.5 mL/min were used, because of the sub-2µm particles. The temperature of the column was set to 25°C and the autosampler temperature to 10°C. The wavelength used for UV detection was 220 nm and the dead time was marked by the first baseline disturbance after injection of a blank (methanol).

The mobile phase consisted of CO_2 quality 4.5 (Messer, Sint-Pieters-Leeuw, Belgium, purity ≥ 99.995%) to which methanol (HPLC grade, VWR international) was added as organic modifier. The fraction of modifier depended on the results of the gradient test. For the weak and strong wash, methanol and isopropyl alcohol (VWR international, HPLC grade) were used, respectively.

5.3. Methods

5.3.1. General Gradient

A gradient run was performed to determine the best fraction of modifier in the isocratic mobile phase. Nine compounds of the test set were selected covering the log P range (low, intermediate and high log P values). They were eluted applying a gradient of methanol, varying from 5% up to 40% in a time period of 10 min. The 40% fraction was kept for 5 min, then that the fraction was brought back to 5% within 0.5 min and held for 2 min to recondition the stationary phase (figure 9).
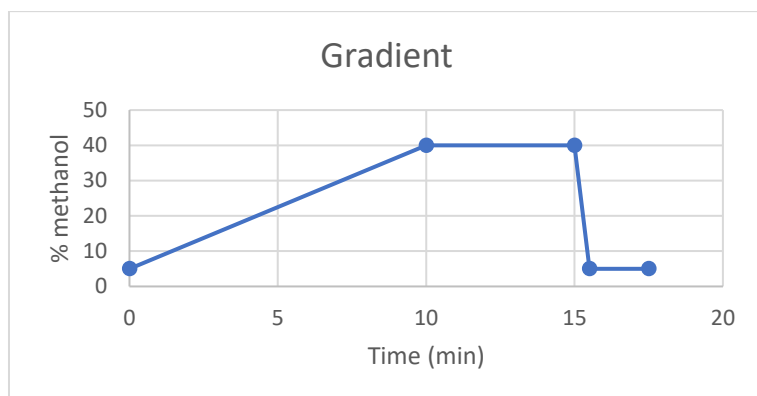
Figure 9: *Gradient profile during the gradient test.*

The final fraction of organic modifier (Ce) where each compound eluted was calculated according to equation 6 (50).

$$Ce = Ci + \frac{Cf - Ci}{t_G} \; x \; (t_r - t_d - t_0)$$ (eq. 6)

where Ci and Cf are the fractions of methanol at the beginning and end of the gradient respectively. The gradient time is represented by t_G (= 10 min), the retention time by t_r (in minutes) and the system dwell time by t_d (= 0.058 min). t_0 represents the dead time (in minutes) which is indicated with a methanol injection.

After all methanol fractions (Ce) are determined, the average was taken as the fraction of methanol used to screen the whole test set of 58 compounds.

5.3.2. Sample injections

When the fraction of modifier was determined from the gradient test, all individual standards are injected in triplicate. After obtaining the chromatograms, handled by the Empower software (Waters) of the SFC instrument, the retention factor (k) for each compound was calculated by equation 7:

$$k = \frac{(t_r - t_0)}{t_0}$$ (eq. 7)

5.3.3. Data analysis

To build a model for predicting skin permeability, molecular descriptors (theoretical and experimental) were used. Both MLR (*Multiple linear regression*) and PLS (*partial least*

squares regression) modelling were used. The software programs to model the skin permeability and calculate the descriptors were Matlab (version r2019b, The Mathworks, Natick, Massachusetts, USA), VEGA ZZ (version 3.2.1, Alessandro Pedretti & Giulio Vistoli) and E-Dragon (version 1.0, Virtual Computational Chemistry Laboratory).

First, a set of 29 molecular descriptors was generated using Vega ZZ (by importing data from PubChem, Bethesda, Maryland, USA). Log k data, obtained from the SFC experiments, were added as an experimental descriptor. The log K_p values for almost all compounds were extracted from a fully validated database (51) and have been measured *in vitro* through human skin. The coefficients had to meet multiple conclusion criteria: 1) the experiment was performed between 20-40°C, 2) the fraction of nonionized molecule must be higher than 10%, 3) the measurements had to be executed by a steady state flux, 4) the donor- and receptor fluids should not affect or damage the skin (more than water) and, 5) a decent log P must represent the measured compound. Some log K_p values were extracted from other studies, such as for acetylsalicylic acid (52), haloperidol (53), lidocaine (54) and thiourea (55). All these studies used *in-vitro* permeation techniques, involving human skin, to obtain the log K_p data.

The correlation between all descriptors was checked and if r > 0.1 with log K_p, models with one up to seven variables were build using automatic linear regression in Vega ZZ. If r < 0.1, log k was not automatically taken into account and thus was added manually for linear regression in Vega ZZ. To ensure that two variables do not contain the same information, the variance inflation factor (VIF) is determined by equation 8 (56). Generally, a VIF value above 5 might indicate collinearity and thus two independent variables contain the same information. Obviously, these two variables cannot be included in the same model. The variable with the highest correlation to log K_p is kept and the other discarded.

$$VIF = \frac{1}{1-r^2}$$
(eq. 8)

The RMSECV (*root mean squared error of cross validation, using leave-one-out CV*) and RMSEC (*root mean squared error of calibration*) were determined for each model. The RMSEC will decrease the more variables are added. The RMSECV decreases when more variables are added to the model, but can increase when too many variables are added (caused by overfitting of the model). Thus, the best model is a compromise between a good RMSEC (the best fit) and a low RMSECV (prediction properties). r² (determination coefficient) of the regression curve indicates the fitting of the model.

Stepwise MLR was performed using Matlab. Here, the correlation (r) between variables is evaluated; when a correlation coefficient above 0.95 between two variables is observed, the descriptor with the lowest correlation to log K_p is discarded. The difference with simple MLR lies in the fact that stepwise MLR adds, evaluates and possibly eliminates independent

variables one by one instead of adding and evaluating all at once. This stepwise evaluation of variables is done by an F-test (57). These models are evaluated with cross-validation (RMSECV), r² and RMSEC, as for MLR models.

PLS models are built with set of new latent variables, called PLS factors, which are combinations of the original variables, and represent new variables in the data space. An optimal number of PLS factors was determined again by cross-validation: the model with the lowest RMSECV and a low number of factors (least complex) was chosen. A regression coefficient was also determined for each individual original variable, which says something about its influence on the model.

After building PLS and MLR models for log K_p based on the 27 Vega ZZ descriptors supplemented with measured data for log k and melting point data (extracted from PubChem), modelling was also done for a larger set of molecular descriptors generated by E-dragon (1666 descriptors, of which 408 remained after deletion of the (almost) constant and the highly correlated descriptors) supplemented with log k for each column.

6. Results and discussion

6.1. Empirical models

First, models based on theoretical descriptors (so without a chromatographic descriptor) were built. This was useful to compare with the ones taking into account retention on given stationary phases (discussed in the next sections). In this way, it was determined whether the addition of a chromatographic descriptor (log k) has of an added value for the models.

The best models built using MLR techniques are given in table 2. Equation 9 describes an MLR model with Vega ZZ descriptors containing three variables, melting point, virtual log P and gyration radius. It has a relatively low accuracy (indicated by the standard error, determination coefficient and RMSEC). The predictive power of this model is also quite low according to the RMSECV. All generated MLR Vega models can be found in table A1 (appendix).

When stepwise MLR modelling was performed, two sets of molecular descriptors were used; one containing Vega ZZ descriptors, another E-dragon descriptors. Both best models are also presented in table 2.

Table 2: *Best empirical automatic MLR with VEGA ZZ and stepwise MLR models (n = 58)*

Equation	r^2	SE	RMSEC	RMSECV	Equation (log K_p = …)
Eq. 9	0.68	0.732	0.706	0.758	-1.61 - 0.0036 Melting point + 0.64 Virtual log P - 0.61 Gyration radius
Stepwise MLR					
Eq. 10	0.68	0.730	0.704	0.757	-2.46 - 0.04 Atoms + 0.68 Geometry center (1) + 0.60 Virtual log P
Eq. 11	0.91	0.420	0.378	0.452	-7.64 -0.01 D/Dr06 - 0.001 SRW09 + 1.63 BEHm1 + 5.95 JGI4 -0.34 RDF020e + 0.26 RDF055p + 0.73 Mor26u + 0.72 Mor32e - 1.20 G1u + 0.77 C-025

The model containing Vega ZZ descriptors (eq. 10) consists of three variables: atoms, geometry center 1 and virtual log P. This model does not differ much from the MLR model (eq. 9) in terms of predictability and accuracy. The model containing E-dragon descriptors (eq. 11) consists of ten variables: D/Dr06, SRW09, BEHm1, JGI4, RDF020e, RDF055p, Mor26u, Mor32e, G1u and C-025. This one is considered superior to the one based on Vega ZZ descriptors because of a higher accuracy and predictive power. For the latter model, more potential theoretical molecular descriptors were taken into account (408 versus 29). In general, the more variables added to the model, the better the fit. Additionally, the predictive properties also need to be evaluated.

PLS regression models were also built using the same two sets of molecular descriptors (table 3).

Table 3: *Best theoretical PLS models (n = 58)*

Number PLS factors	Factors with a high regression coefficient:	r^2	RMSEC	RMSECV
3 (Vega ZZ)	Virtual log P and PSA-based HLB	0.63	0.757	0.832
6 (E-dragon)	TPSA(Tot), PCD, Mor02e, D/Dr09, QXXm and RDF035e	0.72	0.674	0.860

When the Vega ZZ descriptors were used, a model containing three PLS factors was built. The molecular descriptors that have most influence on the model are virtual log P and PSA-based HLB. When E-dragon descriptors were used, a model consisting of six PLS factors was built in which the descriptors with the largest influence on the response are TPSA(Tot), PCD, Mor02e, D/Dr09, QXXm and RDF035e. When compared, the model based on E-dragon descriptors had a better fit over the model based on Vega ZZ descriptors, but the predictive power was similar (RMSECV). Since models with a low complexity are preferred, the model with Vega ZZ descriptors might be a better choice.

Overall, the PLS regression models were found inferior compared to the MLR models because of a higher value for RMSECV. The best model built with theoretical descriptors was the stepwise MLR model based on E-dragon descriptors (eq. 11).

6.2. Gradient screening

To determine which fraction of methanol should be used to obtain optimal retention times, a gradient screening was performed on each stationary phase. This screening was executed by varying the methanol fraction from 5% to 20% for the NH_2 and the HILIC column, from 5% to 35% for the BEH column and from 5% to 40% for the phenyl column. The upper limits of the methanol fraction differ because some columns were not able to handle a high system pressure and a higher modifier fraction means a higher viscosity of the mobile phase, resulting in an increased system pressure.

The retention times with associated methanol fractions (Ce, calculated according to eq. 6) of each compound in the test set are given in table 4. The system dwell time (t_d) equaled 0.058 min. The final fraction of organic modifier used to screen the test set was defines as 10% methanol for the four stationary phases.

