UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

Scienze del Farmaco

CICLO XXXVIII

CHROMATOGRAPHIC METHODS FOR PHYSICO-CHEMICAL EVALUATION AND PERMEABILITY PREDICTION IN EARLY DRUG DISCOVERY

Coordinatore:

Chiar.mo Prof. Fabio Sonvico

Tutore:

Chiar.mo Prof. Marco Mor

Co-tutore:

Dr. Valentina Mileo

Dr. Elisa Moretti

Dottorando:

Francesco Castagnini

Anni Accademici 2022/2023 – 2024/2025

1 Introduction

The process of developing and bringing a drug, from *Research and Development* (R&D) to the market, is complicated and involves multiple phases. Throughout this process, the characteristics of the compounds are accurately studied to gain comprehensive knowledge of the *New Chemical Entity* (NCE) discovered.

The preclinical phase plays a crucial role in this process. During this stage, all features of the compounds are meticulously evaluated, including physico-chemical properties, pharmacokinetic profiles, pharmacodynamic aspects, and toxicological profiles. This complete evaluation is essential for progressing to the subsequent clinical phases leading to commercialization at the end. This entire process aims to ensure the efficacy and safety of a new drug before it is introduced to the market.

In the preclinical phase, developing new analytical methods and new assays to improve the throughput, success rates, and the efficiency of the discovery projects, to minimize waste of resources and the attrition, which is the loss of the compound during the late stages of development, is very important and crucial.

The physico-chemical properties determine the performance of the *New Chemical Entities* (NCEs) during absorption, distribution, metabolism and excretion processes. Experimental methods remain the most widely used and reliable approach to measure the physico-chemical properties; however, nowadays many software tools are available to calculate and predict physico-chemical properties of the compounds. Despite these advances, mathematical models and software cannot fully replace the experimental measurements; as well as being more reliable, experimental data are necessary for refining predictive algorithms through iterative processes.

This project, carried out in collaboration with *Chiesi Farmaceutici S.p.A.*, is the result of studies and analysis conducted within the *Analytics and Early Formulations* (A&EF) department, under the supervision of Dr. Valentina Mileo, Dr. Elisa Moretti, Prof. Marco Mor from University of Parma. This department is located in the preclinical area, in the *Global Research and Preclinical Development* site based in Parma.

This work stems from my Master's thesis, and it is aimed at the setup and implementation in the Chiesi's screening cascade of chromatographic methods for physico-chemical evaluation in early drug discovery stage. Specifically, 19 pharmaceutical standards and hundreds Chiesi's compounds, belonging to *Rule of Five* (Ro5) and *beyond Rule of Five* (bRo5) classes, including peptides and macrocycles, were fully evaluated to assess the capability of these experimental methods to describe passive permeability better than, or as well as, the *gold standard* assays.

The main physico-chemical properties usually determined to adequately characterize NCEs and to guide chemical synthesis, which include acid dissociation constant (pK_a), lipophilicity, polarity, and propensity to form *Intramolecular Hydrogen Bonds* (IMHBs), will be described in detail below.

Each physico-chemical property investigated will be presented in a dedicated chapter, outlining the aims, materials and methods used, and the results obtained. These results include comparison with established *golden standard*.

Furthermore, lipophilicity and polarity will be analyzed to evaluate their capability to predict the formation of IMHBs using the NMR technique as "positive control assay".

A specific section of this thesis will also provide a brief description of the experimental method used to determine permeability, performed in collaboration with the Department of *Pharmacokinetics, Biochemistry and Metabolism* (PKBM) at Chiesi Farmaceutici S.p.A and will be followed by an assessment of polarity and lipophilicity as predictors of passive permeability.

The results will be presented, including comparison with the *gold standard* data, and conclusions will be drawn based on these findings.

1.1 Estimated research and development costs

Over the past four decades, the R&D area has experienced significant growth, particularly in the search for promising NCEs. As highlighted by Di Masi and colleagues in a widely cited study on drug development costs [1], the estimated expenses associated with bringing a new drug to the market have increased dramatically, reaching nearly 3 billion dollars in recent years. Specifically, the average cost to develop a single new therapeutic drug rose from 802 million dollars in 2003 to 2.56 billion dollars in 2013. This extraordinary increase, based on a real capital cost rate of 10.5% per year, as outlined in the same study, represents an approximate 150% rise over a decade.

A separate study examining new therapeutic agents approved by the U.S. *Food and Drug Administration* (FDA) between 2009 and 2018 aimed to assess the R&D expenditure needed to bring a new medicine to market. The results showed a median capitalized R&D investment of 1,787.2 million dollars, with a mean average investment estimated at 2,600 million dollars, when accounting for preclinical costs and expenses of failed trials. Additionally, this study provided median R&D cost estimates by therapeutic area (for categories with ≥ 5 drugs), ranging from $765.9 million for nervous system agents to $2771.6 million for antineoplastic and immunomodulating agents [2].

This surge in costs is driven by several factors, with the main contributors being the increased complexity and scale of clinical studies, persistent inflationary pressures that is affecting all markets, and a notable shift toward developing treatments for rare diseases, all of which are driving up R&D expenses [1]. Contrary to common belief, the shift transition from small molecules to biologics does not necessarily increase R&D costs. Although biologics involve higher component costs, they tend to exhibit a comparatively higher clinical success rate compared to small molecule [1]. The specific cost dynamics within R&D can heavily influence investment decisions. When projected R&D costs are high, investors may prefer to allocate resources to other research areas or pharmaceutical sectors. Another crucial factor contributing to high R&D costs is *attrition*, the loss of drug candidates in late development stages. Effectively reducing attrition rates has become essential, necessitating the development of new analytical methods to identify the most promising molecules. By more accurately and rapidly identifying candidates with optimal characteristics such as absorption, bioavailability, potency, and toxicity, R&D costs could be significantly reduced.

Accurate cost estimation in drug development requires comprehensive data on the expenses tied to failures, specifically R&D costs incurred for drug candidates that ultimately do not gain approval. The clinical success rate has declined, with the clinical success estimated at 11.83%, down from 21.5%. This decrease contributes to the increased costs per approved new drug, as the financial

implications of failed projects are now more pronounced. According to recent findings, only 13.8% of therapeutic agents entering Phase 1 (involving around 100 healthy volunteers), 21% of those entering Phase 2 (involving a few hundred patients with the disease), and 59% of those in Phase 3 (often involving several thousand participants) ultimately obtain FDA approval [3]. Complementing these approval rates, a study noted that the FDA approves approximately 83.2% of biologics license applications and new drug applications [2], [4]. These combined insights deepen our understanding of the challenges posed by attrition and underscore the importance of refining strategies to enhance the identification and advancement of promising drug candidates.

1.2 Finding the most promising compounds

Over the past two decades, small molecules developed as potential drugs have become increasingly complex compared to those optimized before 2000's. Recent drug candidates tend to have higher molecular weight (MW), a greater number of heavy atoms, and more intricate structures. They are generally more flexible and hydrophobic, which often results in lower solubility, a phenomenon described by Hann et al. as *molecular obesity* [5]. A notable example of this trend is represented by ampholytes. These compounds, with multiple ionizable sites, can exist in different charged states, influencing not only their physico-chemical properties but also their biological activity and biopharmaceutical performance [6], [7].

The growing complexity of drug candidates has driven medicinal chemists to adopt a complete approach, optimizing ADMET alongside target affinity [8]. This integrated strategy combines structure-based design, which enhances biological activity and potency, with property-based design, which tailors structural features to improve absorption and pharmacokinetics.

The concept of "*drug-likeness"*, introduced by Lipinski in 1997, evaluates a compound's likelihood of exhibiting favorable permeability and oral absorption based on key physico-chemical parameters. According to Lipinski's *Rule of Five* (Ro5), compounds are likely to have good oral bioavailability if they meet criteria such as *molecular weight* (MW) ≤ 500, *calculated Log P* (clog *P*) ≤ 5, *hydrogen bond donors* (HBDs) ≤ 5, and *hydrogen bond acceptors* (HBAs) ≤ 10. These guidelines have significantly influenced medicinal chemistry, especially in the fields of combinatorial chemistry and *high-throughput screening* (HTS) [9].

However, the Ro5 has its limitations. Some orally administered drugs, like cyclosporin-1, tetracycline or saquinavir are absorbed despite violating these guidelines. Both natural products and syntetic compounds, such as cyclosporine A and saquinavir also challenge these rules by acting as *chemical*

or *molecular chameleons*, altering their shape, polarity, and hydrogen bonding patterns to adapt to different environments [10], [11].

To address these challenges, molecular descriptors (e.g., Log *P*/*D* and polarity) and constitutional descriptors (e.g., counts of heteroatoms, rings and hydrogen bond donors/acceptors) have been incorporated into chemometric models. These models aim to provide actionable insights for designing active compounds.

Furthermore, an integration of the Ro5, called *beyond Rule of five* (bRo5), has been introduced, to explore a broader chemical space for orally available compounds screening. According to this rule, compounds are likely to have good oral bioavailability if they meet criteria such as MW ≤ 1000, clog *P* ≤ 7.5, HBDs ≤ 5, and HBAs ≤ 15.

While useful, these simplified parameters cannot fully capture the complexity of permeability-related behavior, often resulting in outliers and misinterpretations. However, relying solely on them risks overlooking compounds that exhibit favorable properties beyond traditional thresholds. This underscores the importance of developing new experimental analytical methods and integrating advanced physico-chemical profiling. A deeper understanding of these properties is essential for selecting promising drug candidates and optimizing their permeability early in development [12].

Recent literature reflects an evolution of drug-likeness criteria within a holistic framework. New filters include parameters such as potency, selectivity, solubility, aromatic ring count, permeability metrics (i.e. Caco-2 assays), metabolic stability (i.e. cytochrome P450 interactions), and toxicity predictions (i.e. hERG channel interactions) [13]. These refined criteria guide the selection of candidates for advanced stages of drug development.

1.2.1 The drug discovery and preclinical development process

The preclinical stage, which involves drug discovery and preclinical development, aims to narrow down thousands of compounds identified during the initial screening phase to a select few, ultimately identifying a lead candidate for clinical development studies. If these studies prove successful, the drug can eventually be clinically approved and brought to the market.

The primary strategies to discover a NCE with specific pharmacological activity include:

- Modification of existing molecules (*me-too strategy*)
- Identifying new therapeutic uses by leveraging the pharmacological activity of natural substances or the side effects of existing drugs

- Rational drug design, which involves the design of compounds based on the structure of a biological target using computational modeling
- HT screening, which systematically tests large compound libraries

During drug discovery, a significant number of compounds are synthesized and evaluated by correlating their structural properties with biological activity data, through the structure-activity relationship (SAR) studies, physico-chemical characterization and ADMET profiling. Ideally, the screening of ADMET features should occur in the early stage of drug discovery and development, alongside efficacy screening. This dual-track approach supports medicinal chemists and project teams through the hit-to-lead and lead optimization stages, with the goal of improving both the physico-chemical and ADMET profiles of the compounds under evaluation [14].
Ultimately, drug design aims to identify biologically active molecules that possess favorable *pharmacokinetic* (PK) and *pharmacodynamic* (PD) properties. Even if a compound is biologically active, it can only advance as a therapeutic candidate if it meets stringent pharmacological and developability criteria. Such compounds are then formulated for in vitro and in vivo testing, assessing their PK, PD and toxicological profiles. The formulations play a crucial role as they can significantly influence the biological properties of molecules, particularly by improving their exposure in the body. Special attention is given to physico-chemical properties that can influence bioavailability of drugs, such as solubility and permeability. Poor physico-chemical properties can complicate formulation development, making the process more challenging [15].

In the development of orally administrated drugs, achieving efficient membrane passive permeability is a key requirement, as it drives intestinal absorption. Poor intestinal permeability is often associated with low oral absorption and limited bioavailability, which can hinder the achievement of sufficient drug exposure at the site of action. Importantly, the oral administration remains the preferred route among patients due to its convenience and high compliance [14], and tablets are generally favored as a dosage form because of their ease of manufacture and handling [16].

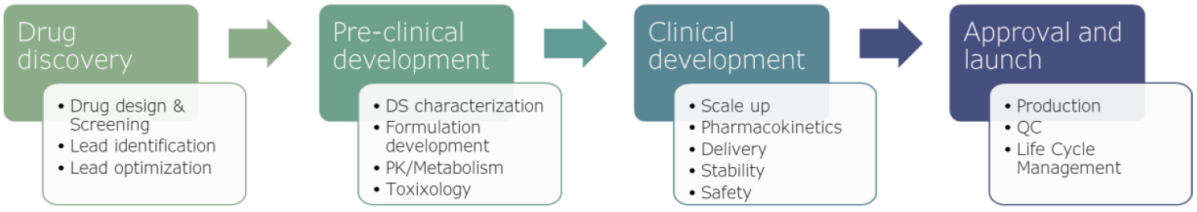


Figure 1. Flow that a candidate follows from drug discovery to the approval and launch in the market

1.3 The importance of physico-chemical characterization in early drug discovery

Mechanistically, passive membrane permeability can be described as a two-step process: first, the partitioning of a molecule between the aqueous phase and the lipid bilayer, and second, the passage across the lipid bilayer. Therefore, lipophilicity, polarity and size are widely recognized as primary determinants of passive permeability. For decades, lipophilicity has been the go-to descriptor with traditional parameters such as lipophilicity expressed as Log *P*/Log *D* being widely used. However, multiple studies have shown that the *n*-octanol/water-based approaches do not always sufficiently guide design strategies toward enhanced absorption and permeability. Other molecular features such as the number of HBDs and molecular size also play critical roles, especially when evaluating permeability using Caco-2 (human epithelial colorectal adenocarcinoma cells) permeability assay. An increased number of HBDs, a larger molecular size, and higher Log *P*/Log *D* values (particularly above 4) are all factors that tend to negatively impact passive permeability [17], consistent with Lipinski's rule described before.

Furthermore, peptide-based therapeutics gain traction in the pharmaceutical market, it has become increasingly important to monitor and understand permeability in these more complex molecules. Several *in-vitro* models have been developed and are now routinely used as surrogates for evaluating permeability. These include Caco-2 cells, Madin-Darby Canine Kidney (MDCK) cells, and artificial lipid bilayer systems like the Parallel Artificial Membrane Permeability Assay (PAMPA) and Permeapad®. Among them, PAMPA has proven particularly useful due to its ability to provide reasonable in vitro/in vivo correlation for drug-like compounds, while Permeapad® is very easy to use.

However, all these permeability assays have limitations, especially for molecules that fall into bRo5 space, such as peptides, macrocycles and PROTACs. These larger, more flexible molecules face challenges like susceptibility to peptidase degradation, non-specific binding, and complex transport mechanisms. While these assays can capture a broad range of permeation mechanisms, they are relatively labor-intensive, time-consuming, and require significant investment to achieve the throughput necessary for early-stage screening. Moreover, they often suffer from low recovery rates when used with poorly soluble, highly lipophilic compounds, such as PROTACs. Therefore, there is a clear need for high-throughput experimental approaches that can reliably evaluate the permeability of bRo5 compounds and generate meaningful data to guide drug discovery in this space [14].

Historically, molecules that violated Lipinski's Ro5 were generally considered poor candidates for oral drug development. However, as mentioned in the previous chapter many *rule-breaking*

molecules have demonstrated sufficient bioavailability and have even been approved as oral drugs. A notable example is *Cyclosporin A* cited earlier, which exhibits oral bioavailability around 30% despite having high polarity, high molecular weight, and relative low lipophilicity (Log *P* ~1) [18]. The increasing opportunity to target previously *undruggable* proteins using bRo5 molecules has sparked growing interest in understanding the molecular principles that govern their oral bioavailability.

Another relevant example is *Paritaprevir*, a compound capable of switching between high and low polarity conformations through IMHBs formation. This dynamic behavior has been directly linked to improvements in permeability, potency, solubility, and target-binding capacity [14]. This has led to a significant rise in interest among medicinal chemists toward the concept of IMHBs, as highlighted in several recent publications, many of which stem from industrial research efforts.

Numerous medicinal chemistry studies have since demonstrated the use of IMHBs formation as a strategy to achieve candidate molecules that are not only potent and permeable but also remain adequately soluble. IMHBs allow molecules to hide their polar groups when in non-polar environments like lipid membranes, enhancing lipophilicity and permeability, while exposing those same polar groups in aqueous environments, promoting solubility [19]. This so-called *molecular chameleon* behavior is highly desirable in drug design, as it enables dual optimization of aqueous solubility and membrane permeability. To successfully incorporate IMHBs strategies into drug discovery workflows, it is essential to include experimental tools capable of assessing a compound's propensity to form these internal hydrogen bonds [20]. While computational tools for predicting IMHBs formation exist, reliable high-throughput experimental validation is critical to verify this important molecular feature during screening cascades.

Due to the ever-increasing cost of R&D and the growing interest in complex molecules such as peptides, macrocycles and PROTACs, developing new high-throughput methods for physico-chemical characterization of these compounds, to understand and predict their behavior, is becoming increasingly crucial in the pharmaceutical field.

2 Acid dissociation constant (pK_a)

2.1 Why is pK_a so important in drug discovery?

Knowing the dissociation constant (K_a) of NCEs, typically expressed as the negative logarithm (pK_a), is crucial for understanding drugs behavior post-administration within the human body. In the absence of active transport mechanisms, only neutral compounds can diffuse through biological membranes. Weak acids (HA) in aqueous environments follow the dissociation equilibrium:

$$\boldsymbol{HA + H_2O \rightleftarrows H_3O^+ + A^-}$$

And it has the same meaning as:

$$\boldsymbol{HA \rightleftarrows H^+ + A^-}$$

The acid dissociation constant represents how an acid dissociates at a temperature of 25°C.

$$K_a = [H^+][A^-]/[HA]$$

$$pK_a = -\log_{10}K_a$$

For bases, the dissociation constant typically refers to the conjugate acid. Generally, the pK_a of a compound with a single ionizable group is defined as the pH at which 50% of the compound is ionized, protonated for a base or deprotonated for an acid. For acids, a lower pK_a indicates a higher K_a and, consequently, a stronger acid. In contrast, for a base, a higher pK_a of the conjugate acid corresponds to a lower K_a, indicating a stronger base. In the pharmaceutical field and in drug discovery, pK_a is used to determine the percentage of ionized molecules at a specific pH, which is essential for understanding drug absorption and distribution.

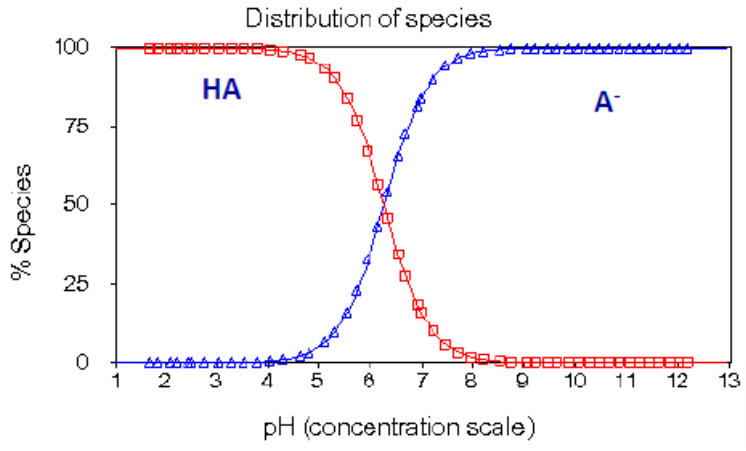


Figure 2. Dissociation profile of flumequine (weak acid); pK_a = 6.27

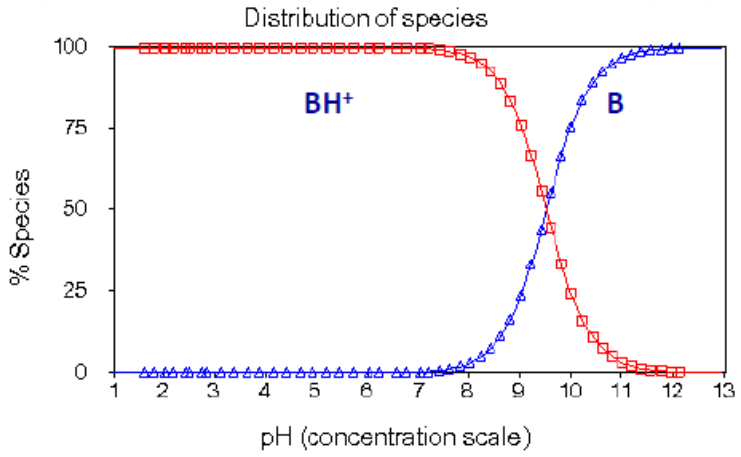


Figure 3. Dissociation profile of propranolol (base); pk_a = 9.53

At pH values approximately 1-1.5 units away from the pK_a, a fraction of both ionized and neutral species will be present in the solution. When the pH is more than 2 units above the pK_a, an acid will predominantly exist in its ionized form [A^-], while a base will be largely neutral. Conversely, when the pH is more than 2 units below the pK_a, an acid will mainly be in its non-dissociated neutral form [HA], whereas a base will exist primarily in its protonated form [BH^+] [21].

Many drugs are ionizable compounds, and their pK_a values play a crucial role in determining their ADME properties. About 75% of available drugs are bases, 20% are acids, and only 5% are neutral compounds lacking ionizable groups [22]. This key physico-chemical property, which is strongly pH-dependent, affects solubility, lipophilicity (often expressed as Log *D*), permeability, compound stability, and plasma protein binding. For instance, both aqueous solubility and membrane permeability are influenced by the molecule's charge state, which is directly affected by its pK_a. The pK_a of a drug significantly influences its bioavailability and absorption in biological environments. Moreover, multifunctional drugs often exhibit multiple ionization centers, making it necessary to understand the distinct pK_a values for each to predict which species will predominate in different compartments. In these cases, Log *D* profiles over a range of pH values provide crucial insights into the compound's absorption and distribution behavior.

2.1.1 Influence of pK_a in ADME properties

The stomach has a pH ranging from 1 to 3.5, while in the intestine the pH varies between 5.5 in the duodenum, 6-7.5 in the colon and 8 in the terminal ileum [23]; in each of these environments, acidic and basic molecules will exhibit different charges and absorption rates. In the cellular compartments the pH of interstitial fluid and of cytoplasm is 7.2, in the mitochondria the pH is close to 8 and in the lysosomes, there is a pH of 4.5-5. Furthermore, the urine has an average pH of 6 and can affect renal

clearance, cerebrospinal fluid has a pH of 7.3 and it is also important consider the pH of inflamed tissues and of tumor microenvironment, 6.5-7 and 5.5-6.5 respectively [23]. Therefore, the pH of the environment where the drug is administered is one of the most critical factors to consider in drug discovery, as it directly affects the conditions under which the compound is most bioavailable. Generally, neutral species more readily pass through lipophilic barriers such as the gastrointestinal epithelium, the blood-brain barrier, and cellular membranes. Therefore, molecules with pK_a values that allow them to remain neutral at intestinal or systemic pH (typically pH 6-8) are favored for oral bioavailability.

For example:

- Acids with $pK_a < 4$ will be mostly charged at pH 7.4, and thus poorly absorbed
- Bases with $pK_a \sim 9$ will remain mostly protonated in the stomach but become mostly non-charged in the intestine, which may enhance their absorption [24].

Ionization also significantly impacts aqueous solubility**.** While non-ionized species favor membrane permeability, ionized forms are more soluble in aqueous media**,** including gastrointestinal fluids and plasma. A well-designed drug must often strike a delicate balance between solubility and lipophilicity**,** which is heavily influenced by pK_a. This concept is frequently exploited in prodrug strategies**,** where ionizable groups are temporarily masked to enhance permeability or solubility [24]. As mentioned before, due to the different pH values in tissues and compartments (e.g., stomach pH ~2, plasma pH ~7.4, intestine pH ~8), the pK_a of a compound can lead to pH-dependent partitioning known as *ion trapping*. Basic drugs tend to accumulate in more acidic compartments (due to protonation), while acidic drugs may remain ionized and confined to extracellular fluids. Modifying the pK_a can therefore be used to modulate distribution patterns**,** including targeted delivery to inflamed or cancerous tissues with aberrant pH microenvironments [25].