Table 4: *Retention times and associated methanol fractions (Ce) from the gradient screening on the four stationary phases*

Compound	Stationary phases							
	NH$_2$		HILIC		Phenyl		BEH	
	Rt (min)	Ce (%)	Rt (min)	Ce (%)	Rt (min)	Ce (%)	Rt (min)	Ce (%)
Testosterone	1.83	7.0	1.57	6.6	2.26	11.2	1.36	7.7
Estriol	9.11	17.9	7.26	15.1	2.52	12.1	3.15	13.1
Corticosterone	3.67	9.8	2.75	8.4	2.75	12.9	1.89	9.2
Benzoic acid	3.44	9.4	0.92	5.6	0.64	5.5	0.63	5.5
Acetylsalicylic acid	7.19	15.0	1.13	5.9	0.71	5.8	0.78	5.9
Barbital	1.52	6.5	1.34	6.2	0.62	5.4	0.62	5.5
Caffein	0.71	5.3	0.82	5.5	1.49	8.5	0.88	6.2
Diclofenac	9.68	18.8	1.58	6.6	1.24	7.6	0.85	6.1
β-estradiol	5.74	12.9	4.33	10.7	2.34	11.5	2.02	9.7
Average Ce (%)		11.4		7.8		8.9		7.6
t$_0$ (min)	0.45		0.45		0.43		0.41	

6.3. HILIC column
6.3.1. Screening of the test set

The retention factors on the HILIC column were in the range [0.13; 113.33]. All retention times and factors can be found in table A2 (appendix). Compounds that elute early are ethyl nicotinate, methyl nicotinate and progesterone. Compounds that elute rather late are chlorpheniramine maleate, haloperidol and estriol. The dead time for this column is 0.413 min. No signal was obtained for ephedrine and p-phenylenediamine, while haloperidol required multiple injections with higher concentrations due to a weak signal; the results of 0.1 mg/mL was used in modelling. When log k is plotted against log K$_p$ (figure 10), r of the regression curve (dotted line) equals -0.615 and shows a fairly good correlation knowing that molecular descriptors were not added at this point.
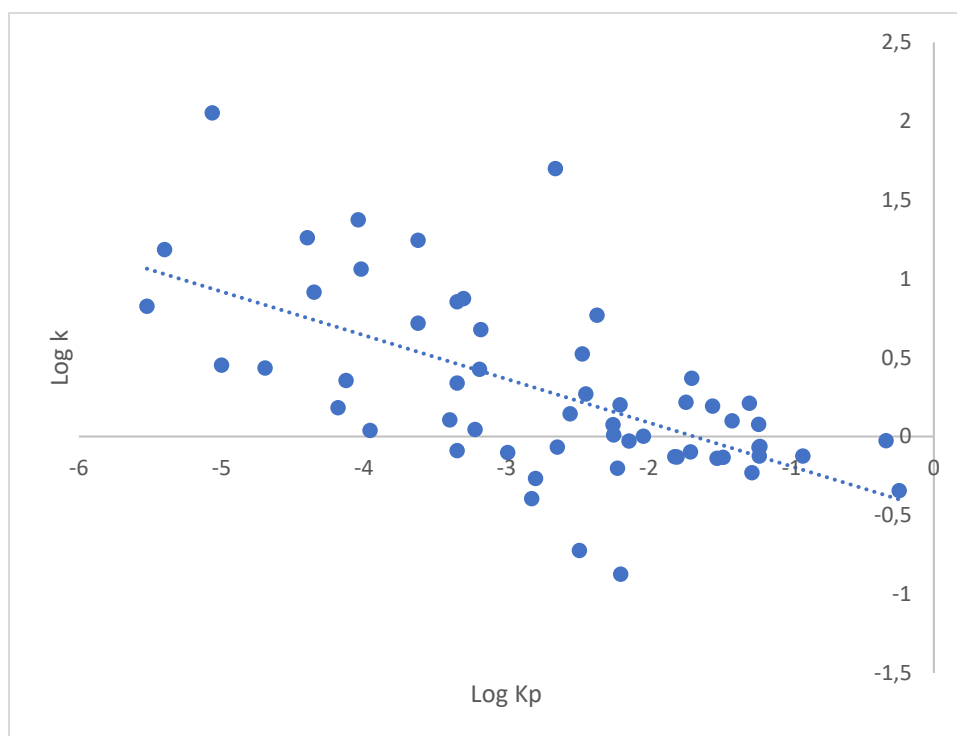
Figure 10: *The correlation between log k and log Kp for the HILIC column, together with the regression line (dotted line)*

6.3.2. Modelling the skin permeability

In this section, only the best models are given. An overview of all models built for the HILIC stationary phase can be found in the appendix table A4.

Table 5 shows the best model built using the MLR modelling. This model (eq. 12), 56 compounds are taken into account. The equation includes four different variables (molecular descriptors), namely virtual log P, log k_{HILIC} , atoms and lipole (Broto). To evaluate whether the experimentally obtained data (log k_{HILIC}) provides added value, this model is compared to the best one based on only theoretical descriptors (table 2). The determination coefficient, the standard error and the RMSEC indicate that the model including log k data from the HILIC stationary phase (eq. 12) is superior and better fitted for modelling skin permeability. The RMSECV shows also a better predictive power for the HILIC model than for the empirical one.

Table 5: *Best automatic MLR with VEGA ZZ and stepwise MLR models for the HILIC column (n = 58)*

Eq. X	r^2	SE	RMSEC	RMSECV	Equation (logKp = ….)
Eq. 12	0.81	0.581	0.555	0.607	-2.75 + 0.49 Virtual log P - 0.93 Log k_{HILIC} - 0.029 Atoms + 0.15 Lipole (Broto)
Stepwise MLR					
Eq. 13	0.82	0.559	0.529	0.602	-2.79 – 0.93 Log k_{HILIC} – 0.028 Atoms + 13.33 Mass center (3) + 0.16 Lipole (Broto)+ 0.48 Virtual log P
Eq. 14	0.94	0.343	0.307	0.368	-1.22 -1.02 Log k_{HILIC} + 0.61 Jhetp -0.82 EEig01r + 9.55 JGI4 + 0.055 G(N..N) + 0.48 Mor13m - 1.38 HATS2u + 63.30 R8v+ + 0.46 C-025 + 0.19 ALOGPS_logP

When stepwise MLR was performed (table 5), two sets of molecular descriptors were used, both including experimental log k data.

The model based on the E-dragon descriptors (eq. 14) is more accurate to predict skin permeability in comparison to the one based on Vega ZZ descriptors (eq. 13), but it is also more complex (ten variables in comparison to five). Log k_{HILIC} is included in both models. A list of the selected E-dragon descriptors with their definition and classification is listed in the appendix (table A3). When both models were compared to those based on only theoretical descriptors (eqs. 10 - 11), the model with the HILIC retention shows a higher predictive power considering the RMSECV. When compared to the previous MLR model (eq. 12), equation 13 is a less useful model, because the predictive power is nearly the same, while it is more complex (five variables versus four).

The best models built using PLS regression are presented in table 6. Two sets of molecular descriptors were again used, both including experimental log k data.

Table 6: *Best PLS models for HILIC stationary phase (n = 56)*

Number PLS factors	Factors with a high regression coefficient:	r^2	RMSEC	RMSECV
5 (Vega ZZ)	Virtual log P and PSA-based HLB	0.69	0.749	0.803
6 (E-dragon)	TPSA(Tot), PCD, D/Dr09, Mor02e, QXXm, RDF035e	0.73	0.674	0.877

When the Vega ZZ descriptors were used, a model containing five PLS factors was built. The molecular descriptors that have the most influence on the model are virtual log P and PSA-based HLB. When E-dragon descriptors were used, a model with six PLS factors was built in which the descriptors with the most influence on the response are TPSA(Tot), PCD, D/Dr09, Mor02e, QXXm, RDF035e. When compared to the previous models (table 3), it is

seen that log k_{HILIC} had no added value in these models, resulting in similar predictive power and accuracy.

Overall, the PLS model can be regarded as the model with the least predictive power. The determination coefficient, RMSEC and RMSECV of these models suggest that they are less accurate for skin permeability than those obtained by Vega MLR or stepwise MLR modelling. The stepwise MLR model (based on E-dragon descriptors, table 5) is considered the best because of its low RMSEC and RMSECV. The determination coefficient for the regression curve between the predicted and the *in-vitro* measured log K_p is given in figure 11 and shows a good fit for in sample data.
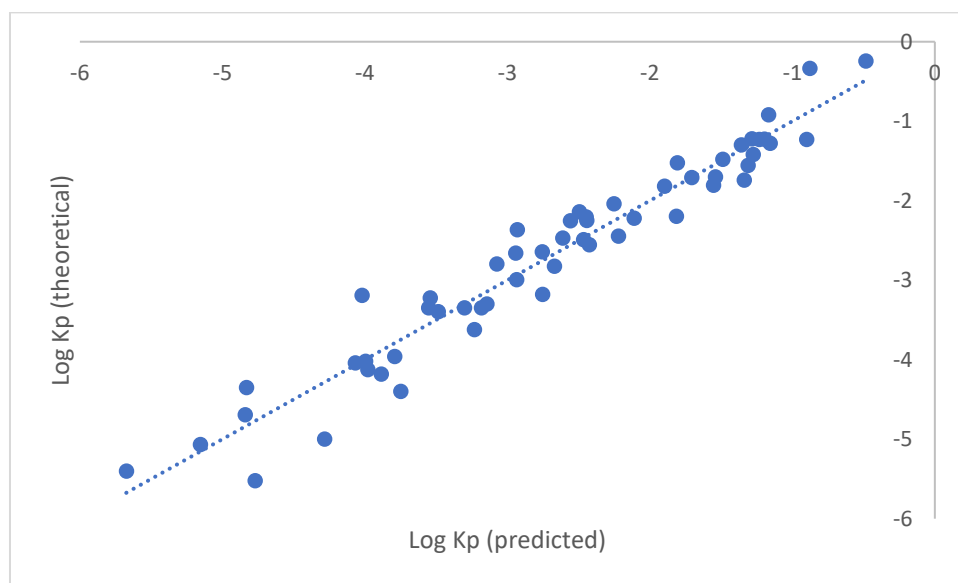
Figure 11: *The relation between predicted log K_p for the HILIC stationary phase and in-vitro measured log K_p (n = 56)*

6.4. Amino (NH₂) column
6.4.1. Screening of the test set

The retention factors for the NH₂ column range between [0.17; 66.46]. In table A2 (appendix) all retention times and factors can be found. Compounds that elute early are ethyl nicotinate, methyl nicotinate and caffeine. Final eluting compounds are diclofenac, triamcinolone and estriol. The dead time for this column equals 0.381 minutes. No signal was obtained for salicylic acid, resulting in excluding this compound when modelling. Piroxicam, p-phenylenediamine, atropine and ephedrine required multiple injections with higher concentrations due to a weak signal; the results of 1 mg/mL have been used for modelling, except for piroxicam, where the result of 0.1 mg/mL was selected. The correlation between log k and log K_p is given in figure 12 and is very poor; r of the regression curve equals 0.225 which is considered low and results in data spread like a cloud.
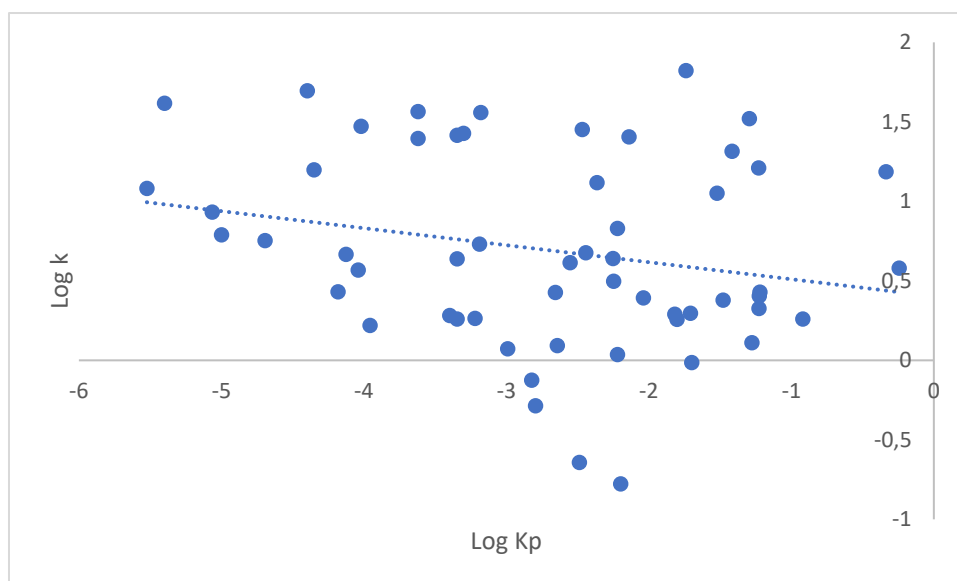
Figure 12: *Correlation between log k and log K_p for the NH_2 column, together with the regression line (dotted line)*

6.4.2. Models

Models were built using MLR, stepwise MLR and PLS regression with 57 compounds taken into account. Because the correlation coefficient between log k and log K_p was below 0.1 (figure 12), MLR models including log k were not automatically generated with the automatic linear regression module in Vega ZZ. Log k was manually added to each found MLR model where log k was not part of the model (going from one up to seven variables, table A5) and evaluated. Because the HILIC MLR models gave good results (table A4), the same selected Vega ZZ descriptors were also tested combined with the log k_{NH2} (table A5). The best model was selected and presented in table 7 (eq. 15). This model, made in analogy with the HILIC MLR model, described three variables; log k_{NH2}, atoms and virtual log P. As can be seen, the addition of log k_{NH2} compared to a empirical MLR model (eq. 9) does not make a large difference.