Ionization state can influence binding affinity to plasma proteins such as albumin, which preferentially binds acidic drugs in their ionized forms. Similarly, interactions with enzyme active sites**,** receptors, or transporters often require specific ionization states for optimal electrostatic interactions and hydrogen bonding. Therefore, controlling a compound's pK_a enables medicinal chemists to fine-tune binding specificity and selectivity [25]**.**

Drugs excreted via the kidney undergo pH-dependent renal clearance, where ionized forms are less likely to be reabsorbed in the nephron. Designing compounds with appropriate pK_a values can promote or delay excretion according to therapeutic needs. Additionally, pK_a can influence microsomal enzyme recognition and metabolic stability**,** particularly for functional groups such as amines, carboxylic acids, and phenols [25].

In summary, pK_a is a cornerstone parameter in molecular drug design**,** impacting every phase of the ADMET process. Its influence on ionization governs drugs biological behavior, from membrane permeability and solubility to distribution and clearance. Medicinal chemists leverage this knowledge to optimize structure-property relationships, ensuring that novel compounds reach their intended targets with appropriate bioavailability and safety profiles. For this reason, a compound's pK_a is one of the primary physico-chemical properties measured in early drug discovery.

2.2 Main analytical techniques for pK_a determination

This important physico-chemical property is routinely predicted *in-silico* and confirmed experimentally by two primary methods: potentiometric titration and UV-metric titration. A number of reviews have comprehensively summarized the advantages, limitations, and applicability domains of these techniques across different chemotypes, including weak acids/bases and complex pharmaceutical compounds [26]. In potentiometric titration, the pK_a value is determined through mathematical analysis of acid-base titration curves. In contrast, in UV-metric titration, pK_a is measured by monitoring the pH-dependent changes in absorbance of chromophore groups near the ionizable site [21]. These two techniques will be described in detail in the following chapters.

2.2.1 The potentiometric technique

Potentiometry (or pH-metric) is an analytical technique based on the measurement of the potential difference of an electrochemical cell under current-free conditions before, after and during titration. This method enables indirect quantification of a substance by recording changes in electrochemical potential after adding a titrant of known concentration. Acid-base titration specifically allows us to measure the amount of acid or base in solution by determining the volume of titrant required for complete neutralization. By using the added volume of titrant, we can calculate the moles needed to achieve neutralization, which equals the moles of the analyte.

The equivalence point is reached when the moles of titrant added equal the moles of analyte in the sample. This point is identified by changes in physical properties, such as pH in acid-base titrations.

Acid dissociation constant (pK_a)

The standard setup for potentiometric titration consists of a reference electrode with a stable and known potential that is unaffected by the sample composition, an indicator electrode whose response varies with analyte concentration, and a device for measuring the potential difference. This potential can be linearly converted into pH values through calibration with standard solutions of known pH [27].

The Nernst equation allows us to calculate the electrode potential or the difference between the potentials across an electrochemical cell depending on the activity of the species involved in the redox process:

$$E=E^0+RT/nF\ln a_{ox}/a_{red}$$

Normally, in the Nernst equation, the concentrations are used instead of activities. As a result, a change in the ionic concentration causes a corresponding change in the electrode potential.

$$E=E^0\text{-}RT/nF*\ln C_{ox}/C_{red}$$

The parameters of the Nernst equation are the following:

- E: redox potential at the working temperature expressed in Kelvin (K)
- E^0: standard electrode potential
- R: universal gas constant equal to 8.314 J/(mol K)
- T: absolute temperature in Kelvin
- n: number of electrons transferred in the half-reaction by each mole of reagents
- F: Faraday constant

It is important to note that temperature should be carefully controlled during titration, as the ionization constant varies with temperature changes. pH-metric titration is considered the gold standard method due to its reliability, repeatability, and applicability to ionizable compounds with pK_a values in the pH range of 2 to 11, regardless of the presence of conjugate chromophores near ionizable groups. This method is also advantageous because it requires only a small amount of compound (~1 mg based on MW).

Due to its simplicity and low cost, potentiometric titration is one of the most widely used methods in the pharmaceutical field to determine the ionization constant of NCEs.

2.2.1.1 pH-metric pK_a determination

From the titration curve, which shows pH as a function of the volume of titrant added, we can obtain information about the ionization degree of the compound based on the pH. The final graph, known as the *Bjerrum plot* (Figure 4), represents the molecular charge as a function of the fraction of ionized molecules at different pH values during the analysis.

For monoprotic acid (HA):

$$Z = -1 / (1 + 10^{\wedge(pH - pKa)})$$

For monoprotic base (BH+):

$$Z = +1 / (1 + 10^{\wedge(pKa - pH)})$$

For polyprotic compounds:

$$Z = \Sigma \text{ (charges from all ionizable groups)}$$

The pK_a is identified where $|Z| = 0.5$ (50% ionization).

From the graph, the pK_a value can be determined, typically by optimizing it through the minimization of the sum of the squares of the residuals (S):

$$S = \sum_i^{N0} \frac{(pHiobs - pHicalc)^2}{\sigma^{2i}(pH)}$$

N_0 = number of experimentally measured pH
σ_i^2 = estimated variance for each measured pH
pHicalc = pH calculated by the model, based on the pK_a and independent variables
pHiobs = observed pH during the experiment

$$\sigma i^2(pH) = \sigma c^2 + (\sigma v * dpH/dV)^2$$

σc and $\sigma v =$ fixed contributions correlated to the measured pH and the incremental titrant volume, equal to $5*10^{-3}$ pH units and $2*10^{-4}$ mL, respectively.

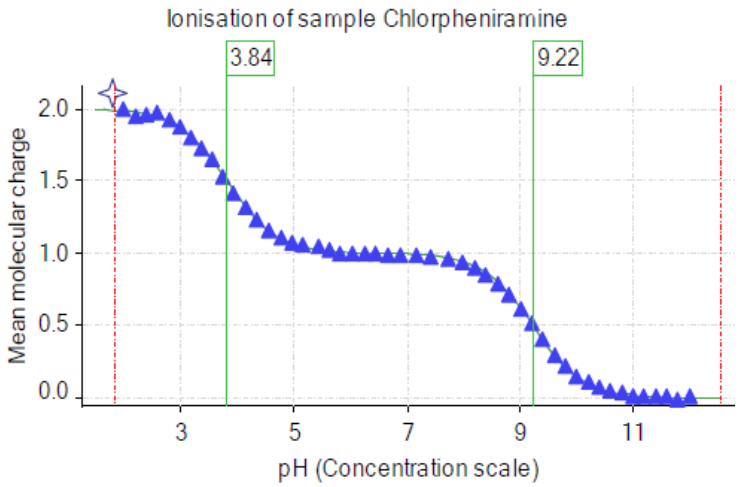


Figure 4. Example of a theoretical Bjerrum curve for Chlorpheniramine

Next, the refinement process is applied, which is an iterative calculation method used to achieve the best fit between the experimental data and the mathematical model. At the end of each iteration, the progress of the refinement is quantified by the *Goodness of Fit* (GOF), which is calculated as:

$$GOF=\sqrt{S/(N_0-Nr)}$$

N_0 = number of experimentally measured pH
N_r = number of parameters refined.

The ideal value of the Goodness of Fit (GOF) is equal to 1, which indicates that the calculated pH curve and the observed pH curve perfectly fit each other. Convergence of the non-linear equation is achieved when the change in GOF (ΔGOF) between two consecutive cycles is less than 10^{-5}, or when all refined parameters have a calculated shift of less than 10^{-4}. Typically, between 15 and 50 iterations are required to meet these criteria, and each titration is repeated at least three times. Finally, the Multiset polynomial function is used to refine all the data. During the refinement process, the *Four Plus Parameters* determined during electrode calibration must remain fixed, while the *concentration factor*, *carbonate concentration*, and *acidity error* are optimized.

2.2.1.2 Concentration factor, carbonate concentration and acidity error

The sample concentration factor can be determined during the titration by measuring the volume of titrant added to ionize/neutralize the sample. Alternatively, it can also be determined empirically.

Sample concentration factor = empiric concentration/theoretical concentration

This value should be between 0.7-1.3.

The acidic error is defined as the difference between the theoretical volume of the titrant and the empirical volume, normalized to the average volume of the titrant, and then multiplied by the molarity of the titrant. If there is a discrepancy between the empirical and theoretical volumes, it results in a deviation in the ionization curve. The range of acidity error can be broad, and it is still possible to determine the pK_a value even with a high acidity error.

The pK_a measurement can also be influenced by the concentration of carbon dioxide in the sample. When carbon dioxide dissolves in water, it forms carbonic acid:

$$CO_2+H_2O = H_2CO_3 = H^+ + HCO_3^- = 2H^+ + CO_3^{2-}$$

Carbonic acid is a weak acid, and its ionization percentage depends on the pH of the solution. Therefore, its ionization profile can interfere with that of the sample. To eliminate the effect of carbonic acid, its concentration is estimated from its ionization profile and subtracted from the results. To prevent excessive carbonic acid presence and carbonation of both the sample and the basic titrant, the system is maintained under a nitrogen atmosphere. Additionally, the carbonate concentration must be kept below 0.1 mM when measuring the aqueous pK_a, although this value can be slightly higher in the presence of a cosolvent.

2.2.1.3 pH-metric pK_a determination using cosolvents

As mentioned earlier, to measure the pK_a value using the pH-metric technique, the sample concentration should be at least 1 mM. However, poorly soluble compounds may precipitate in such cases, a cosolvent, typically methanol, is used to enhance the solubility of these compounds [28].
The pK_a measured in the presence of a cosolvent is considered an apparent pK_a (p_sK_a). The presence of a cosolvent can influence the measurement, so it is always preferable to use it only when necessary and in the smallest possible amount. Three concentrations of methanol were used for the pH-metric pK_a determination of compounds with low solubility: 35%, 45% and 55%.

The apparent pK_a (p_sK_a) should be rationalized in terms of aqueous conditions. To determine the aqueous pK_a, three or more titrations are performed using different cosolvent-water ratios. The p_sK_a

is measured at various cosolvent concentrations, and the data is plotted accordingly. Finally, an extrapolation to 0% cosolvent is performed to obtain the aqueous pK_a.

There are two methods to perform the extrapolation:

- The Yasuda-Shedlo*vs*ky method
- Linear extrapolation

The first one relates the observed value of p_sK_a to the dielectric constant ε of the mixture between water and the organic cosolvent, considering the molar concentration of the water:

$$p_sK_a + \mathrm{Log}_{10}[H_2O] = A/\varepsilon + B$$

Where A and B are the slope and intercept of the linear extrapolation, respectively.
This extrapolation strategy has been successfully applied to a variety of water-insoluble drugs, demonstrating that accurate aqueous pK_a values can be obtained from cosolvent titrations when electrode calibration and dielectric corrections are properly handled [29].
The addition of an organic solvent reduces the dielectric constant of the water/cosolvent mixture compared to pure water. This reduction weakens the strength of acidic and basic sites ($p_sK_a < pK_a$ for a base; $p_sK_a > pK_a$ for an acid). Consequently, the slope value (A) will be positive ($m > 0$) for an acid and negative ($m < 0$) for a base.

Linear extrapolation to 0% cosolvent yields the aqueous pK_a value with an accuracy of 0.2 logarithmic units. When extrapolating from high cosolvent percentages, corrections are necessary. It is advisable to work with the lowest possible cosolvent concentration, depending on the compound's solubility, as the presence of cosolvent can negatively affect electrode performance.

2.2.2 UV-metric technique

In this work the pK_a values of 18 pharmaceutical standards were determined using the pH-metric technique, while one of them was analyzed by the UV-metric technique. When one or more chromophore groups are located near an ionizable group, UV/Vis spectroscopy can be used to determine pK_a values. To use this technique having the ionizable group close to the chromophore ($n < 3$ bonds distance) is mandatory. This methodology has been developed to improve sensitivity, to decrease the amount of sample needed to perform the analysis and in general to broaden the range of analyzable molecules for pK_a determination.

Acid dissociation constant (pK_a)

During an acid-base titration, changes in pH can lead to distinct absorbance spectra for the ionized and neutral forms of the compound [21]. Changes in ionization should result in a measurable difference in the molar extinction coefficient of the chromophore in the UV-Vis range.
In the SiriusT3™ instrument, used for the experimental pK_a determinations in this Ph.D. project, a deuterium lamp serves as the radiation source, and an optical fiber with a dip probe is connected to a diode array detector. After each addition of titrant and at every pH measurement point during titration, the absorbance spectrum is recorded across a wavelength range of 200 to 700 nm, encompassing the full UV-Vis spectrum.

2.2.3 The automatic titrator SiriusT3™

There are numerous automatic titrators available on the market. Among them, there is the one proposed by Sirius Analytical Instrument Ltd, the SiriusT3™, which is specifically designed to determine key physico-chemical properties in drug discovery for NCEs with ionizable groups, including pK_a, lipophilicity and solubility.
The instrument is composed of three main modules (Figure 5):

- Dispenser module: this module includes micro dispensers for ISA water, acid titrant (HCl 0.5 N) and basic titrant (KOH 0.5 N), cosolvent (methanol) and partition solvent (*n*-octanol) . It is also featuring a UV-Vis spectrophotometer with an optical fiber for measuring UV-Vis spectra in solution;
- Titration module: this module features a mobile mechanic arm equipped with multiple probes, including a combined glass electrode (consisting of a constant potential Ag/AgCl reference electrode, and a glass electrode for H^+ ion activity and pH measurements), an immersion UV probe for UV-metric measurements, an agitator, a temperature sensor and a series of quartz micro capillaries for reagents addition. These probes are miniaturized to operate with sample volumes as small as 1mL. Along the path of the mobile arm, there are vials containing buffer solution at pH 7 for electrode conditioning and calibration, as well as washing solvents (IPA and surfactants) for cleaning the probes after analysis.

The reference electrode typically consists of a glass tube encased by an outer containing a silver wire in contact with its sparingly soluble salt (AgCl), immersed in a KCl solution with constant activity. The potential of this Ag/AgCl electrode depends on the Cl^- activity in the solution. By maintaining a constant Cl^- concentration, the electrode potential remains stable. Inside the inner tube, there is also a silver wire, coated with AgCl and surrounded by KCl and a buffer solution; this inner component is the electrode used for measurements. It contacts the sample solution through a glass membrane at

its tip, while the outer tube contacts the external solution via a porous diaphragm that functions as a salt bridge.

The measurement electrode is a glass electrode, with its pH-sensitive component being the glass membrane itself. Ion exchange occurs at the membrane's surface layers, between the Na^+ ions in the glass and H^+ ions in the solution. Each silicon atom in the glass is bonded to four oxygen atoms. The numerous silicon negative charges are counterbalanced by mobile monovalent cations (Na^+ and Li^+) within the 3D structure, which are responsible for electrical conduction. Before functioning as pH sensor, both surfaces of the glass membrane must be hydrated. The glass hydration involves an ion exchange reaction between the monovalent cations and the protons in the solution:

$$H^+ + Na^+SiO^- \leftrightarrow Na^+ + H^+SiO^-$$

The electric potential measured by the electrode is due to the imbalance between the H^+ ions present on the internal and external surface layers of the membrane. The H^+ ions in the solution influence and modify the hydrated silicon layer of the membrane to an extent that depends on their activity in the solution. The charge variation on the external side is transmitted to the inner side by the Na^+ or Li^+ ions in the glass. This potential is sent to the pH sensor through a coaxial cable, where it is amplified and displayed as corresponding pH units.

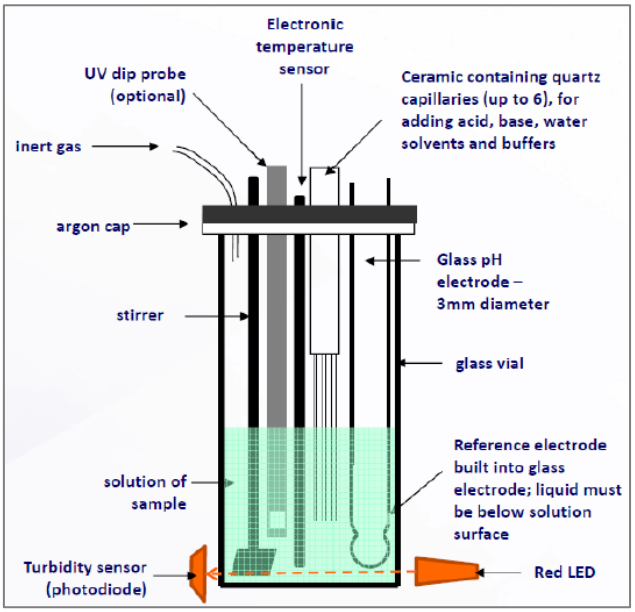


Figure 5. SiriusT3™ analytical cell components. Image taken from "SiriusT3™ training course: pK_a theory and definitions" slide n. 6 (2011)

3 Lipophilicity

3.1 Importance of lipophilicity in early drug discovery

Lipophilicity is defined as the affinity of a molecule for a lipid phase, measured by its extent of distribution in a biphasic system normally composed by *n*-octanol and aqueous buffer solution [30]. A higher partition coefficient often indicates greater lipophilicity, which can enhance a drug's ability to cross cell membranes and reach its target site. It is integral to *quantitative structure activity relationship* (QSAR) studies, which correlate chemical structure with biological activity. This relationship helps in predicting the efficacy and safety of new compounds based on their partitioning behavior. This knowledge aids in the formulation of drugs, influencing how they are delivered and absorbed in the body. For instance, drugs with optimal partition coefficient are more likely to be effectively absorbed in gastrointestinal tract. Partition coefficient also plays a role in assessing the environmental impact of pharmaceuticals. Compounds with high lipophilicity may accumulate in biological systems and ecosystems, raising concerns about toxicity and bioaccumulation [31].

It is commonly expressed as the logarithm of the partition coefficient in *n*-octanol/water (Log $P_{oct/wat}$) [32], and it remains one of the most fundamental molecular properties influencing drug discovery. *n*-octanol has been widely adopted as the reference solvent for partition coefficient measurements due to its effectiveness as a good biological membrane model. It affects the ADMET profile of a drug, and it is thus tightly linked to pharmacokinetics and safety profiles [33], [34]. It also influences off-target effects and toxicity, particularly for central nervous system (CNS) drugs and highly lipophilic structures [25].

3.1.1 The difference between Log P and Log D

Log *P* is defined as the lipophilicity of neutral compounds, while for the molecules having ionizable groups in their structure, at certain pH, the Log $D_{oct,pH}$ is normally used. It is important to differentiate between these values, as they provide different information about the molecule's behavior in different environments. For ionizable compound at physiological pH, the Log *D* value is generally considered more important and relevant than Log *P*, especially for the pharmacokinetic, because Log *D* considers the molecule's degree of ionization based on its pK_as [35].

$$P = \frac{[\text{neutral}]_{\text{octanol}}}{[\text{neutral}]_{\text{water}}}$$

$$D = \frac{[\text{ionized + neutral}]_{\text{octanol}}}{[\text{ionized + neutral}]_{\text{water}}}$$

Consequently, Log *D* provides a more accurate representation of the compound's behavior under physiological conditions, marking it a valuable parameter in pharmaceutical research. The relationship to ionization is as follows:

- for non-ionizable compounds, Log *D* is equal to Log *P* across all pH values
- For acidic compounds, they exhibit identical Log *D* and Log *P* values at low pH
- For basic compounds they show equal Log *D* and Log *P* values at high pH
- For multiprotic compounds there are more complex equilibria that are described by more complex equations.

Almost all the commercial drugs contain one or more ionizable groups in their structure and so at physiological pH they are often ionized, making their lipophilicity pH dependent [22]. This necessitates careful consideration of ionization when predicting their properties.
For monoprotic acids and bases, the Log *D* value can be estimated using the Log *P* value and pK_a, assuming that the neutral form is the predominant one in the partition. This estimation provides a valuable tool for predicting lipophilicity under various pH conditions.

$$\text{Log } D \sim \text{Log } P - \text{Log}(1+10^{\text{pH-pKa}}) \qquad \text{for a monoprotic acid}$$

$$\text{Log } D \sim \text{Log } P - \text{Log}(1+10^{\text{pKa-pH}}) \qquad \text{for a monoprotic base}$$

Understanding the relationship between Log *P* and Log *D* is crucial in pharmaceutical research and development. It allows for more accurate predictions of drug behavior in different physiological compartments, each with its unique pH environment. This knowledge is essential for optimizing drug formulations, predicting bioavailability, and understanding drug distribution within the body.

3.1.2 How does ionization affect the lipophilicity of drugs?

Ionization significantly affects the lipophilicity of drugs. As said before, the lipophilicity of a drug is often pH-dependent, as many compounds contain ionizable groups. The degree of ionization can change with pH, with in turn affects the distribution of the drug between lipophilic (non-polar) and hydrophilic (polar) environments. For instance, at lower pH values, acid groups may be non-ionized while basic groups may be protonated, altering the molecule's overall lipophilicity [36].

The presence of these groups means that lipophilicity can vary widely depending on the ionization state of the molecule at a given pH [22].
The lipophilicity profile of a drug, which plots Log *D* against pH, illustrates how the lipophilicity changes as the drug transitions between ionized and non-ionized states. This profile is crucial for understanding how the drug will behave in biological systems, particularly regarding absorption and membrane permeability [36].

Some drugs may exhibit zwitterionic characteristics, where they possess both positive and negative charges (such as ammonium group and carboxylate group) within the same structure, resulting in a net neutral charge, which can influence their solubility and interaction with biological membranes. These ampholytes can complicate the study of lipophilicity and pharmacokinetics, as their behavior in lipophilic environments can be less predictable. Measuring the lipophilicity of zwitterionic compounds can be challenging. Traditional methods may not accurately capture the behavior of these molecules due to their unique ionization patterns, needing the development of specialized techniques to access their lipophilic profiles effectively. The zwitterionic nature of a drug can influence ADMET properties. Zwitterion may have decreased solubility in aqueous environments due to their neutral net-charge, which can facilitate absorption in the gastrointestinal tract. However, their ability to cross lipid membranes may be reduced compared to purely non-ionized lipophilic compounds. Understanding the balance between these properties is crucial for drug design [36].

3.2 How lipophilicity influences ADME properties

From a pharmacokinetic perspective, lipophilicity directly affects a compound's ability to permeate cellular membranes, influencing its absorption, distribution, metabolism, and excretion. High lipophilicity may enhance membrane permeability, tissue distribution, plasma protein binding, and affinity for CYP450 enzymes, thus increasing clearance and potentially toxicity. However, excessive lipophilicity can compromise metabolic stability. Conversely, compounds with low lipophilicity often show poor membrane permeability and limited bioavailability [37].

Nevertheless, measuring and interpreting lipophilicity remains challenging, especially for ionizable, poorly soluble, or ampholytic/amphiphilic compounds. Traditional partition systems may not fully capture dynamic biological interactions. This has prompted the adoption of alternative methods offering better high-throughput capabilities and biological relevance [37], [38].
Lipophilicity is now viewed not just as a bulk solvent partition but as a dynamic interaction at biological interfaces such as membranes and proteins [39].

Lipophilicity

During lead optimization, compounds are modified to enhance potency, selectivity, and pharmacokinetics while minimizing toxicity. Here, fine-tuning lipophilicity and pK_a are key factors. Highly lipophilic molecules often show poor solubility and increased toxicity, while overly hydrophilic compounds may suffer from poor permeability and oral bioavailability [25].
Structural adjustments, such as changing alkyl chains or substituents, are used to modulate Log *P* and pK_a, with computational tools like Abraham descriptors supporting rational design [40].

Additionally, the difference in partition coefficients between organic solvents (ΔLog *P*), measured via shake-flask or potentiometric methods, provides insight into solute properties like hydrogen bond donor acidity. However, each assay captures different aspects of lipophilicity and passive permeability. For example, liposome/water systems better reflect biological behavior than octanol/water systems and seem to be more aligned with Caco-2 permeability because liposomes better mimic the composition and complexity of biological membranes, including both hydrophobic and electrostatic interactions. Meanwhile, ΔLog *P* seems to correlate well with skin and BBB permeation. Importantly, the predictive value of lipophilicity depends on how well the test system models biological membranes [32]. That said, some systems, like liposome/water, lack reproducibility and are not suitable for high-throughput screening.