Table 7: *Best MLR and stepwise MLR models for the NH_2 stationary phase (n = 57)*

Eq. X	r²	SE	RMSEC	RMSECV	Equation (log K_p = ….)
Eq. 15	0.67	0.753	0.726	0.774	-2,40 – 0,25 log k_{NH2} – 0,041 Atoms + 0,60 Virtual log P
Stepwise MLR					
Eq. 16	0.68	0.732	0.706	0.759	-2.51 -0.043 Atoms + 0.71 Geometry center (1) + 0.61 Virtual log P
Eq. 17	0.90	0.427	0.388	0.457	-7.62 - 0.34 log k_{NH2} - 0.0075 D/Dr06 - 0.0010 SRW09 - 0.48 MATS4e + 1.65 BEHm1 - 0.32 RDF020e + 0.25 RDF055p - 0.90 G1u + 0.80 C-025

Table 7 also presents models built using stepwise MLR: two sets of molecular descriptors were used, both including log k_{NH2} data from the experiments.

The model based on Vega ZZ descriptors (eq. 16) solely exists of theoretical descriptors and obviously gave the same accuracy and predictive power as the empirical stepwise MLR model (eq. 10). When this model (eq.16) is compared to the previously discussed MLR model (eq. 15), it can also be concluded that log k_{NH2} data does not improve modelling for skin permeability.

Only the stepwise MLR model based on E-dragon descriptors (eq. 17) includes log k_{NH2} data. This model consists of nine variables; log k_{NH2}, D/Dr06, SRW09, MATS4e, BEHm1, RDF020e, RDF055p, G1u and C-025. A definition of the selected E-dragon descriptors can be consulted in the appendix (table A3). When this model is compared to the empirical model obtained by stepwise MLR (eq. 11), the added chromatographic descriptor log k_{NH2} seems to be of no added value to the model. When compared to the MLR model (eq. 15), a greater accuracy is observed due to a larger number of molecular descriptors taken into account, but also a greater predictive power is obtained.

Models built using PLS regression are presented in table 8. Two different sets of molecular descriptors were used, both including log k_{NH2} data from the experiments.

Table 8: *Best PLS models for NH$_2$ stationary phase (n = 57)*

Number PLS factors	Factors with a high regression coefficient:	r²	RMSEC	RMSECV
5 (Vega ZZ)	Virtual log P and PSA-based HLB	0.63	0.761	0.828
6 (E-dragon)	TPSA(Tot), PCD, Mor02e, D/Dr09, QXXm and RDF035e	0.69	0.679	0.867

When E-dragon descriptors are used, a model consisting of six PLS factors was built in which the descriptors with the biggest influence on the model are TPSA(Tot), PCD, Mor02e, D/Dr09, QXXm and RDF035e. By using Vega ZZ descriptors, a model containing five PLS factors was generated for the NH$_2$ stationary phase. The molecular descriptors causing the largest influence on the model are virtual log P and PSA-based HLB. Log k does not have a major influence on the modelling of the skin permeability. As can be seen, factors with a large influence on the model are the same for the empirical models (table 3) as well as the RMSEC, RMSECV and the r². This means that the addition of log k_{NH2} is not causing any influence on the PLS model.

The determination coefficient of the relationship between the predicted log K_p for the best model and *in-vitro* measured log K_p (figure 13) seems like a good fit at first sight. However,

because this is the same determination coefficient (and with even a slightly lower value) as for the best empirical model (eq. 11), this model can be considered useless.
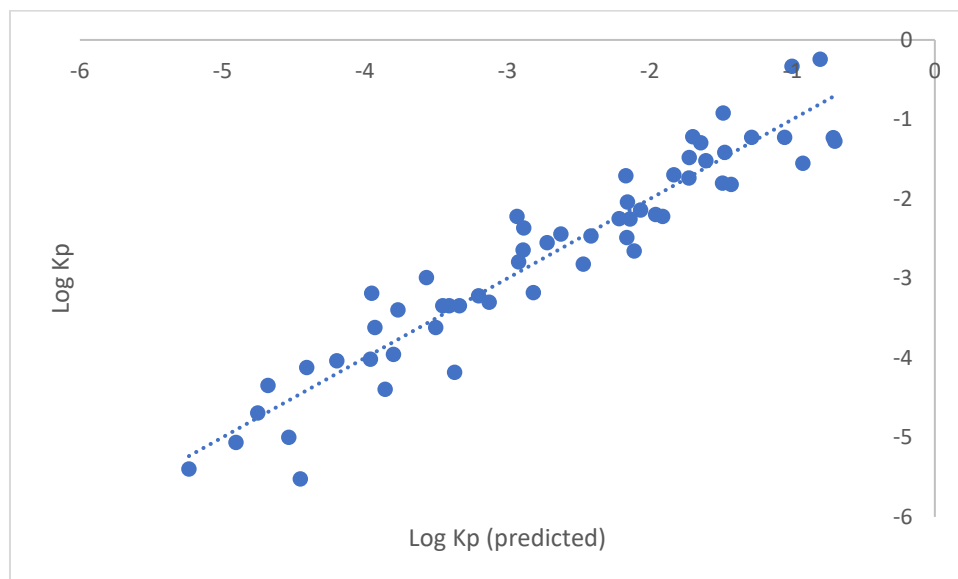
Figure 13: *The relation between predicted log K_p for the NH$_2$ stationary phase and in-vitro measured log K_p (n = 57)*

Overall, this stationary phase does not represent skin permeability well, even when multiple descriptors were added. The results of all three types of modelling are comparable to the models including only empirical models, suggesting log k_{NH2} does not offer any added value to modelling skin permeability. This NH$_2$ stationary phase is considered not being suitable to measure of skin permeability.

6.5. Phenyl column
6.5.1. Screening of the test set

All 58 compounds gave a notable signal that can be used for modelling. The retention factors for the phenyl column lay between the range [0.404; 14.144] and the dead time for this column was 0.395 minutes. Table A2 (appendix) shows all retention times and factors. First-eluting substances were barbital, amylobarbital and phenol. Compounds that elute late are chlorpheniramine maleate, atropine and haloperidol. P-phenylenediamine, haloperidol, ephedrine and atropine required multiple injections with intermediate concentrations due to a weak signal. For p-phenylenediamine and ephedrine, the results of 1 mg/mL were selected for modelling. The result of 0.75 mg/mL atropine was taken into account and for haloperidol the result of 0.5 mg/mL. Log k was plotted against log K_p (figure 14), and the correlation coefficient of the regression line gave a value of 0.614, which indicates a fairly good correlation between these 2 variables.
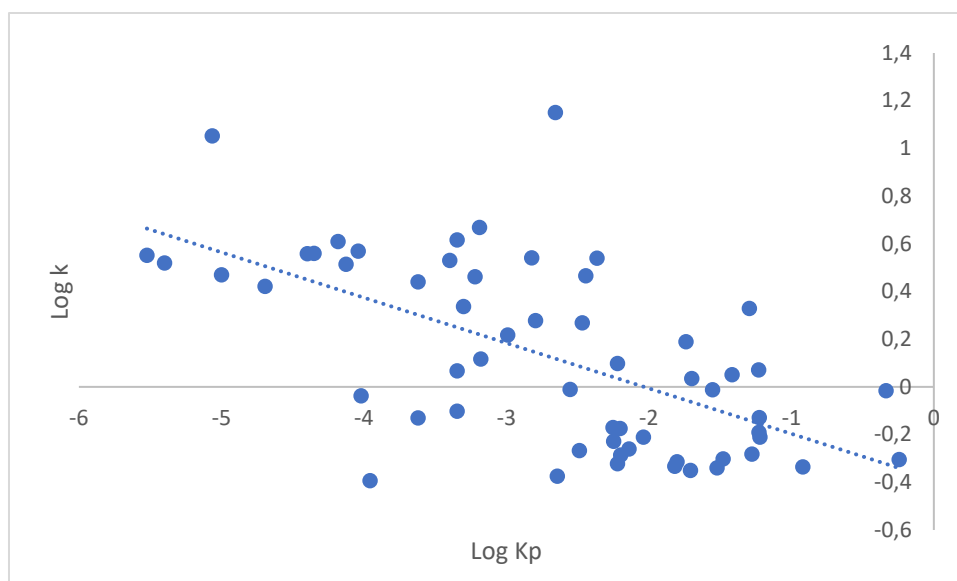
Figure 14: *Correlation between log k and log K$_p$ for the phenyl column, together with the regression line (dotted line)*

6.5.2. Models

The best models generated using the MLR modelling technique are shown in table 9. In this model, 58 compounds are taken into account. The model itself includes three different variables: melting point, log k$_{phenyl}$ and virtual log P. This model is compared to the best one (based on only theoretical descriptors (table 2)) to evaluate whether the experimentally obtained data (log k$_{phenyl}$) is an addition to an empirical equation. The determination coefficient, the standard error and the RMSEC indicate that the model including log k$_{phenyl}$ data from the phenyl stationary phase (eq. 18) is superior and more fitted for modelling skin permeability with a smaller error. Above that, the RMSECV equals a lower value, meaning that the MLR model for the phenyl column is better at predicting skin permeability than the empirical MLR model.

Table 9: *Best MLR and stepwise MLR models for the phenyl stationary phase (n = 58)*

Eq. X	r^2	SE	RMSEC	RMSECV	Equation (log K_p = ...)
Eq. 18	0.75	0.647	0.625	0.681	-3.12 – 0.0029 Melting point + 0.55 Virtual log P – 1.59 Log k_{phenyl}
Stepwise MLR					
Eq. 19	0.73	0.663	0.645	0.686	-3.56 - 1.91 Log k_{phenyl} + 0.60 Virtual log P
Eq. 20	0.98	0.207	0.167	0.238	-3.82 -1.79 Log k_{phenyl} -1.07 nBnz + 0.22 S3K -0.022 D/Dr06 -0.00059 SRW09 + 1.07 MATS1m + 0.78 EEig13x + 15.48 JGI4 + 23.51 JGI7 -0.034 G(O..O) -0.13 RDF020e + 0.076 RDF090e -0.26 Mor09e -0.83 Mor20e + 0.41 H1m + 0.25 nCconj + 0.74 C-025 + 0.24 H-047 -0.80 ALOGPS_logS

When the stepwise MLR approach is performed, two different sets of molecular descriptors were used, both including log k_{phenyl} data from the experiments. The best models are given in table 9.

A model of only two variables (eq. 19) can be obtained when the set with Vega ZZ descriptors (including the chromatographic descriptor log k_{phenyl}) is applied: log k_{phenyl} and virtual log P. This model is a lot less complex than the one based on E-dragon descriptors. The latter contains 19 variables, namely Log k_{phenyl}, nBnz, S3K, D/Dr06, SRW09, MATS1m, EEig13x, JGI4, JGI7, G(O..O), RDF020e, Mor09e, Mor20e, H1m, nCconj, C-025, H-047 and ALOGPS_logS. This one possesses a great predictive power according to the RMSECV and an almost perfect fit indicated by r^2, SE and RMSEC. When both models were compared to the empirical stepwise MLR ones, there is a clear improvement for predictability (regarding RMSECV) and in accuracy (regarding r^2, SE and RMSEC) for the ones including log k_{phenyl} data (table 9). The appendix provides a list of the selected E-dragon descriptors with their associated definition (table A3).

Models built using PLS regression are presented in table 10. Two sets of molecular descriptors were used, both including log k data from the experiments.

Table 10: *Best PLS models for the phenyl stationary phase (n = 58)*

Number PLS factors	Factors with a high regression coefficient:	r^2	RMSEC	RMSECV
5 (Vega ZZ)	Virtual log P and PSA-based HLB	0,63	0.755	0.824
6 (E-dragon)	TPSA(Tot), PCD, Mor02e, D/Dr09, QXXm and RDF035e	0,72	0.674	0.860

A model containing five PLS factors can be built when relying on Vega ZZ descriptors

including the chromatographic descriptor log k_{phenyl} (table 10). The molecular with the biggest influence on the model are virtual log P and PSA-based HLB. Another PLS regression model can be built containing six factors, using the E-dragon descriptors including log k_{phenyl} (table 10). The PLS factors with biggest influence here are TPSA(Tot), PCD, Mor02e, D/Dr09, QXXm and RDF035e. Overall, log k_{phenyl} does not seem to have a big influence on these models and can be determined by comparing these models to empirical models (table 3); the determination coefficient and the RMSECV of the latter are equal to the obtained PLS regression models including log k_{phenyl}. When compared to MLR and stepwise MLR modelling, the determination coefficient and the RMSECV of these suggest that both models are less accurate than those obtained by MLR or stepwise MLR modelling (table 9).

The best model for predicting skin permeability is considered the stepwise MLR model (with E-dragon descriptors, eq. 20). The predicted log K_p for the column is plotted against log K_p data of *in-vitro* measurements and a regression line was set up (figure 15). This model is a really good fit for in sample data because the determination coefficient shows high correlation. It also has a great predictive power according to the low RMSECV.
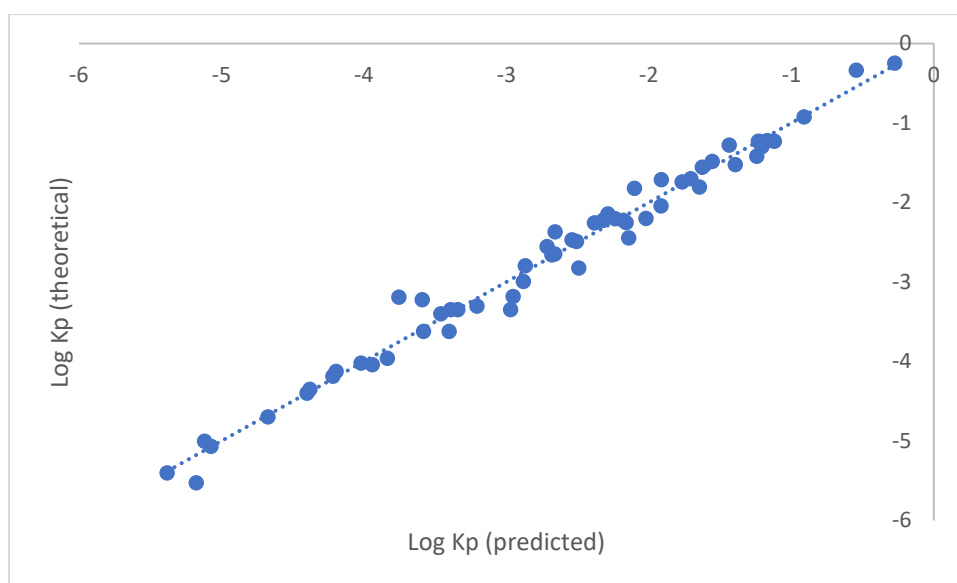
Figure 15: *The relation between predicted log K_p for the best model of the phenyl stationary phase and in-vitro measured log K_p (n = 58)*

The stepwise MLR model based on Vega ZZ descriptors (eq. 19) is also useful. Although it has a higher RMSECV over the MLR model (eq. 18), it contains 1 variable less (which comes down to the use of only two variables), which make this an interesting approach.

6.6. BEH column
6.6.1. Screening of the test set

Table A2 (appendix) shows all retention factors and times; all retention factors for the BEH column were in the range [0.134; 25.827]. Early eluting compounds are ethyl nicotinate,

methyl nicotinate and ibuprofen. Substances that elute last are haloperidol, p-phenylenediamine and piroxicam. P-phenylenediamine, piroxicam and haloperidol gave a bad signal and thus multiple injections with higher concentrations were required; the results of 1 mg/mL were chosen to model with. The dead time for this column equals 0.380 minutes and no signal was obtained for atropine, chlorpheniramine maleate and ephedrine, leading to their omission from the model. The correlation between log k and log K_p is given in figure 16, r of the regression curve (dotted line) equals 0.605, which is a fairly good fit.
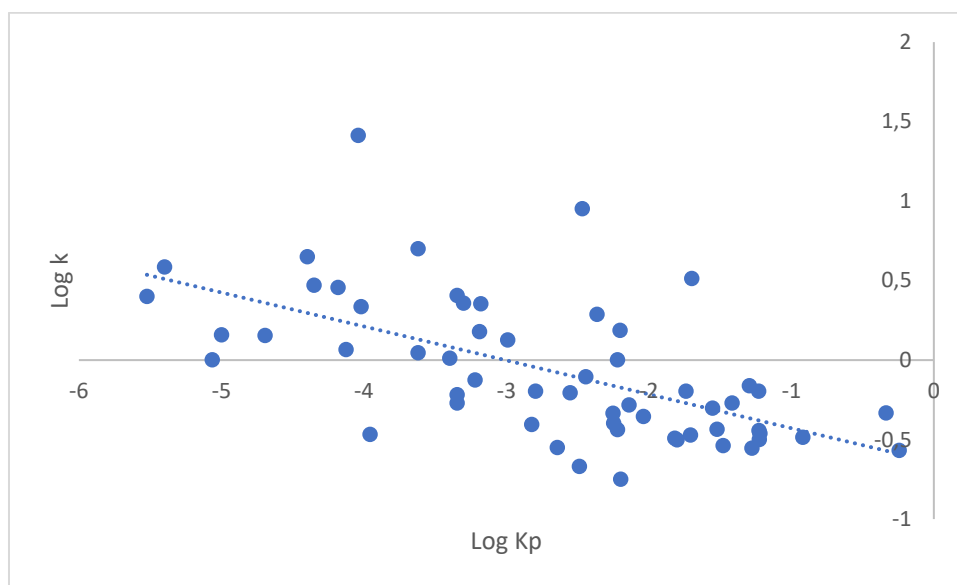
Figure 16: *Correlation between log k and log K_p for the BEH column, together with the regression line (dotted line)*

6.6.2. Models

The best model (eq. 21) built using the MLR modelling method is presented in table 11. This model includes 3 different variables (molecular descriptors) namely virtual log P, log k_{phenyl} and atoms. When compared to an empirical model (eq. 9), the RMSEC, r² and SE of the BEH MLR model show a better fit in modelling skin permeability. This suggests that the added chromatographic descriptor, log k_{BEH}, is an added value to an empirical model. The RMSECV of this model is also considered better than the one from the empirical model, meaning that the MLR model for the BEH column (eq. 21) has a greater predictive power.

Table 11: : Best MLR and stepwise MLR models for the BEH stationary phase (n = 55)

Eq. X	r²	SE	RMSEC	RMSECV	Equation (log K_p = ….)
Eq. 21	0.76	0.625	0.602	0.654	-2.86 + 0.56 Virtual log P - 0.87 Log k_{BEH} - 0.029 Atoms
Stepwise MLR					
Eq. 22	0.78	0.603	0.575	0.645	-2.88 - 0.90 Log k_{BEH} -0.027 Atoms + 14.00 Mass center (3) + 0.55 Virtual log P
Eq. 23	0.93	0.364	0.326	0.396	-3.47 -1.12 Log k_{BEH} -0.62 MATS8e + 0.19 GATS5m - 0.22 RDF020e -1.60 Mor18m + 0.75 Mor08e + 14.07 H8p -0.26 HATSp + 8.24 R4e+ + 0.41 C-025

Two different sets of molecular descriptors, both including log k_{BEH} data from the experiments, were used to generate stepwise MLR models (table 11).

The model based on the E-dragon descriptors including log k_{BEH} (eq. 23) is more accurate to predict skin permeability than the one with Vega ZZ descriptors (eq. 22) but is also more complex (ten variables versus four variables). The model based on Vega ZZ descriptors (eq. 22) includes three variables: virtual log P, log k_{BEH} and atoms. When E-dragon were used, the obtained model (eq. 23) contains ten variables; log k_{BEH}, MATS8e, GATS5m, RDF020e, Mor18m, Mor08e, H8p, HATSp, R4e and C-025. A list of selected E-dragon descriptors and accompanying definitions is provided in the appendix (table A3). Both models came out as superior when they were compared to the stepwise MLR models based on theoretical descriptors (table 2) due to a greater predictive power (RMSECV) and fit (r², RMSEC, SE). This again confirms that log k_{BEH} is an added value to the model. The same was concluded when compared to the previous obtained MLR model (eq. 21).

Models built using PLS regression are presented in table 12. Two different sets of molecular descriptors were used, both including log k_{BEH} data from the experiments.

Table 12: : Best PLS models for the BEH stationary phase (n = 55)

Number PLS factors	Factors with a high regression coefficient:	r²	RMSEC	RMSECV
5 (Vega ZZ)	Virtual log P and PSA-based HLB	0.65	0.730	0.766
6 (E-dragon)	TPSA(Tot), QXXm, PCD, RDF035e, Mor02e	0.73	0.654	0.916

The first PLS model contains 5 PLS factors and is based on the set of Vega ZZ descriptors. The molecular descriptors with the biggest influence on the model are virtual log P and PSA-based HLB. When the E-dragon descriptors are taken into account, another PLS model could be built, consisting of 6 PLS factors, in which the descriptors with the biggest

influence on the model are TPSA(Tot), QXXm, PCD, RDF035e, Mor02e. Again, log k_{BEH} does not seem to have a big influence on the final PLS models and the determination coefficient, RMSEC and RMSECV of these PLS models indicate that these are less accurate for skin permeability than those obtained by MLR or stepwise MLR modelling (table 11). When compared to the empirical PLS models (table 3), they seem to possess almost the same predictive power and accuracy as the empirical ones.

The stepwise MLR model based on E-dragon descriptors (eq. 23) turned out to be the best one to predict skin permeability for the BEH stationary phase, due to its high r² and low RMSEC, RMSECV. The correlation between the predicted values for log K_p and the log K_p data from *in-vitro* measurements in the literature (51) were given by the regression curve in figure 17. The determination coefficient shows a high correlation meaning that this model (eq. 22) is a good fit for in sample data.
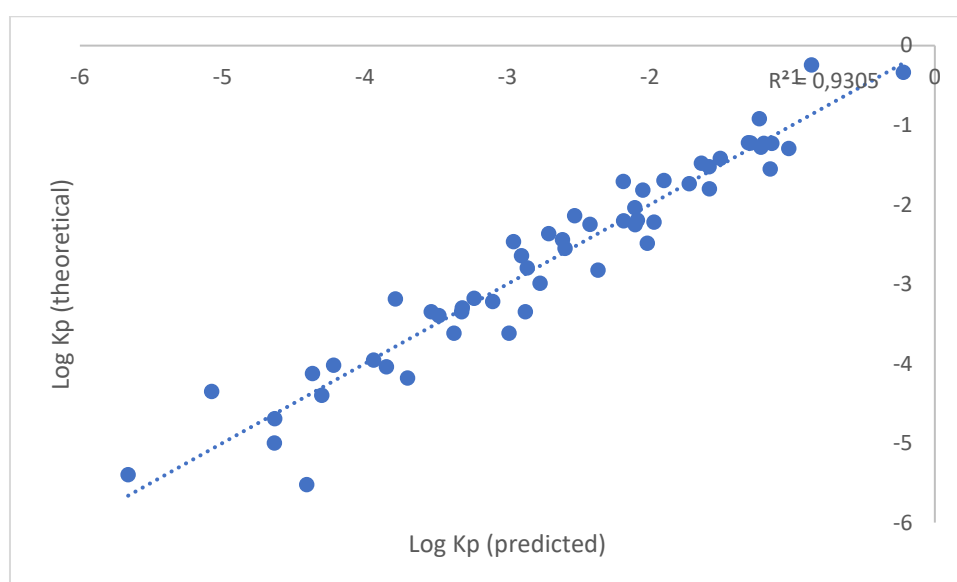
Figure 17: *The relation between predicted log K_p for the best model of the BEH stationary phase and in-vitro measured log K_p (n = 55)*

6.7. Comparison of the four different stationary phases

To decide which stationary phase is the best option for modelling skin permeability, the four stepwise MLR models based on E-dragon descriptors are compared to each other since this type of modelling gave the best results.