Despite advances in drug delivery, oral administration remains the most preferred route due to patient compliance. For oral drugs, membrane permeability is essential for absorption. This process can occur via passive diffusion, carrier-mediated transport, or active transport requiring energy (e.g., ATP). Efflux mechanisms, like those involving P-glycoprotein (P-gp), serve as a defense against xenobiotics. If a compound is a P-gp substrate, it may show low intracellular accumulation and reduced efficacy or bioavailability. Passive diffusion is considered the main contributor to absorption and is governed by molecular properties like size, polarity, and lipophilicity. These characteristics are encapsulated by the *Rule of Five*/*beyond Rule of Five* framework [12].

Data from Caco-2 and MDCK cell assays show a bi-linear relationship between lipophilicity and permeability: permeability rises with hydrophobicity up to an optimal point, after which it drops due to poor solubility and membrane trapping. Thus, it is important to define the optimal lipophilicity range for maximal transport. This discovery led to the formulation of the Wellcome/GSK "4-box" model, which combines Log $D_{7.4}$ with "cmr" (calculated molar refraction), an estimate of three-dimensional molecular size more accurate than simple molecular mass, to stratify compounds into quadrants predictive of oral bioavailability. The linear discriminant function "ldf" derived from this

two-dimensional space predicts oral absorption more accurately than the PAMPA or Caco-2 cell assays [37], [41], [42].

Lipophilicity also predicts protein binding, especially to human serum albumin (HSA). Higher lipophilicity (and acidic features) values correlate with stronger binding to HAS columns, reducing free drug levels and potentially efficacy. Plasma protein binding increases with Log $D_{7.4}$ and aromaticity so incorporating aromaticity further enhances predictive power. The distribution volume (Vd) is strongly influenced by tissue binding, experimentally surrogated by binding to immobilized phosphatidylcholine columns (IAM). Lipophilic compounds (and with basic features) show high Log K_wIAM values, indicating accumulation in membrane-rich tissues [37], [43], [42]

Regarding metabolism, high lipophilicity values are linked to greater affinity for CYP450 enzymes (2D6, 2C9, 2C19, 3A4), leading to faster turnover and affecting clearance and drug-drug interactions. Both very hydrophobic and very hydrophilic compounds may show metabolic liabilities, as indicated by bi-linear trends between lipophilicity and enzyme inhibition profiles, but in general lipophilic compounds more easily penetrate the microsomes and the catalytic pockets of P450s prefer hydrophobic substrates. A Property Forecast Index (PFI = Chrom Log $D_{7.4}$ + number of aromatic rings) > 7 seems to increase P450-mediated hepatic metabolism [37].

Compounds with PFI > 7 were over 50% more likely to inhibit 5 or more proteins with pIC50 ≥ 5. Similarly, hERG risk, QT-dependent cardiotoxicity, increased with intrinsic lipophilic (Log *P*) and positive charge, with critical cut-off at PFI > 7 [37], [41].

In summary, the integration of lipophilicity into ADMET profiling provides comprehensive and operationally accessible metrics, such as the Ligand Lipophilicity Efficiency (LLE = pIC50 - Log $D_{7.4}$), the PFI or the AB-MPS (AbbVie Multiparameter Score = |CLog $D_{7.4}$ - 3| + Ar + NRB) for assessing drug-likeness and bioavailability also for bRo5. Its strong correlations with key developability parameters make it an invaluable tool in modern drug design, particularly when combined with structural descriptors such as aromatic ring count [37], [42], [44].

3.3 The gold standard techniques to measure lipophilicity

There are a lot of methods available to measure this important physico-chemical property, and to correlate lipophilicity with passive permeability is almost always the principal aim. The shake-flask

and pH-metric assay are the *gold standard* but, especially during the last years, other innovative methods are becoming popular in the analytical field.

3.3.1 Shake-Flask assay

The *shake-flask* assay is considered as the *gold standard* technique for lipophilicity measurements. A sample is dissolved in a biphasic system normally composed by *n*-octanol and aqueous buffer solution using MOPS (3-(N-Morpholino)propanesulfonic acid sodium salt) for pH 7.4 and CAPS (3-(cyclohexylamino)-1-Propanesulfonic acid sodium salt) for pH 9.3 with controlled ionic strength (usually KCl 0.15 M). The system is stirred to promote and to achieve the equilibrium between the concentrations of the compound in each phase [36].

As said before, the partition coefficient is expressed as Log *P* for the neutral form, and depending on the pH of the aqueous buffer solution it is possible to calculate the Log *D* of the compound measuring the concentration of it in each phase. This allows us to obtain a complete lipophilicity profile even if the pK_a is unknown. To measure Log *P* or Log *D* more than one volumetric ratio should be used. Normally, the assay is performed using 3 different volumetric ratios (i.e. 1:1, 1:2 and 1:4) to verify that the lipophilicity value doesn't change based on the environment. Furthermore, because of the expected lipophilicity of the compound, it is mandatory to invert the ratio (1:1, 2:1 and 4:1) especially if the lipophilicity is very low, as for peptides.

Due to its simplicity and for its capacity to give us a direct measure of the concentration of the compound in each phase, the *shake-flask* technique is considered a reference method against which other methods are validated. Unfortunately, there are some experimental limits to this technique. Normally, a high amount of compound is needed to perform the assay (compared to the other techniques) and, in general, it takes a lot of time because of reaching equilibrium of the phases. Precision can be affected by factors such as incomplete phase separation or the presence of impurities. Furthermore, amphiphilic molecules can't be analyzed due to their interface accumulation between the biphasic system. For highly lipophilic and poorly water-soluble compounds, classical *shake-flask* measurements can be experimentally challenging due to low aqueous concentrations, emulsions, and long equilibration times. In these cases, HPLC-based Log *P* determination using gradient or isocratic reversed-phase methods have been shown to provide reliable lipophilicity data over extended Log *P* ranges, including very hydrophobic aromatic systems [45]. Chromatographic methods will be discussed in the next section.

It provides high accuracy and reproducibility for neutral and moderately soluble compounds. To overcome limitations in scalability and solubility, miniaturized and automated variants have been developed. Additionally, partitioning systems using alternative solvents such as alkanes or cyclohexane have been proposed to better model lipophilic environments like cell membranes or the blood-brain barrier [40] but with more solubility issues.

3.3.2 Liposome-water biphasic system

Traditionally, lipophilicity has been evaluated using the *n*-octanol/water partitioning system, which estimates the distribution of a compound between a hydrophobic organic phase and an aqueous buffer. However, this system has limitations in mimicking the structural and electrostatic complexity of biological membranes. To address this, alternative models such as the liposome-water biphasic system have been developed, offering a more physiologically relevant representation of the drug-membrane interaction.

In this system, the lipid phase is composed of small uni-lamellar vesicles made from dimyristoylphosphatidylcholine (DMPC), a phospholipid that closely resembles those found in cell membranes. The aqueous phase consists of a buffered solution, and partitioning experiments are performed at controlled pH values to study the distribution of ionizable compounds in both their neutral and charged forms. The distribution coefficient between the liposome and aqueous phases ($\mathrm{Log}\, D_{\mathrm{mem}}$) is then determined, typically using ultrafiltration followed by HPLC analysis of the free (unbound) drug.

One of the main advantages of the liposome/water system is its ability to reveal membrane interactions of charged species, something that the *n*-octanol system often fails to capture. In fact, studies have shown that ionized compounds, particularly protonated amines, can still partition significantly into the phospholipid bilayer. This is contrary to the conventional assumption that only neutral forms of drugs can interact with hydrophobic environments. Notably, the observed partitioning of charged amines does not appear to be due to ion pairing, as increasing the ionic strength of the buffer had little effect on their membrane affinity. Instead, it is likely driven by electrostatic interactions with the phospholipid headgroups or by partial insertion of the molecule into the bilayer, where the hydrophobic regions are embedded, and the charged moiety remains solvated near the membrane interface.

These findings suggest that the membrane affinity of ionizable compounds cannot be predicted solely by *n*-octanol/water partition coefficients. For example, amlodipine, a calcium channel blocker with a

protonated amine group, exhibits high log D_{mem} values even at physiological pH, despite relatively moderate *n*-octanol/water partitioning. This behavior may contribute to its pharmacodynamic profile, including prolonged onset and duration of action, by enhancing membrane retention and slow release. Overall, the liposome/water system provides a more nuanced and mechanistically relevant insight into the lipophilicity of drug candidates, particularly those that are ionizable. By accounting for membrane-specific interactions, this model aids in the optimization of compounds for improved absorption, distribution, and target engagement in vivo [46].

3.3.3 Potentiometric technique

Another important technique to measure the lipophilicity of compounds is potentiometric analysis (also called pH-metric). This technique uses a single acid-base titration, always using a biphasic system composed by *n*-octanol and aqueous buffer solution with controlled ionic strength (KCl 0.15 M). Also in this technique, the phases have saturated each other before the experiment (mutual saturation of the phases). The partition of the compound in *n*-octanol can modify the equilibrium between the phases.

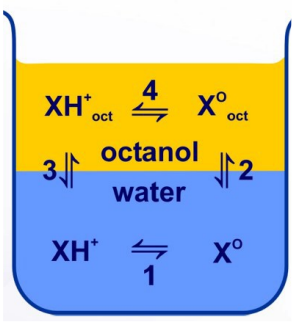


Figure 6. Partition equilibrium between two phases, n-octanol/water. The equilibrium n.1 represents the aqueous pK_a, the 2nd one represents the lipophilicity of the neutral specie (Log P), the 3rd one represents the lipophilicity of the ionized form (Log P of ion) and the 4th one represents the pK_a in n-octanol. Image taken from "SiriusT3™ training course: Log P measurement" slide n. 5 (2011)

To know the aqueous pK_a value of the compound we want to analyze is mandatory, so it should be determined. This assay is performed using an electrode and it is important to calibrate it before starting the measurements. The technique is based on the principle that the ionization state of compounds affects its solubility in aqueous phases. By titrating the compound in a two-phase system, the changes in pH can be monitored to determine how the compound partitions between the phases.
In a typical pH-metric titration, the neutral species will tend to distribute in the *n*-octanol phase, causing a shift in the inflection of the titration curve (apparent pK_a o p_oK_a if *n*-octanol is used) if it is referred to an aqueous pK_a.

The ionization equilibrium will become weaker and so the pK_a for an acid compound shift to higher values, while the pK_a for a base shift to lower values, and these shifts will be greater the more the neutral species go into *n*-octanol phase. As for the shake-flask assay, different volumetric ratios should be used, based on the expected lipophilicity of the compound. For compounds with high lipophilicity, a low *n*-octanol volume should be used, on the contrary for hydrophilic compounds a higher *n*-octanol volume should be used [36].

The method can be automated, allowing for the analysis of multiple samples in a relatively short time. It is particularly useful for studying the lipophilicity of ionizable compounds, which are common in drug development. Factors such as temperature, ionic strength, and the presence of other solutes can affect the results and must be carefully controlled.