The model with the lowest RMSECV is selected because this suggests the best approach for out of sample data, while r² and the RMSEC are only useful for in-sample data and only gives an estimation of the accuracy of the model. Therefore, the best predictive model is the stepwise MLR model for the phenyl stationary phase based on E-dragon descriptors (eq. 20). With an RMSECV of 0.238 and an r² of 0.98, this is a nearly perfect model to

predict skin permeability. A disadvantage of this model is the large number of variables (= 19) and thus the complexity.

If a simpler model is desired, a compromise could be the MLR model of the HILIC column (eq. 12). This model contains only four variables and has a RMSECV of 0.607 (note that this is still three times the magnitude of the RMSECV of equation 20). Although there is a stepwise MLR model with a slightly lower RMSECV for this stationary phase (eq. 13), the difference is too small to be held against one more variable. Another interesting approach is the use of the stepwise MLR model based on Vega ZZ descriptors of the phenyl stationary phase (eq. 19) where only two variables are present with an RMSECV of 0.686.

Stepwise MLR gave the best results for modelling skin permeability for each stationary phase so it is only logical to choose the best column based on these models. In conclusion, the phenyl stationary phase gave the best (based on E-dragon descriptors) and the simplest (based on Vega ZZ descriptors) approach to predict skin permeability when modelled with stepwise MLR.

When the obtained data for log k of each stationary phase is compared to each other (figure 18), some interesting trends can be observed. When trying to draw a calibration line, some datapoints exceed the expected value. This is the case for chlorpheniramine maleate, haloperidol, atropine and piroxicam. It is no coincidence that these are some of the compounds that gave difficulties when measuring; they did not generate (a good) signal or the peak shape was not at all Gaussian. On the other hand, ethyl nicotinate and methyl nicotinate can easily be distinguished on the graphs, because they both have a very small log k due to low retention times. This is probably due to the fairly non-polar properties of these two substances and all four stationary phases are used to separate polar or highly polar compounds. Another finding while plotting log k data of different stationary phases to each other, is that the data obtained by the BEH and HILIC column are very much alike. This is determined by the correlation coefficient of the calibration curve, with a value for r of 0.870 (table 13). The inverse can also be concluded for the NH_2 and the phenyl column, with an r of 0.279 (table 13).

Table 13: *The correlation coefficient (r) when log k of each column is plotted against log k from another column*

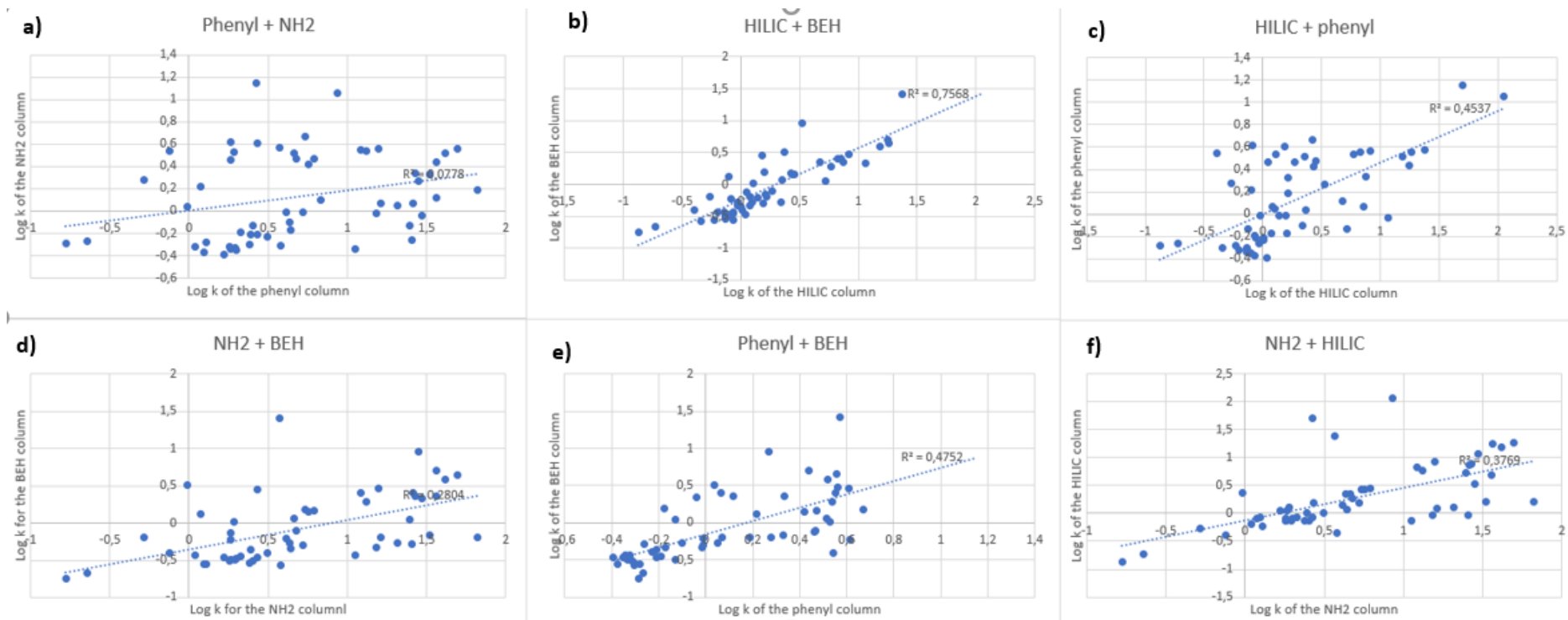| Correlation (r)	Amino	HILIC	Phenyl	BEH
Amino	1	0.614	0.279	0.530
HILIC	0.614	1	0.674	0.870
Phenyl	0.279	0.674	1	0.689
BEH	0.530	0.870	0.689	1

Figure 18: Log k for each stationary phase plotted against all three other stationary phases.

7. Conclusion

The retention times and factors of fifty-eight compounds were measured on four stationary phases using SFC. Afterwards, these results were used for modelling for skin permeability. Three different types of models were built, namely MLR, stepwise MLR and PLS regression for each column and the best for each modelling technique was selected.

After comparing all models obtained with the different stationary phases, the most suitable column to measure skin permeability was chosen. This turned out to be the phenyl column. Certain models from the BEH (eq. 23) and the HILIC column (eq. 14) also gave decent results in comparison to the best model based on only theoretical molecular descriptors (eq. 11), but displayed a lower accuracy and predictive power in comparison to the best model from the phenyl column (eq. 20). The NH_2 column was not suitable since these models (tables 8 and 9) gave no added value to the empirical ones (tables 2 and 3). The best approach to model skin permeability turned out to be stepwise MLR modelling for each chromatographic descriptor in combination with E-dragon descriptors.

This research suggests that supercritical fluid chromatography in combination with QSAR modelling can be used to respectively measure and model skin permeability as an alternative method. This method is considered a fast, easy, cheap, ethically correct and ecological method. This alternative testing method cannot measure compounds in pharmaceutical formulations or in their salt forms, making it thus only applicable in screening tests in the field of toxicology or getting a first idea of the skin permeability of a new active pharmaceutical compound. It certainly offers opportunities to the future of skin permeability assessments of pharmaceuticals and/or cosmetics.

For now, the phenyl stationary phase gave the best results but more research needs to be performed to find a stationary phase that leans closest to predicting human skin permeability. This research was limited to commercially available stationary phases, but maybe in the future the packing of SFC columns for this specific application could be looked into. Another approach could be to change the mobile phase by adding other types of organic modifier or additives in different concentrations in order to further optimize the correlation with the skin permeability.

REFERENCES

1. Abd E, Yousef SA, Pastore MN, Telaprolu K, Mohammed YH, Namjoshi S, Grice JE,S Roberts MS. Skin models for the testing of transdermal drugs [Internet]. Vol. 8, Clinical Pharmacology: Advances and Applications. Dove Medical Press Ltd; 2016 [cited 2021 Apr 25]. p. 163–76. Available from: /pmc/articles/PMC5076797/

2. Neupane R, Boddu SHS, Renukuntla J, Babu RJ, Tiwari AK. Alternatives to biological skin in permeation studies: Current trends and possibilities. Pharmaceutics. 2020;12(2).

3. Ng KW, Lau WM. Skin deep: The basics of human skin structure and drug penetration. Percutaneous Penetration Enhancers Chemical Methods in Penetration Enhancement: Drug Manipulation Strategies and Vehicle Effects. 2015(1). 3–11 p.

4. ChangWong DJ, Chang HY. Skin tissue engineering. Melton D, Cowan CA, editors. StemBook [internet]. 2009 [accessed 2021 May 22]. Available from: http://www.stembook.org/node/582

5. Gorzelanny C, Mess C, Schneider SW, Huck V, Brandner JM. Skin barriers in dermal drug delivery: Which barriers have to be overcome and how can we measure them? Pharmaceutics. 2020;12(7):1–31.

6. Zsikó S, Csányi E, Kovács A, Budai-Szűcs M, Gácsi A, Berkó S. Methods to evaluate skin penetration in vitro. Sci Pharm. 2019;87(3).

7. Schönsee CD, Bucheli TD. Experimental Determination of Octanol-Water Partition Coefficients of Selected Natural Toxins. J Chem Eng Data. 2020;65(4):1946–53.

8. Bhal SK. Application Note LogP-Making Sense of the Value [Internet]. [cited 2021 May 7]. Available from: www.acdlabs.com

9. Bos JD, Meinardi MMHM. The 500 Dalton rule for the skin penetration of chemical compounds and drugs. Exp Dermatol. 2000;9(3):165–9.

10. Pereira GR, Luís A, Ruela M, Gravinez Perissinato A, Esselin De Sousa Lino M, Mudrik PS. Evaluation of skin absorption of drugs from topical and transdermal formulations. Artic Brazilian J Pharm Sci. 2016;52(3). Available from: http://dx.doi.org/10.1590/S1984-82502016000300018

11. Couto A, Fernandes R, Cordeiro MNS, Reis SS, Ribeiro RT, Pessoa AM. Dermic diffusion and stratum corneum: A state of the art review of mathematical models. J Control Release. 2014;177(1):74–83.

12. OECD. OECD Guideline for testing of chemicals. Skin Absorption: in vitro Method (427). Test. 2004. p. 1–8. Available from: http://www.oecd-ilibrary.org/docserver/download/9742801e.pdf?expires=1455016488&id=id&accname=guest&checksum=A084AE65CC5B740E047613C22407B0B9

13. Animals used for scientific purposes - Environment - European Commission. 2009. [accessed 2021 May 1]. Available from: https://ec.europa.eu/environment/chemicals/lab_animals/3r/alternative_en.htm

14. Franz Cell - The Original - PermeGear. [accessed 2021 Apr 30]. Available from: https://permegear.com/franz-cells/

15. Franz Flow Cell - PermeGear. [accessed 2021 Apr 30]. Available from: https://permegear.com/franz-flow-cell/

16. Finnin B, Walters KA, Franz TJ. In vitro skin permeation Methodology. In: Transdermal and Topical Drug Delivery. 2012. p. 85–108.

17. Holmgaard R, Nielsen JB. Dermal absorption of pesticides – evaluation of variability and prevention. Danish Environ Prot Agency. 2009;(124):116. Available from:

https://www2.mst.dk/udgiv/publications/2009/978-87-7052-980-8/pdf/978-87-7052-981-5.pdf

18.   OECD. OECD Guideline for the testing of chemicals. Skin Absorption: in vitro Method (428). Test. 2004;(April):1–8.

19.   Parallel Artificial Membrane Permeability Assay (PAMPA) Kit | PAMPA-096 | Universal Biologicals. [accessed 2021 Apr 30]. Available from: https://www.universalbiologicals.com/parallel-artificial-membrane-permeability-assay-pampa-kit-pampa-096

20.   Henning A, Schaefer UF, Neumann D. Potential pitfalls in skin permeation experiments: Influence of experimental factors and subsequent data evaluation. Eur J Pharm Biopharm. 2009;72(2):324–31. Available from: http://dx.doi.org/10.1016/j.ejpb.2008.07.016

21.   Uchida T, Kadhum WR, Kanai S, Todo H, Oshizaka T, Sugibayashi K. Prediction of skin permeation by chemical compounds using the artificial membrane, Strat-M^{TM}. Eur J Pharm Sci. 2015;67:113–8.

22.   Kaur L, Singh K, Paul S, Singh S, Singh S, Jain SK. A Mechanistic Study to Determine the Structural Similarities Between Artificial Membrane Strat-M^{TM} and Biological Membranes and Its Application to Carry Out Skin Permeation Study of Amphotericin B Nanoformulations. AAPS PharmSciTech. 2018;19(4):1606–24.

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

24.   Chang YC, Chen CP, Chen CC. Predicting skin permeability of chemical substances using a quantitative structure-activity relationship. In: Procedia Engineering. Elsevier Ltd (Amsterdam, The Netherlands); 2012. p. 875–9.

25.   Basak SC, Mills D, Mumtaz MM. A quantitative structure-activity relationship (QSAR) study of dermal absorption using theoretical molecular descriptors. SAR QSAR Environ Res. 2007 [accessed 2021 May 1];18(1–2):45–55. Available from: https://pubmed.ncbi.nlm.nih.gov/17365958/

26.   Chen LJ, Lian GP, Han LJ. Prediction of human skin permeability using artificial neural network (ANN) modeling. Acta Pharmacol Sin. 2007;28(4):591–600.

27.   Abraham MH, Martins F. Human skin permeation and partition: General linear free-energy relationship analyses. J Pharm Sci. 2004;93(6):1508–23.

28.   Soriano-Meseguer S, Fuguet E, Port A, Rosés M. Estimation of skin permeation by liquid chromatography. ADMET DMPK. 2018;6(2):140–52.

29.   Turowski M, Kaliszan R. Collagen immobilised on silica derivatives as a new stationary phase for HPLC. Biomed Chromatogr. 1998;12(4):187–92.

30.   Waters LJ, Shahzad Y, Stephenson J. Modelling skin permeability with micellar liquid chromatography. Eur J Pharm Sci. 2013;50(3–4):335–40.

31.   El-Shaheny RN, El-Maghrabey MH, Belal FF. Micellar liquid chromatography from green analysis perspective. Open Chem. 2015;13(1):877–92.

32.   De Vrieze M, Lynen F, Chen K, Szucs R, Sandra P. Predicting drug penetration across the blood-brain barrier: Comparison of micellar liquid chromatography and immobilized artificial membrane liquid chromatography. Anal Bioanal Chem. 2013;405(18):6029–41.

33.   Hidalgo-Rodríguez M, Soriano-Meseguer S, Fuguet E, Ràfols C, Rosés M. Evaluation of the suitability of chromatographic systems to predict human skin permeation of

neutral compounds. Eur J Pharm Sci. 2013;50(5):557–68.

34.   Lázaro E, Rà C, Abraham MH, Rosés M. Chromatographic Estimation of Drug Disposition Properties by Means of Immobilized Artificial Membranes (IAM) and C18 Columns. 2006 [accessed 2021 May 10]; Available from: https://pubs.acs.org/sharingguidelines

35.   Doležal R, Karásková N, Musil K, Novák M, Maltsevskaya N V., Maliňák D, et al. Characterization of the Penetration of the Blood–Brain Barrier by High-Performance Liquid Chromatography (HPLC) Using a Stationary Phase with an Immobilized Artificial Membrane. Anal Lett. 2018;51(15):2401–14. Available from: https://doi.org/10.1080/00032719.2018.1424175

36.   Courtois C, Allais C, Constantieux T, Rodriguez J, Caldarelli S, Delaurent C. Cholesteric bonded stationary phases for high-performance liquid chromatography: Synthesis, physicochemical characterization, and chromatographic behavior of a phospho-cholesteric bonded support. A new way to mimic drug/membrane interactions? Anal Bioanal Chem. 2008;392(7–8):1345–54.

37.   Pesek JJ, Matyska MT, Dawson GB, Wilsdorf A, Marc P, Padki M. Cholesterol bonded phase as a separation medium in liquid chromatography: Evaluation of properties and applications. J Chromatogr A. 2003;986(2):253–62.

38.   Janicka M, Sztanke M, Sztanke K. Reversed-phase liquid chromatography with octadecylsilyl, immobilized artificial membrane and cholesterol columns in correlation studies with in silico biological descriptors of newly synthesized antiproliferative and analgesic active compounds. J Chromatogr A. 2013;1318:92–101.

39.   Turowski M, Kaliszan R. Keratin immobilized on silica as a new stationary phase for chromatographic modelling of skin permeation. J Pharm Biomed Anal. 1997;15(9–10):1325–33.

40.   Makansi M, R. Barron A. 3.4 Supercritical Fluid Chromatography. In: Physical Methods in Chemistry and Nano Science. 2019. p. 362–72.

41.   Liu J, Regalado EL, Mergelsberg I, Welch CJ. Extending the range of supercritical fluid chromatography by use of water-rich modifiers. Org Biomol Chem. 2013;11(30):4925–9.

42.   Witkowski A, Majkut M, Rulik S. Analysis of pipeline transportation systems for carbon dioxide sequestration. Arch Thermodyn. 2014;35(1):117–40.

43.   HPLC Column | Phenomenex | Luna. [accessed 2021 Apr 30]. Available from: https://www.phenomenex.com/Products/Detail/Luna/NH2

44.   HPLC Column | Synergi | Phenomenex. [accessed 2021 Apr 30]. Available from: https://www.phenomenex.com/products/detail/Synergi

45.   HPLC Column | Phenomenex | Luna. [accessed 2021 May 1]. Available from: https://www.phenomenex.com/Products/Detail/Luna/HILIC

46.   Synthesis and Applications of BEH Particles in Liquid Chromatography. [accessed 2021 May 1]. Available from: https://www.chromatographyonline.com/view/synthesis-and-applications-beh-particles-liquid-chromatography-0

47.   SFC Columns | Torus, Viridis SFC Column Separations | Waters. [accessed 2021 Apr 30]. Available from: https://www.waters.com/waters/en_US/SFC-Columns---Trefoil%2C-Torus%2C-Viridis-/nav.htm?locale=en_US&cid=134696052

48.   BEH (Ethylene Bridged Hybrid) Technology | Waters. [accessed 2021 Apr 30].

Available from: https://www.waters.com/waters/en_US/BEH-(Ethylene-Bridged-Hybrid)-Technology/nav.htm?cid=134618172&locale=en_US

49. Galea C, Mangelings D, Vander Heyden Y. Method development for impurity profiling in SFC: The selection of a dissimilar set of stationary phases. J Pharm Biomed Anal. 2015;111:333–43.

50. D'Attoma A, Grivel C, Heinisch S. On-line comprehensive two-dimensional separations of charged compounds using reversed-phase high performance liquid chromatography and hydrophilic interaction chromatography. Part I: Orthogonality and practical peak capacity considerations. J Chromatogr A. 2012;1262:148–59.

51. Vecchia B, Bunge A. Skin absorption databases and predictive equations in transdermal drug delivery. In: Guy RH, Hadgraft J, editors. Skin absorption databases and predictive equations In Transdermal Drug Delivery. 2nd ed. New York: Marcel Dekker; 2003. p. 57–171.

52. Degim IT, Pugh WJ, Hadgraft J. Skin permeability data: Anomalous results. Int J Pharm. 1998;170(1):129–33.

53. Azarbayjani AF, Lin H, Yap CW, Chan YW, Chan SY. Surface tension and wettability in transdermal delivery: a study on the in-vitro permeation of haloperidol with cyclodextrin across human epidermis. J Pharm Pharmacol. 2010;62(6):770–8.

54. Thomas J, Majumdar S, Wasdo S, Majumdar A, Sloan KB. The effect of water solubility of solutes on their flux through human skin in vitro: An extended Flynn database fitted to the Roberts-Sloan equation. Int J Pharm. 2007;339(1–2):157–67.

55. Hoang KT. Dermal Exposure Assesment: Principes and Applications. Washington D.C.; 1992.

56. Variance Inflation Factor (VIF) - Overview, Formula, Uses. [accessed 2021 Apr 30]. Available from: https://corporatefinanceinstitute.com/resources/knowledge/other/variance-inflation-factor-vif/

57. Deconinck E, Zhang MH, Petitet F, Dubus E, Ijjaali I, Coomans D, Vander Heyden Y. Boosted regression trees, multivariate adaptive regression splines and their two-step combinations with multiple linear regression or partial least squares to predict blood-brain barrier passage: A case study. Anal Chim Acta. 2008 Feb 18;609(1):13–23.

8. Abstract

As the interest in dermal absorption and permeation of medicines rises in the pharmaceutical and cosmetic industry, alternatives to animal testing are highly in demand. Next to *in vitro* skin permeation methods, chromatographic techniques can be used. The use of supercritical fluid chromatography (SFC) has never been tested for modelling skin permeability, although it is a fast, cheap and ethical approach. In this thesis, the skin permeability is modelled using the retention obtained in SFC. Four orthogonal stationary phases were selected based on previous studies, namely a HILIC, amino, phenyl and BEH column. A gradient screening was performed on nine compounds that represent the log P range of the test set, to determine the fraction of modifier (methanol) needed in the mobile phase to screen the complete test set. This resulted in an average of 10% methanol for all stationary phases. Once retention times and retention factors were obtained for all 58 compounds, models were built using multiple linear regression (MLR) stepwise MLR and partial least squares (PLS) regression to predict skin permeability. In order to optimize these models, theoretical molecular descriptors were added. These descriptors were extracted from Vega ZZ or E-dragon software. Skin permeability coefficients, log K_p, were extracted from the literature containing databases from *in-vitro* experiments on human skin. The models obtained with the chromatographic retentions on the different stationary phases were compared to select which column gave the best results and predictive power. The best model turned out to be a stepwise MLR model containing E-dragon descriptors and the chromatographic descriptor from the phenyl column. This model contained 19 different variables, which makes it very accurate but also complex. It showed a high predictive power due to a low RMSECV. When compared to empirical models, this model turned out to be much more accurate and predictive than the empirical model based on E-dragon descriptors, meaning that these log k data provided an added value to the model. This suggests that SFC could be used as an alternative testing method for measuring skin permeability.