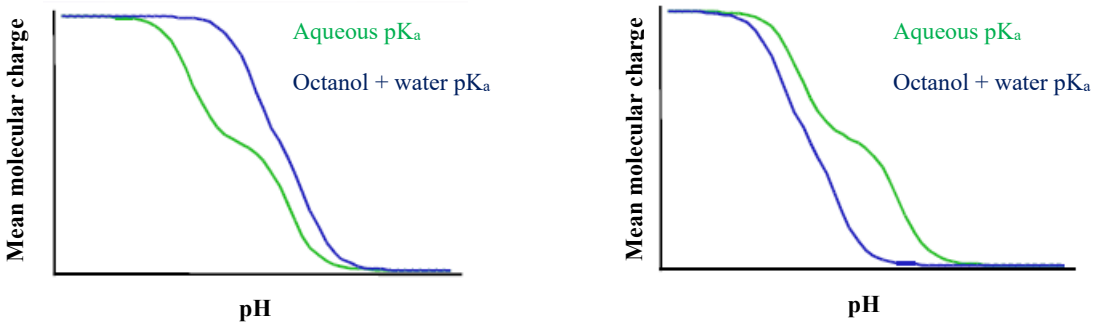


Figure 7. Ionization profiles representing the aqueous pK_a (green line) and the apparent pK_a with n-octanol (blue line), for an acid (left plot) and for a base (right plot)

3.3.4 Chromatographic technique

All the methods explained above have some critical points and sometimes are tedious. Using the *shake-flask* to measure the lipophilicity of amphiphilic molecules could be challenging because they tend to accumulate in the interface of biphasic system due the hydrophobic and hydrophilic moieties. Furthermore, if you do not have a miniaturized assay the amount of compound you need is high. Finally, this method takes a lot of time in terms of throughput. The same as for the pH-metric methods, where the solubility is an important parameter and, in this case, it is not possible measure the lipophilicity of amphiphilic compounds due emulsion formation after strong stirring. On the contrary chromatographic methods, are faster, more reproducible, require less amount of compound to perform the analysis and amphiphilic molecules can also be analyzed without any issue.

Particularly, reversed-phase high performance liquid chromatography (RP-HPLC) is increasingly used to measure lipophilicity. These methods separate compounds based on their interaction with a stationary phase and a mobile phase. Chromatographic methods have many advantages, such as the

low amount of compound needed to perform the analysis, the speed of the analysis, the precision and automation [39].

Normally, for RP-HPLC, the stationary phase can be composed of a non-polar long carbon chain (C8 or C18) or phospholipids as IAM columns, while the mobile phase is typically a polar solvent (water with organic modifier like acetonitrile). Compounds are separated based on their hydrophobicity; more lipophilic compounds elute later than less lipophilic ones. The Log *P* can be estimated from the retention time using calibration curves established with known standards.

With chromatography it is possible to analyze multiple samples quickly, making it suitable for high throughput screening. The method offers high precision and reproducibility, especially when automated systems are used. It can be adapted to measure a wide range of compounds, including those that may not partition well using the shake-flask method.

This technique is used to indirectly measure lipophilicity profiles and can handle a variety of solvents and pH ranges [36].

3.3.4.1 The Chromatographic Hydrophobicity Index (CHI)

To address the limitations of traditional partitioning methods, the *Chromatographic Hydrophobicity Index* (CHI) has emerged as a robust and high-throughput alternative. The impact of lipophilicity on ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profiles is profound and multifaceted. In this context, CHI serves not just as a surrogate for Log *P*/*D* but as a central descriptor capable of explaining and predicting a range of pharmacokinetic behaviors [37].

Derived from RP-HPLC, CHI leverages the correlation between a molecule's retention time and its hydrophobicity. Specifically, the CHI value reflects the percentage of organic modifier (usually acetonitrile) required to elute with Log k = 1. A set of standard compounds with known CHI values, ranging from low to high hydrophobicity, is injected, and their retention times are used to build a calibration curve. These standard compounds (e.g., theophylline, benzimidazole, indole, propiophenone, butyrophenone) cover a broad range of hydrophobicity and allow the establishment of a linear relationship between retention time and CHI value.

For each test compound, the percentage of organic solvent required to achieve a retention factor (k) of 0 is calculated. This percentage corresponds to the CHI value. The retention factor k = 0 represents a retention time double that of an unretained marker, ensuring consistency across runs. This correlates to equal distribution between stationary, a C18 column, and mobile phases, aqueous buffer solution composed of 50mM ammonium acetate at pH 7.4, during a gradient HPLC run and is directly influenced by the compound’s hydrophobicity [34], [38]. To express CHI on a scale comparable to

traditional Log $D_{7.4}$, Valkó et al. developed linear conversion formulas. The most widely used equation is:

$$\text{CHI Log } D = 0.054*\text{CHI} - 1.467$$

This relationship was derived from experimental data on 98 compounds with both measured CHI and *shake-flask* Log $D_{7.4}$ values.

A critical insight from Valkó's work is that CHI measured with acetonitrile as the organic modifier minimizes interference from hydrogen bond donor groups. Acetonitrile is a weak hydrogen bond acceptor compared to solvents like *n*-octanol or methanol. This characteristic makes CHI measurements more reflective of hydrophobic interactions alone, without the confounding stabilization seen in *n*-octanol/water systems where *n*-octanol can act as an H-bond acceptor. To further refine lipophilicity estimates, the hydrogen bond donor acidity ($\Sigma\alpha^2H$) of the compound can be incorporated into predictive equations when necessary [34], [39].

Another chromatographic value is used, the Chrom Log D, derived from a much larger data set (>40,000 compounds) where CHI is aligned to calculated (non-experimental) Log D, and the correlation is extended over a broader base of compounds, exploiting Log D from fragment or QSAR calculations. Chrom Log D is thus also a chromatographic scale, but calibrated (rescaled) to optimize predictivity and readability against calculated values and modern pharmaceutical screening experience [34], [39].

$$\text{Chrom Log } D = 0.0857*\text{CHI} - 2$$

The CHI methodology is flexible and scalable, it enables rapid profiling of compound hydrophobicity, up to hundreds of compounds per day, with minimal sample requirements, making it ideal for combinatorial libraries and high-throughput environments. In terms of solubility, CHI values provide superior resolution compared to traditional *n*-octanol-water derived Log *D* values, especially in the low-solubility regime. Poorly soluble molecules often yield unreliable Log *D* values in shake-flask assays due to phase separation and incomplete dissolution [37].

Importantly, CHI values have been demonstrated to correlate well with Log *P* across diverse compound sets [38]. This correspondence provides an empirical chromatographic measure of lipophilicity that is operationally simpler and often more reliable than shake-flask methods,

particularly for molecules that are difficult to analyze using conventional techniques [37]. Another important aspect is the operating pH ranges, from 2 to 10, providing a dynamic view of lipophilicity [33]. By measuring retention times at pH 2.0, 7.4, and 10.5, it becomes possible to estimate the behavior of both ionized and unionized species, enhancing the pharmacokinetic relevance of the data. In practice, CHI-derived Log *D* values offer a better resolution in correlation with solubility, permeability, and cytochrome P450 inhibition [37].
Its dynamic interface-based partitioning could better mimics tissue and membrane distribution and surface interaction than *n*-octanol/water partition system [39].

Recent large-scale studies, such as those conducted at GlaxoSmithKline, have further validated the utility of CHI-based measurements. In datasets encompassing over 100,000 compounds, CHI values consistently outperformed traditional Log *P/D* predictions in modeling drug-like properties [37].
By enabling a more nuanced and operationally efficient measurement of lipophilicity, the CHI framework not only improves predictive modeling but also aligns with contemporary needs in pharmaceutical research where speed, reliability, and multidimensional profiling are paramount [37], [38]. This approach could address one major cause of late-stage drug development failure: poor in vivo behavior due to unintended tissue binding and toxicity, often driven by excessive lipophilicity. Biomimetic retention measurements offer insight into these liabilities and enable better compound selection without the need for extensive animal testing [34].
Advanced models derived from biomimetic chromatography combine CHI values with solvation parameters (e.g. hydrogen bond acidity, polarizability) to further enhance predictivity of Log P_{oct} and biological partitioning [33].

In summary, CHI represents a shift from classical, oversimplified models of lipophilicity toward dynamic, realistic, and predictive models of compound behavior. These chromatographic methods hold the promise of optimizing drug candidates earlier, reducing attrition and aligning screening tools more closely with the complexities of in vivo pharmacokinetics. CHI circumvents these limitations by offering a chromatographic method that is less sensitive to solubility artifacts, thus better capturing the compound's true behavior in biological systems [37].
Overall, CHI Log $D_{7.4}$ serves as a valuable metric in modern ADMET profiling, particularly for challenging chemical space such as bRo5 compounds.

4 Polarity

4.1 Polar Surface Area (PSA)

Biological and physico-chemical processes are governed by surface molecular interaction and due to this the importance of the calculation of these surface properties has always been one of the key factors to understand the mechanism interaction between drug and target in computational medicinal chemistry. In the pursuit of efficient and cost-effective drug discovery, computational prediction of pharmacokinetic properties has become indispensable. Among the molecular descriptors used to estimate absorption and bioavailability, *Polar Surface Area* (PSA) stands out as a particularly robust and informative metric.

During the last 20-30 years, PSA has become very important in the pharmaceutical and in the computational chemistry field.

It is used to measure the polarity of compounds during the drug discovery phase and PSA can be calculated using fragmental approaches starting from the structure of the molecule or using molecular dynamic techniques and computational analysis, in which area associated to heteroatoms like oxygen (O), nitrogen (N), phosphorus (P), sulfur (S) and the hydrogen binding to them are considered as surface polar moieties of the compounds [47].

It is demonstrated that low PSA values are correlated to better pharmacokinetics properties, increasing the intestinal absorption and passive permeability through biological membrane and BBB as well [48]. Furthermore, the PSA influences solubility, molecular geometry, and the conformation of the structure [49].

The relevance of this measurement has increased during the last years, and it is still important for the prediction of the intestinal and BBB permeability behavior in the vitro model like Caco-2 or MDCK cellular monolayers [50]. It seems PSA is a better permeability and absorption descriptor than the calculated Log *D* or Log *P* values [51]. For example, compounds with a PSA value over 140 A^2 tend to have lower permeability than compounds with a PSA value under 60-90 A^2 [51].

The polar surface area is very sensitive to the structure of compounds, especially to the molecular conformation. Due to this, it is possible to calculate the PSA considering the contribution of each conformer weighing the average based on their abundance. However, it can take a lot of time.

In summary, PSA offers a rapid, cost-effective, and reliable computational method for predicting passive drug absorption and *Central Nervous System* (CNS) penetration**,** helping to streamline early-stage drug design. It should be considered a core descriptor in virtual screening and lead optimization,

particularly when used in conjunction with other properties such as lipophilicity and molecular weight.

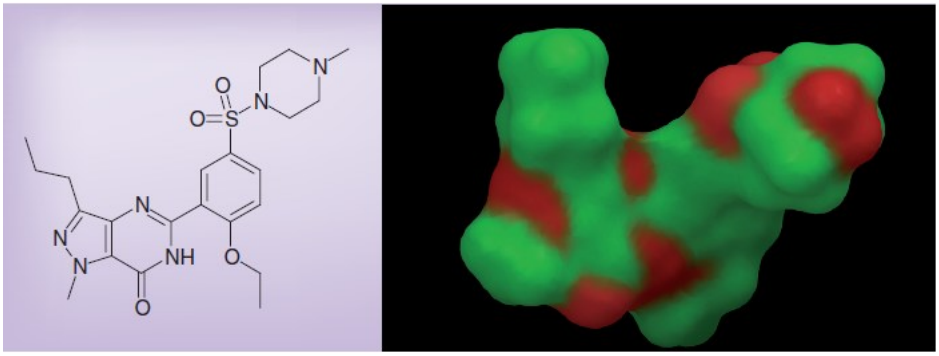


Figure 8. Van der Waals surface of Sildenafil and its polar surface area in red. [48]

4.1.1 First PSA studies

Some researchers first used this concept as a useful descriptor in predictive models of ADME properties of NCEs. In this research, the authors focused on the "hydrophilic portion of the Van der Waals surface", defined as the surface area subtended between nitrogen and oxygen atoms, while all other types of atoms were defined as hydrophobic [52].

The computational method to calculate PSA has evolved significantly. *Dynamic Polar Surface Area* (PSAd) introduced by Palm et al., which averages multiple conformers weighted by their thermodynamic stability expanded the concept of PSA by introducing PSAd**,** which considers the Boltzmann-weighted average of all low-energy conformers of a molecule [51]. This approach significantly improved correlations between PSA and drug permeability across both Caco-2 monolayers and rat intestinal segments, outperforming traditional lipophilicity metrics such as Log $D_{7.4}$ [47]. This suggests that incorporating molecular flexibility into PSA calculations yields a more accurate prediction of intestinal absorption, especially for more flexible compounds.

Also, in another study conducted at the University of Uppsala in Sweden, the concept of PSAd was introduced and correlated with experimentally measured intestinal permeability in Caco-2 [51] cell monolayers.