9. Samenvatting

Naarmate de interesse in dermale absorptie en permeatie van geneesmiddelen toeneemt in de farmaceutische en cosmetische industrie, is er veel vraag naar alternatieven voor dierproeven. Naast *in-vitro* methoden voor huidpermeatie kunnen chromatografische technieken worden gebruikt. Het gebruik van superkritische vloeistofchromatografie (SFC) is nooit getest voor het modelleren van de doorlaatbaarheid van de huid, hoewel het een snelle, goedkope en ethische benadering is. In deze thesis wordt de doorlaatbaarheid van de huid gemodelleerd met behulp van de retentie verkregen door SFC. Op basis van eerdere studies werden vier orthogonale stationaire fasen geselecteerd, namelijk een HILIC-, amino-, fenyl- en BEH-kolom. Er werd een gradiëntscreening uitgevoerd op negen stoffen die de log P-spreiding van de test set vertegenwoordigen om de fractie organisch solvent (methanol) te bepalen die nodig is in de mobiele fase voor het meten van de volledige test set. Dit resulteerde in een afgerond gemiddelde van 10% methanol voor iedere stationaire fase. Deze fractie werd vervolgens gebruikt (isocratisch) en zodra retentietijden en retentiefactoren waren verkregen voor alle 58 stoffen in de test set, konden modellen gebouwd worden met behulp van multiple linear regression (MLR), stepwise MLR en partial least squares (PLS) regressie om de doorlaatbaarheid van de huid te voorspellen. Om deze modellen te verbeteren, werden theoretische moleculaire descriptoren toegevoegd aan de chromatografische descriptor. Deze descriptoren zijn verkregen uit Vega ZZ of E-dragon software. De coëfficiënten voor de doorlaatbaarheid van de huid, log K_p, werden uit de literatuur (databanken) gehaald en werden verkregen op basis van *in-vitro* experimenten op de menselijke huid. De verschillende manieren van modelleren, verkregen op basis van de resultaten van iedere stationaire fase, werden binnen elke stationaire fase met elkaar vergeleken en de beste modellen werden onderling tussen de verschillende stationaire fasen vergeleken om te selecteren welke kolom het beste resultaat en voorspellend vermogen gaf. Het beste model bleek een stepwise MLR-model te zijn op basis van E-dragon-descriptoren met de chromatografische descriptor van de fenyl kolom. Dit model bevatte 19 verschillende variabelen, wat het zeer nauwkeurig maar ook complex maakt. Het bezit een hoog voorspellend vermogen dankzij een lage RMSECV. In vergelijking met uitsluitend theoretische modellen bleek dit model veel nauwkeuriger en beter voorspellend te zijn dan het theoretische model op basis van E-dragon-descriptoren, wat betekent dat deze toegevoegde log k-data een meerwaarde zijn aan het model. Dit suggereert dat SFC kan worden gebruikt als een alternatieve testmethode voor het meten van de doorlaatbaarheid van de huid.

APPENDIX

Table A1: All generated MLR models based on theoretical descriptors from Vega ZZ (n = 58)

Number of independent variables	r^2	SE	RMSEC	RMSECV	Equation (log K_p = ...)
1	0.39	0.988	0.971	1.002	-1.26 - 0.011 Melting point
2	0.66	0.751	0.731	0.768	-1.66 + 0.73 Virtual log P - 0.82 Gyration radius
3	0.68	0.732	0.706	0.758	-1.61 - 0.0036 Melting point + 0.64 Virtual log P - 0.61 Gyration radius
4	0.69	0.721	0.734	0.801	-2.26 + 0.42 Virtual log P - 0.040 Atoms + 0.19 Lipole (Broto) - 0.25 H-bond donor
5	0.71	0.711	0.691	0.771	-2.54 + 0.49 Virtual log P + 0.0060 Polar area (PSA) - 0.0074 Volume + 0.23 Lipole (Broto) - 0.44 H-bond donor
6	0.73	0.693	0.663	0.764	-2.66 - 0.0041 Melting point + 0.43 Virtual log P - 0.032 Atoms + 0.0076 Polar area (PSA) + 0.26 Lipole (Broto) - 0.42 H-bond donor
7	0.73	0.698	0.660	0.776	-2.78 - 0.0041 Melting point + 0.42 Virtual log P - 0.036 Atoms + 0.0075 Polar area (PSA) + 0.25 Lipole (Broto) - 0.42 H-bond donor + 0.054 Appx. Dimensions (3)

Table A2: Retention times and retention factors for 58 compounds (concentration of 0.1 mg/mL, unless mentioned otherwise) measured on four stationary phases

	Stationary phases							
	HILIC (n = 56)		Phenyl (n = 58)		NH$_2$ (n = 57)		BEH (n = 55)	
Compound	Average Rt (minutes)	k	Average Rt (minutes)	k	Average Rt (minutes)	k	Average Rt (minutes)	k
17α-Hydroxyprogesterone	0.871	1.109	1.539	2.896	1.082	1.840	0.664	0.747
2.4.6-Trichlorophenol	0.724	0.753	0.688	0.743	1.351	2.545	0.500	0.316
2.4-Dichlorophenol	0.770	0.865	0.638	0.615	1.405	2.687	0.511	0.345
2-Amino-4-nitrophenol	2.382	4.768	0.912	1.308	14.20	36.28	1.239	2.261
2-Nitro-p-phenylenediamine	3.521	7.525	1.253	2.171	10.59	26.81	1.244	2.275
4-Amino-2-nitrophenol	0.988	1.393	0.780	0.975	1.953	4.125	0.616	0.621
Acetylsalicylic acid	0.801	0.939	0.612	0.549	10.10	25.51	0.578	0.522

Aminopyrine	0.741	0.793	1.047	1.651	0.833	1.186	0.887	1.333
Amylobarbital	0.768	0.860	0.562	0.423	0.853	1.240	0.487	0.282
Antipyrine	1.044	1.527	2.001	4.065	1.409	2.699	1.466	2.857
Atropine	(c = 1 mg/mL) 47.22	113.3	(c = 0.75mg/mL) 4.847	11.27	(c = 1 mg/mL) 3.65	8.586	/	/
Barbital	0.864	1.093	0.555	0.404	1.013	1.658	0.509	0.339
Benzoic acid	0.713	0.727	0.576	0.457	4.670	11.26	0.519	0.367
Benzyl alcohol	0.673	0.630	0.583	0.477	0.796	1.090	0.519	0.365
Caffeine	0.638	0.544	1.144	1.896	0.579	0.519	0.621	0.635
Chloroxylenol	0.766	0.855	0.650	0.646	1.190	2.123	0.517	0.361
Chlorpheniramine (maleate)	(c = 1 mg/mL) 21.15	50.20	(c= 1 mg/mL) 5.982	14.14	1.400	2.675	/	/
Cortexolone	1.350	2.269	1.686	3.268	2.149	4.640	0.822	1.163
Cortexone	0.750	0.817	2.025	4.127	1.076	1.823	0.610	0.604
Corticosterone	1.516	2.672	2.239	4.668	2.437	5.396	0.953	1.508
Cortisone	1.587	2.843	1.561	2.952	2.734	6.177	0.928	1.441
Diclofenac	1.094	1.649	1.005	1.545	25.70	66.46	0.622	0.636
Ephedrine	/	/	(c = 1 mg/mL) 0.889	1.251	(c = 1 mg/mL) 2.956	6.759	/	/
Estriol	7.963	18.28	1.822	3.613	19.30	49.65	2.075	4.461
Estrone	1.183	1.864	1.550	2.923	2.195	4.761	0.678	0.783
Flurbiprofen	0.802	0.943	0.775	0.962	6.230	15.35	0.556	0.464
Haloperidol	10.21	23.73	1.862	3.713	1.794	3.710	(c = 1mg/mL) 10.19	25.83
Hydrocortisone	3.189	6.722	1.801	3.559	4.991	12.10	1.332	2.506
Ibuprofen	0.600	0.454	0.591	0.495	1.833	3.812	0.482	0.269
Indomethacin	1.086	1.630	1.238	2.133	13.03	33.19	0.641	0.687
Ketoprofen	0.908	1.198	0.860	1.177	6.577	16.26	0.622	0.638
Lidocaine	1.381	2.343	0.823	1.084	0.751	0.971	1.617	3.255
m-Cresol	0.722	0.747	0.579	0.465	1.125	1.953	0.502	0.322
Methyl-4-hydroxybenzoate	0.828	1.006	0.638	0.615	1.323	2.472	0.548	0.442

m-Nitrophenol	0.836	1.023	0.628	0.591	1.578	3.143	0.532	0.401
Naproxen	0.932	1.257	0.839	1.123	8.250	20.65	0.583	0.535
Nicotinate. ethyl	0.468	0.134	0.600	0.518	0.445	0.167	0.448	0.178
Nicotinate. methyl	0.491	0.190	0.608	0.540	0.468	0.228	0.461	0.214
o-Chlorophenol	0.718	0.739	0.592	0.499	1.295	2.398	0.490	0.289
o-Cresol	0.719	0.741	0.586	0.484	1.071	1.811	0.499	0.314
Paracetamol	3.375	7.172	0.856	1.167	10.32	26.09	1.346	2.543
p-cresol	0.723	0.751	0.577	0.462	1.074	1.820	0.504	0.326
Phenobarbitone	1.315	2.183	0.707	0.790	2.037	4.346	0.584	0.536
Phenol	0.744	0.801	0.571	0.446	1.135	1.980	0.508	0.337
Piroxicam	1.792	3.338	1.127	1.853	11.19	28.38	3.779	8.945
p-Nitrophenol	0.905	1.190	0.662	0.676	2.046	4.369	0.555	0.461
p-phenylenediamine	/	/	(c = 1 mg/mL) 1.482	2.752	(c = 1 mg/mL) 14.39	36.77	(c = 1 mg/mL) 2.287	5.018
Prednisolone	3.813	8.232	1.830	3.634	6.416	15.84	1.502	2.954
Progesterone	0.580	0.404	1.768	3.477	0.668	0.752	0.529	0.392
Resorcinol	2.579	5.244	0.688	0.742	9.862	24.89	0.803	1.112
Salicylic acid	1.070	1.592	0.659	0.668	/	/	0.963	1.535
Testosterone	0.941	1.278	1.733	3.388	1.110	1.913	0.770	1.025
Thiourea	5.184	11.55	0.758	0.918	11.68	29.65	1.203	2.166
Thymol	0.657	0.591	0.602	0.523	0.874	1.294	0.486	0.279
Triamcinolone	6.744	15.33	1.702	3.309	16.16	41.40	1.841	3.846
Triamcinolone acetonide	1.539	2.726	1.436	2.635	2.539	5.663	0.922	1.425
β-estradiol	2.848	5.897	1.762	3.460	5.388	13.14	1.115	1.933
β-Naphthol	1.057	1.559	0.780	0.974	2.378	5.241	0.569	0.498

Table A3: *List of selected E-dragon descriptors*

Descriptor	Definition	Category
ALOGPS_logP	Log P calculated by the ALOGPS application	Molecular properties
ALOGPS_logS	Log S calculated by the ALOGPS application	Molecular properties
BEHm1	Highest eigenvalue n. 1 of Burden matrix / weighed by atomic masses	Burden eigenvalues
C-025	R--CR--R	Atom-centred fragments
D/Dr06	distance/detour ring index of order 6	Ring descriptors
D/Dr09	distance/detour ring index of order 9	Ring descriptors
EEig01r	eigenvalue n. 1 from augmented edge adjacency mat. weighted by resonance integra	Edge adjacency indices
EEig13x	eigenvalue n. 13 from augmented edge adjacency mat. weighted by bond order	Edge adjacency indices
G(N..N)	sum of geometrical distances between N..N	3D Atom Pairs
G(O..O)	sum of geometrical distances between O..O	3D Atom Pairs
G1u	1st component symmetry directional WHIM index / unweighted	WHIM descriptors
GATS5m	Geary autocorrelation of lag 5 weighted by mass	2D autocorrelations
H-047	H attached to C1(sp3)/C0(sp2)	Atom-centred fragments
H1m	H autocorrelation of lag 1 / weighted by mass	GETAWAY descriptors
H8p	H autocorrelation of lag 8 / weighted by polarizability	GETAWAY descriptors
HATSp	leverage-weighted total index / weighted by polarizability	GETAWAY descriptors
HATS2u	leverage-weighted autocorrelation of lag 2 / unweighted	GETAWAY descriptors
JGI4	mean topological charge index of order 4	2D autocorrelations
JGI7	mean topological charge index of order 7	2D autocorrelations
Jhetp	2D matrix-based descriptors	Barysz matrix weighted by polarizability (Dz(p))
MATS1m	Moran autocorrelation of lag 1 weighted by mass	2D autocorrelation