The value has been calculated following these steps:

- Detailed conformational studies of compounds
- Low energy conformers selection (2.5 kcal/mol)
- PSA calculation for each conformer
- PSA average considering the existing probability of each conformer (using Boltzman distribution) [48]

The disadvantage of this method lies in the time required to calculate PSAd, which was mainly related to the need to conduct detailed conformational search (i.e. several hours of calculation for each selected compound).
As a result of this, it was very difficult to profile even modest collections of compounds in a reasonable amount of time.
To reduce time and cost of the computational calculation, an alternative approach was later evaluated by other research groups, whereby only a single conformer, rather than a set of conformers, was used for the calculation.

Clark et al. proposed a streamlined algorithm using a single low-energy conformer, avoiding the complexity of Boltzmann-averaged conformational ensembles an approach that retains excellent correlation with absorption data while offering considerable time savings [53]. In this way, the correlation between *static Polar Surface Area* (PSAs) of individual conformers and permeability was only slightly lower than that obtained with PSAd [47], but on the other hand, this avoided the need for a detailed conformational search and the time taken for PSA calculation of a single conformer was thus reduced to a few seconds per compound [53]. This method proved nearly as effective as the more computationally expensive *dynamic Polar Surface Area* (PSAd) introduced by Palm et al., which averages multiple conformers weighted by their thermodynamic stability [51].

After, in the 2000s, several research groups introduced *Topological Polar Surface Area* (TPSA), an index based on the addition of tabulated surface contributions of polar fragments, which allows the analysis of large datasets. This has eliminated the need to generate a three-dimensional structure and calculate the surface area. It is based on PSA data from around 35,000 molecules and based on the TPSA contribution of 43 fragments centered on polar atoms such as oxygen, nitrogen, and sometimes is possible to add in the calculation sulfur and phosphorus, with tabulated values [54].
TPSA calculations are extremely fast. Thousands of compounds per second can be processed, enabling the profiling of extremely large virtual databases in a reasonable amount of time.
Having access to a simple, cheap, and efficient way to assess polarity from a 2D structure, which correlates well with cell permeability, the need for experimental polarity measurement has never really been a priority need in medicinal chemistry.

4.1.2 Polarity in the drug discovery

PSA has found ready application as a descriptor in many *structure-activity relationships* (SARs), particularly those relevant to models of ADMET processes.

In particular, the use of PSA is very useful for optimizing cellular potency, oral absorption, intestinal permeability and across BBB. It is important to clarify that although it is very often useful in medicinal chemistry, achieving a desirable PSA value in NCEs is no guarantee of success.
A classic negative example is a compound with only one quaternary nitrogen, which may be characterized by a low PSA value, but, on the other hand, being permanently charged, has a very difficult passive diffusion across the membrane. Conversely, some compounds with very high PSA values can permeate efficiently through membranes because they are substrates of transporters.
It is therefore essential to consider PSA, and all the other physico-chemical parameters, as a key physico-chemical property that should always be integrated into the multifactorial optimization process of NCEs. At the same time, other properties such as lipophilicity and ionization state, typically represented by Log *P*, Log *D*, and pK_a should be equally important depending on the specific context [55], [56].

The speed and accuracy of TPSA have made it the method of choice for PSA calculations for most researchers. TPSA is available, along with other physico-chemical properties, online via *freeware* applications or in commercial software such as Molinspiration, Chemdraw or ACD/Labs Percepta. The emergence of such tools on the web and the rapid growth of interest *in-silico* ADMET predictions has certainly facilitated the assimilation and application of TPSA by medicinal chemists.
As previously discussed, one of the main limitations of using TPSA lies in its fragment-based nature. This approach, although computationally efficient, fails to discriminate between regioisomers, such as the ortho, meta, and para-analogues of polar aromatic compounds, since these often share identical or nearly identical TPSA values. This limitation underscores the need for complementary analytical methods capable of capturing the full structural and conformational complexity of a molecule, especially when subtle positional changes can influence pharmacokinetic behavior [57].

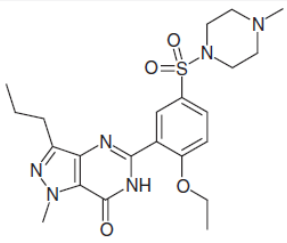
Structure	Fragment	Fragment TPSA Value (Å²)	Frequency of fragment in structure	Sub-total (Å²)
	[n](:*):*	12.89	1	12.89
	[n](-*)(:*):*	4.93	1	4.93
	[O]=*	17.07	3	51.21
	[N](-*)=*	12.36	1	12.36
	[NH](-*)(-*)	12.03	1	12.03
	[O](-*)-*	9.23	1	9.23
	[N](-*)(-*)-*	3.24	2	6.48
	Total TPSA (Å²)			**109.13**

Figure 9. Example of Topological Polar Surface Area calculation for Sildenafil [48]

This implies that, in the design of new drug candidates, molecules with the lowest possible PSA values are generally preferred in order to achieve good bioavailability. In this context, the

development of novel analytical methods for the experimental determination of PSA in NCEs is of paramount importance. Such methods are essential for building predictive models of cellular permeability, guiding rational structural modifications to optimize molecular polarity, and supporting the development and continuous refinement of computational tools for PSA estimation.

This need has become even more critical in recent years due to the growing interest in structurally complex molecules such as peptides, macrocycles, and PROTACs, which often challenge traditional drug-likeness criteria and require a deeper understanding of their physico-chemical properties.

4.1.3 PSA and absorption

As previously mentioned, PSA can serve as a valid alternative to the *n*-octanol/water partition coefficient for estimating a compound's permeability and absorption but considering the combination of this parameter is always the first choice. One of the first studies to propose PSA cut-off values dates back to 1997, when researchers demonstrated, on a chemically diverse set of 20 drugs, those compounds with an absorbed fraction greater than 90% exhibited PSA values ≤ 60-90 Å², whereas those with an absorbed fraction below 10% had PSA values > 140 Å² [51]. A subsequent study further expanded on these findings, showing that a PSA ≤ 140 Å², combined with a number of rotatable bonds ≤ 10, could serve as an effective and selective criterion for identifying compounds with oral bioavailability greater than 20-40% [58]. Regarding permeability across the BBB, more stringent PSA thresholds have been proposed. A PSA value below 90 Å² has been commonly cited as a general upper limit, although other studies suggest an even lower threshold of 60 Å² to ensure central nervous system penetration [59].

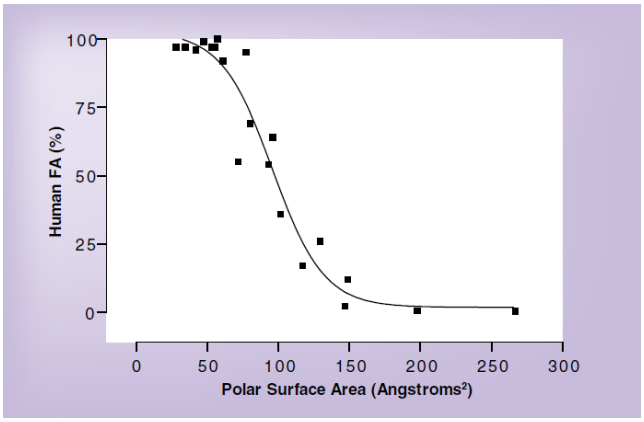


Figure 10. Polar Surface Area vs human absorption [48]

Winiwarter et al. (1998) provided strong human in vivo evidence, correlating jejunal permeability (P_{app}) with PSA and hydrogen bond donor (HBD) count using multivariate analysis. Their findings indicate that PSA combined with HBD and Log *D* at physiological pH creates a reliable model for

predicting passive drug absorption in humans. Importantly, the predictive models were validated on an external set of 34 compounds with consistent accuracy, reinforcing PSA utility in real-world drug development scenarios [55].

Lastly, Veber et al. (2002) analyzed over 1,100 drug candidates and identified that low PSA (≤ 140 Å²) and fewer than 10 rotatable bonds are strong predictors of good oral bioavailability in rats. Interestingly, PSA alone correlated more strongly with membrane permeation rate than calculated lipophilicity (CLog *P*), particularly when considered alongside molecular flexibility. This empirical rule, derived from a large and structurally diverse dataset, offers a practical guideline for early compounds screening [58].

Beyond intestinal absorption, as cited before, polarity is also a key determinant of brain penetration**.** The BBB restricts the entry of polar molecules, and thus low PSA is typically a prerequisite for CNS activity. Compounds with PSA below 60-90 Å² are more likely to cross the BBB, while values above 120-140 Å² tend to indicate CNS exclusion [59].
Furthermore, PSA has demonstrated greater predictive power than traditional empirical filters like Lipinski's Rule of 5. While the Rule of 5 evaluates general drug-likeness based on molecular weight, lipophilicity, and hydrogen bond count, PSA provides a more spatially accurate measure of hydrogen bonding potential, accounting for 3D conformation and intramolecular shielding of polar groups [51]. Sometimes, PSA is included in the Lipinski's Rule of Five and beyond Lipinski's Rule of Five to improve the power in the prediction of drugs absorption [60].

Still, PSA is not a universal predictor. Compounds absorbed via active transport mechanisms may violate the PSA rule, exhibiting high absorption despite large PSA values. For example, methotrexate and zidovudine both exceed 140 Å² but are efficiently absorbed through carrier-mediated pathways. Methotrexate is absorbed by a carrier-mediated process which is responsible for folate absorption (Reduced Folate Carrier SLC19A1 and Proton-coupled Folate Transporter SLC46A1), while zidovudine is absorbed by a recombinant nucleoside transporter which is responsible for the absorption of pyrimidine nucleosides (Equilibrative Nucleoside Transporters ENT1 and ENT2) [53]. But also compounds with PSA above 140 Å² can passive permeate through biological membrane and can be characterized by high permeability even if their PSA is above range and these molecules could be discarded during the screening cascade. With the increase in the number of bRo5 compounds in the pharmaceutical field, such as peptides, macrocycles and PROTAC, a high throughput experimental method to measure the real exposed polarity of the compounds taking into account

conformational freedom and intramolecular hydrogen bond formation (IMHBs) is increasingly needed, to avoid loss of compounds during the screening cascade.

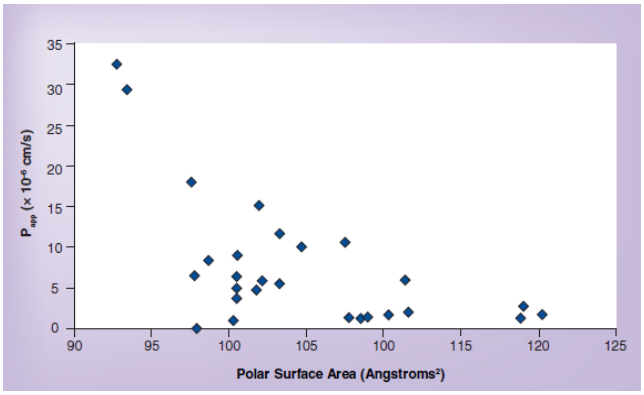


Figure 11. Polar Surface Area vs Caco-2 P_{app} (10^{-6} cm/s) [48]

In this context, recent analyses of orally bioavailable bRo5 drugs have shown that PSA values can extend well beyond the classical threshold of 140 Å², reaching up to 200 or even 250 Å² without necessarily compromising permeability. This apparent contradiction is explained by the formation of intramolecular hydrogen bonds and conformational adaptations that effectively mask polar groups from solvent exposure. These dynamic changes reduce the "exposed" polarity of the molecule, allowing it to behave more like a compound with lower PSA. The study proposes that such internal shielding mechanisms account for the successful absorption of many bRo5 compounds, despite their large size and polar surface area. Accordingly, the concept of exposed PSA (EPSA) emerges as a more meaningful descriptor for permeability in flexible, high-MW molecules compared to static TPSA measures [60].