MATS4e	Moran autocorrelation of lag 4 weighted by Sanderson electronegativity	2D autocorrelation
MATS8e	Moran autocorrelation of lag 8 weighted by Sanderson electronegativity	2D autocorrelation
Mor02e	signal 02 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor08e	signal 08 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor09e	signal 09 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor20e	signal 20 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor32e	signal 32 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor13m	signal 13 / weighted by mass	3D-MoRSE descriptors
Mor18m	signal 18 / weighted by mass	3D-MoRSE descriptors
Mor26u	signal 26 / unweighted	3D-MoRSE descriptors
nBnz	number of benzene-like rings	Ring descriptors
nCconj	number of non-aromatic conjugated C(sp2)	Functional group counts
PCD	difference between multiple path count and path count	Walk and path counts
QXXm	quadrupole x-component value / weighted by mass	Geometrical descriptors
R4e+	R maximal autocorrelation of lag 4 / weighted by Sanderson electronegativity	GETAWAY descriptors
R8v	R autocorrelation of lag 8 / weighted by van der Waals volume	GETAWAY descriptors
RDF020e	Radial Distribution Function - 020 / weighted by Sanderson electronegativity	RDF descriptors
RDF035e	Radial Distribution Function - 035 / weighted by Sanderson electronegativity	RDF descriptors
RDF090e	Radial Distribution Function - 090 / weighted by Sanderson electronegativity	RDF descriptors
RDF055p	Radial Distribution Function - 055 / weighted by polarizability	RDF descriptors
S3K	3-path Kier alpha-modified shape index	Topological indices
SRW09	self-returning walk count of order 9	Walk and path counts
TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions	Molecular properties

Table A4: *All generated MLR models for the HILIC stationary phase based on Vega ZZ descriptors (n = 56)*

Number of independent variables	r^2	SE	RMSEC	RMSECV	Equation (log K_p = ...)
1	0.37	1.015	0.997	1.037	-2.31 - 1.38 Log k_{HILIC}
2	0.67	0.741	0.721	0.760	-3.40 + 0.56 Virtual log P - 1.24 Log k_{HILIC}
3	0.79	0.594	0.572	0.615	-2.54 + 0.56 Virtual log P - 0.93 Log k_{HILIC} - 0.031 Atoms
4	0.81	0.581	0.555	0.607	-2.75 + 0.49 Virtual log P - 0.93 Log k_{HILIC} - 0.029 Atoms + 0.15 Lipole (Broto)
5	0.81	0.580	0.548	0.615	-2.94 + 0.52 Virtual log P - 0.95 Log k_{HILIC} + 0.10 H-bond acceptor - 0.034 Atoms + 0.17 Lipole (Broto)
6	0.81	0.583	0.546	0.631	-2.90 - 0.0010 Melting point + 0.50 Virtual log P - 0.91 Log k_{HILIC} + 0.11 H-bond acceptor - 0.032 Atoms + 0.17 Lipole (Broto)
7	0.81	0.588	0.544	0.651	-2.96 - 0.0013 Melting point + 0.52 Virtual log P - 0.95 Log k_{HILIC} + 0.10 H-bond acceptor - 0.031 Atoms + 0.056 H-bond donor + 0.16 Lipole (Broto)

Table A5: *All generated MLR models for the NH2 stationary phase based on Vega ZZ descriptors (n = 57)*

Number of independent variables	r^2	SE	RMSEC	RMSECV	Equation (log K_p = ...)
1	0.40	0.992	0.974	1.006	-1.26 - 0.011 Melting point
2	0.66	0.754	0.734	1.023	-1.70 + 0.74 Virtual log P - 0.81 Gyration radius
3	0.68	0.732	0.706	0.760	-1.66 - 0.0039 Melting point + 0.64 Virtual log P - 0.59 Gyration radius
4	0.69	0.726	0.694	0.755	-1.65 - 0.0042 Melting point + 0.71 Virtual log P + 0.15 H-bond acceptor - 0.75 Gyration radius
5	0.71	0.714	0.676	0.757	-1.66 - 0.0040 Melting point + 0.66 Virtual log P + 0.0058 Polar area (PSA) - 0.32 H-bond donor - 0.64 Gyration radius
6	0.73	0.698	0.654	0.759	2.67 - 0.0042 Melting point + 0.43 Virtual log P - 0.031 Atoms + 0.0075 Polar area (PSA) + 0.25 Lipole (Broto) - 0.42 H-bond donor

Number of independent variables	r²	RMSEC	RMSECV	Equation (log Kₚ = ...)
7	0.73	0.703	0.652	0.773	-2.79 - 0.0042 Melting point + 0.42 Virtual log P - 0.036 Atoms + 0.0073 Polar area (PSA) + 0.25 Lipole (Broto) - 0.41 H-bond donor + 0.055 Appx. Dimensions (3)

Note: The row "7" above belongs to the preceding table section. Below begins a new section.

MLR models (Vega ZZ) incl. log K (manually added)				

Number of independent variables	r²	RMSEC	RMSECV	Equation (log K_p = ...)
1	0.05	1.223	1.262	$-2.37 - 0.47$ Log k_{NH2}
2 (NH₂ column model))	0.40	0.971	1.023	$-1.29 + 0.14$ Log $k_{NH2} - 0.011$ Melting point
2 (HILIC column model)	0.42	0.953	0.998	$-3.60 - 0.36$ Log $k_{NH2} + 0.62$ Virtual log P
3 (NH₂ column model)	0.66	0.728	0.778	$-1.63 - 0.16$ Log $k_{NH2} - 0.79$ Gyration radius $+ 0.73$ Virtual log P
3 (HILIC column model)	0.67	0.726	0.774	$-2.40 - 0.25$ Log $k_{NH2} - 0.041$ Atoms $+ 0.60$ Virtual log P
4 (NH₂ column model)	0.68	0.706	0.780	$-1.66 - 0.019$ Log $k_{NH2} - 0.59$ Gyration radius $+ 0.64$ Virtual log P $- 0.0038$ Melting point
4 (HILIC column model)	0.68	0.708	0.775	$-2.58 - 0.30$ Log $k_{NH2} - 0.039$ Atoms $+ 0.18$ Lipole (Broto) $+ 0.51$ Virtual log P
5 (theoretical descriptors model)	0.69	0.694	0.777	$-2.29 - 0.054$ Log $k_{NH2} - 0.039$ Atoms $+ 0.18$ Lipole (broto) $+ 0.43$ Virtual log P $- 0.22$ H-bond donor

5 (NH$_2$ column model)	0.69	0.693	0.776	-1.64 - 0.036 Log k$_{NH2}$ - 0.76 Gyration radius + 0.71 Virtual log P + 0.15 H-bond acceptor - 0.0041 Melting point
5 (HILIC column model)	0.69	0.700	0.774	-2.74 – 0.34 Log k$_{NH2}$ – 0.046 Atoms + 0.20 Lipole (Broto) + 0.54 Virtual log P + 0.11 H-bond acceptor
6 (theoretical descriptors model)	0.71	0.678	0.775	-2.56 - 0.033 Log k$_{NH2}$ + 0.0060 Polar area (PSA) - 0.0073 Volume + 0.23 Lipole (Broto) + 0.50 Virtual log P - 0.43 H-bond donor
6 (NH$_2$ column model)	0.72	0.669	0.768	-1.55 + 0.24 Log k$_{NH2}$ - 0.64 Gyration radius + 0.0057 Polar area (PSA) + 0.61 Virtual log P - 0.42 H-bond donor - 0.0045 Melting point
6 (HILIC column model)	0.70	0.686	0.785	-2.61 – 0.21 Log k$_{NH2}$ – 0.037 Atoms + 0.19 Lipole (Broto) + 0.51 Virtual log P + 0.12 H-bond acceptor – 0.0031 Melting point
7 (NH$_2$ column model)	0.73	0.654	0.778	-2.66 + 0.020 Log k$_{NH2}$ – 0.031 Atoms + 0.0075 Polar area + 0.25 Lipole (Broto) + 0.43 Virtual log P – 0.42 H-bond donor – 0.0042 Melting point
7 (HILIC column model)	0.71	0.677	0.792	-2.35 – 0.0038 Log k$_{NH2}$ – 0.036 Atoms + 0.19 Lipole (Broto) + 0.43 Virtual log P – 0.18 H-bond donor + 0.10 H-bond acceptor – 0.0029 Melting point
8 (NH$_2$ column model)	0.73	0.652	0.795	-2.78 + 0.0061 Log k$_{NH2}$ – 0.036 Atoms + 0.055 Appx. Dimensions (3) + 0.0073 Polar area + 0.25 Lipole (Broto) + 0.42 Virtual log P – 0.42 H-bond donor – 0.004 2Melting point

Table A6: *All generated MLR models for the phenyl stationary phase based on Vega ZZ descriptors (n = 58)*

Number of independent variables	r^2	SE	RMSEC	RMSECV	Equation (log K_p = ...)
1	0.38	1.000	0.982	1.025	-2.44 – 1.98 Log k_{phnyl}
2	0.73	0.663	0.645	0.686	-3.56 + 0.60 Virtual log P – 1.91 Log k_{phenyl}
3	0.75	0.647	0.625	0.681	-3.12 - 0.0029 Melting point + 0.55 Virtual log P – 1.59 Log k_{phenyl}
4	0.76	0.640	0.612	0.682	-2.50 + 0.58 Virtual log P - 1.40 Log k_{phenyl} - 0.16 H-bond donor - 0.29 Gyration radius
5	0.76	0.639	0.605	0.689	-2.86 + 0.48 Virtual log P – 1.31 Log k_{phenyl} – 0.016 Atoms + 0.10 Lipole (Broto) - 0.18 H-bond donor
6	0.77	0.642	0.602	0.709	-2.81 – 0.001 Melting point + 0.46 Virtual log P – 1.29 Log k_{phenyl} – 0.01 Atoms + 0.10 Lipole (Broto) - 0.15 H-bond donor
7	0.77	0.645	0.599	0.723	-2.91 – 0.0020 Melting point + 0.47 Virtual log P – 1.16 Log k_{phenyl} – 0.014 Atoms + 0.0026 Polar area (PSA) + 0.14 Lipole (Broto) - 0.24 H-bond donor

Table A7: *All generated MLR models for the BEH stationary phase based on Vega ZZ descriptors (n = 55)*

Number of independent variables	r^2	SE	RMSEC	RMSECV	Equation (log K_p = ...)
1	0.39	0.986	0.968	1.014	-2.77 - 1.70 Log k_{BEH}
2	0.65	0.748	0.727	0.772	-3.71 + 0.53 Virtual log P - 1.31 Log k_{BEH}
3	0.76	0.625	0.602	0.654	-2.86 + 0.56 Virtual log P - 0.87 Log k_{BEH} - 0.029 Atoms
4	0.78	0.607	0.579	0.672	-3.86 + 0.47 Virtual log P - 1.15 Log k_{BEH} - 0.054 Atoms + 0.16 Appx. Dimensions (1)
5	0.79	0.605	0.571	0.675	-4.11 + 0.41 Virtual log P - 1.14 Log k_{BEH} - 0.057 Atoms + 0.10 Lipole (Broto) + 0.18 Appx. Dimensions (1)
6	0.79	0.603	0.563	0.692	-3.84 + 0.39 Virtual log P - 0.0020 Melting point - 1.03 Log k_{BEH} - 0.049 Atoms + 0.11 Lipole (Broto) + 0.16 Appx. Dimensions (1)
7	0.79	0.609	0.563	0.703	-3.79 + 0.41 Virtual log P - 0.0022 Melting point - 0.99 Log k_{BEH} - 0.049 Atoms + 0.033 H-bond acceptor + 0.11 Lipole (Broto) + 0.15 Appx. Dimensions (1)