4.2 Most common Polar Surface Area evaluation techniques

4.2.1 Computational method

4.2.1.1 Topological Polar Surface Area (TPSA)

In the present study, polar surface area was assessed using two complementary approaches: a computational method through the TPSA and a chromatographic experimental method through the *Experimental Polar Surface Area* (EPSA)**.**

TPSA, introduced in the early 2000s, is a purely computational descriptor based on summing predefined surface contributions of specific polar fragments, typically atoms like oxygen, nitrogen and sometimes sulfur, and phosphorus, depending on their bonding context [54]. These fragment contributions were derived from a dataset of approximately 35,000 drug-like molecules and enable the estimation of polarity without requiring three-dimensional structure generation or surface

modeling. The key advantage of TPSA lies in its exceptional computational efficiency. Since the method relies only on the molecule's 2D topology (e.g., SMILES format), it can process thousands of structures per second, making it ideal for high throughput *in silico* screening campaigns. Consequently, TPSA has become a widely accepted surrogate for polarity in medicinal chemistry, available in both commercial (e.g., ACD/Labs Percepta, ChemDraw) and free platforms (e.g., Molinspiration, SwissADME). The descriptor's ability to approximate permeability-related properties such as absorption and BBB penetration has been validated through comparison with experimental datasets [48].

Despite its widespread use, TPSA presents important limitations. Its fragment-based additive nature does not account for conformational differences, which may lead to indistinguishable TPSA values among regioisomers, particularly problematic for aromatic compounds with ortho-, meta-, or para-substitution patterns. These subtle variations, though chemically meaningful, are invisible to the TPSA algorithm, potentially reducing predictive accuracy in certain contexts [57].
For this reason, there is increasing interest in experimental methods, such as EPSA, that incorporate three-dimensional information, intermolecular interactions, and chromatographic behavior.
By comparing TPSA with EPSA in this work, we aim to evaluate not only the computational speed and accessibility of TPSA, but also its limitations in reflecting true molecular polarity, especially for structurally complex or flexible drug-like compounds. The complementary nature of these two methods provides a broader perspective on polarity profiling in modern drug discovery.

4.2.2 Chromatographic Method

4.2.2.1 Experimental Polar Surface Area

Due to the limitations of computational methods in polarity evaluation, and given the importance of obtaining reliable experimental data, a chromatographic approach based on *Supercritical Fluid Chromatography* (SFC) was developed by Goetz et al. [16], [18], [49]. Computational calculation like TPSA often fail to account for the molecular flexibility and the influence of the environment on conformational behavior. In contrast, SFC employs supercritical carbon dioxide (CO_2) as the main mobile phase, creating an apolar environment with a low dielectric constant, conditions that can promote the formation of weak IMHBs, similar to those that may form near biological membranes. This is particularly relevant because molecules with high TPSA values might be prematurely excluded during early-stage screening, despite possessing favorable ADME properties. Structural features enabling the masking of polar groups through IMHBs formation can reduce the effective polarity, potentially improving membrane permeability and bioavailability. In a recent study, Goetz

and colleagues emphasized how the conformational flexibility of complex structures, such as cyclic peptides, facilitates IMHBs formation and decreases polarity, thereby increasing the likelihood of permeability and enhancing their drug-like profile [49].

Goetz and colleagues selected a training set of 118 standard compounds with known, well-characterized TPSA values. These compounds were carefully chosen to cover a broad range of polar surface areas typically encountered in medicinal chemistry and they were unable to form any IMHBs. The standards were injected, and their retention times were measured under controlled SFC conditions. Because the mobile phase was mostly CO_2 with a small organic modifier, retention on the polar stationary phase was driven primarily by the compound's exposed polarity. The TPSA values of the standards were plotted against their observed retention times. The assumption here was that for these relatively rigid and simple standards, TPSA would approximately exposed polarity well. A linear correlation was established, allowing retention time to be translated into a “measured” polar surface area (EPSA). For test compounds, retention times were measured under the same chromatographic conditions and converted into EPSA values using the calibration curve [18], [49].

As mentioned before, an additional advantage of this chromatographic method is its ability to distinguish between compounds capable of forming IMHBs and those that are not. This is typically achieved through pairwise analysis using a reference compound known to be incapable of forming IMHBs. If the test compound exhibits a lower EPSA value under the same conditions, it is indicative of IMHB formation. It is important to note, however, that this remains an indirect method; only spectroscopic techniques such as NMR can provide direct evidence of IMHBs [49].

Recent advances have reinforced the importance of EPSA as a predictive tool in early drug discovery. In particular, Wang et al. (2024) developed a high-throughput supercritical fluid chromatography-tandem mass spectrometry (HT-SFC-MS/MS) method to measure EPSA, offering significantly improved speed, sensitivity, and data quality compared to earlier approaches. Their work demonstrated that EPSA outperforms topological TPSA and PAMPA assays in predicting human intestinal and Caco-2 permeability, particularly for structurally complex, bRo5 compounds such as PROTACs. By enabling more accurate permeability ranking and faster decision-making, EPSA is shown to be a highly valuable physico-chemical descriptor for optimizing membrane permeability during early-stage screening and compound triaging [14].

5 The Hydrogen Bond

A hydrogen bond is a type of molecular interaction that occurs when a hydrogen atom (H), covalently bonded to a highly electronegative atom such as oxygen (O) or nitrogen (N), is attracted to another electronegative atom. Due to the electronegative nature of the atom, it is bonded to, the hydrogen acquires a partial positive charge, making it capable of interacting electrostatically with another lone pair of electrons on a nearby electronegative atom. This interaction can occur either between different molecules (intermolecular hydrogen bonding) or within the same molecule (intramolecular hydrogen bonding), depending on the spatial arrangement of the functional groups.

This type of interaction is classified as a non-covalent bond with a bond strength lower than that of true chemical bonds, such as ionic or covalent bonds, but stronger than Van der Waals forces, which are relatively weak attractive or repulsive interactions between molecules. The strength of a hydrogen bond depends on several factors, including the nature of the hydrogen bond donor (HBD) and acceptor (HBA), their relative orientation (geometry), and the surrounding chemical environment. As a result, hydrogen bond energies can vary significantly, typically ranging from 0.2 to 40 kcal/mol^{-1} [61].

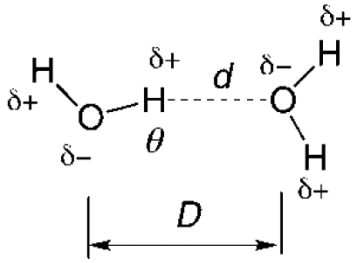


Figure 12. Hydrogen bond in water dimer.
The geometrical parameters are: "d" = H---O distance; "D" = O---O distance; θ = O-H---O angle [61]

HBD is defined as a hydrogen atom covalently bonded to an electronegative atom, such as oxygen and nitrogen. A hydrogen bond is primarily an electrostatic dipole-dipole interaction, yet it also exhibits some covalent character, particularly in terms of directionality. This means that the optimal hydrogen bond strength is achieved when the lone pair of the acceptor atom, the hydrogen atom, and the donor atom (e.g., oxygen or nitrogen) are aligned in a nearly linear geometry. Deviation from this ideal geometry results in weaker interaction, as the orbital overlap and electrostatic attraction are reduced [61].

The ability of molecules to form hydrogen bonds plays a crucial role in biological systems. In nucleic acids such as DNA, hydrogen bonding is the key interaction that stabilizes the double helix, holding

the two complementary strands together through specific base pairing. Similarly, in proteins, hydrogen bonds contribute significantly to the stabilization of both tertiary and quaternary structures, helping to maintain the correct folding and overall conformational equilibrium that are essential for biological function. From a physico-chemical perspective, the presence of hydrogen bonds increases boiling points and intermolecular distances compared to compounds lacking such interactions. This is because hydrogen bonding introduces additional attractive forces that must be overcome during phase transitions, and it also influences the molecular arrangement in condensed phases.

5.1 Intramolecular Hydrogen Bond (IMHB)

IMHBs are non-covalent interactions formed within a single molecule, which play a critical role in modulating conformation, solubility, membrane permeability, and bioavailability of a drug. These interactions enable a molecule to dynamically adapt to different environments by masking polar groups and influencing solvation and interaction with biological membranes [10]. It occurs when a hydrogen bond (HB) is formed between an HBD and an HBA within the same molecule. This process is governed by an equilibrium described by thermodynamic constant, K_{imhb} [20], that quantifies the presence of the molecule in the open-close conformation. Indeed, molecules that can form IMHBs in non-polar environment disguise their polarity. The same molecules can expose the donor groups in polar medium, as the water.

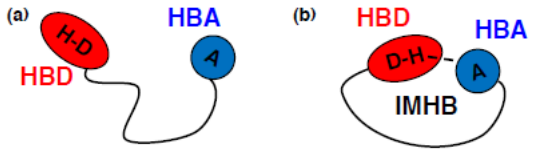


Figure 13. IMHB formation between a HBD and a HBA in the same structure [20]

IMHBs also play a crucial role in altering a molecule's lipophilicity, the tendency to partition into lipid environments rather than aqueous phases. This is typically quantified using partition coefficients such as Log P_{oct} and Log P_{alk}. IMHBs can reduce the hydrogen bonding capability of polar groups by internally satisfying donor-acceptor interactions, thus enhancing apparent hydrophobicity [62]. Interestingly, IMHBs may increase, decrease, or have little effect on Log P_{oct} values, depending on the donor and acceptor strength involved. However, they almost always increase Log P_{alk}, which is more sensitive to internal polarity shielding. This makes them more reliable indicators of true lipophilicity when IMHBs are present [63]. To better assess how much hydrogen bonding within a molecule is internally masked by intramolecular interactions, the *Hydrogen Bonding Capability difference* (ΔHBC) was developed as a descriptor. This value represents the gap between a molecule's theoretical hydrogen bonding potential, calculated based on its chemical structure, and its

experimentally observed hydrogen bonding behavior. When ΔHBC is high, it suggests that some hydrogen bond donors or acceptors are engaged in intramolecular hydrogen bonds (IMHBs), thereby reducing their availability to interact with the surrounding environment, such as solvents. In this way, ΔHBC serves as an indirect indicator of the presence and strength of IMHBs [62].

5.1.1 Intramolecular Hydrogen Bond in drug discovery

As medicinal chemistry expands into the bRo5 chemical space, characterized by large, flexible, and polar molecules, *chameleonicity* becomes a powerful concept. *Chameleonicity* is the ability of a molecule to adapt its conformation in response to environmental polarity, typically by forming IMHBs that shield polar groups in nonpolar environments [64].

Molecules that display *chameleonicity* can retain water solubility while improving membrane permeability, a key advantage in drug development. For example, macrocyclic drugs like cyclosporine and non-macrocyclic bRo5 drugs like saquinavir can adopt closed conformations in lipid membranes and extended conformations in aqueous environments, thanks to IMHB-driven conformational flexibility [64].

Incorporating IMHBs and *chameleonicity* analysis into drug design enables medicinal chemists to balance solubility and permeability, two parameters often at odds in the bRo5 space. Molecules can be engineered to form reversible IMHBs that optimize polarity exposure depending on the surrounding medium, enhancing both solubility in plasma and permeability through membranes [62], [64].

In the beginning of 2000, some researchers defined that the formation of these bonds are fundamental to increase permeability of cyclic peptide [65]. In 2012 other researchers reported the importance of IMHBs to increase permeability and sometimes potency, and to decrease P-gp efflux [66].

Additionally, the ΔHBC and chromatographic methods as the ChameLog k index offer predictive tools that guide scaffold optimization, enabling the early identification of compounds with favorable biopharmaceutical profiles. As chemical space exploration continues to evolve, these tools will likely become essential components in both computational modeling and experimental workflows for next-generation drug discovery. These two values are examples of tools that medicinal chemists can use to understand the behavior and the capability of compound to form IMHBs.

The implementation of this consideration about the IMHBs in drug discovery requires experimental technique to evaluate the tendency of compound to form IMHBs. Indeed, molecular models can suggest structural modification to implement IMHBs in the compound unable to form these interactions to also improve ADME properties [20]. After synthesis this compound may be

experimental high throughput evaluated, and from this need arises the need to have analytical methods able to evaluate this propensity.

5.2 Intramolecular Hydrogen Bond: evaluation techniques

Nuclear Magnetic Resonance (NMR) is the most common technique used to measure the formation of IMHBs and is the most appreciated in the medicinal chemistry field. It was used to confirm the presence of IMHBs of the compounds used for this work. This technique measures the 1H NMR chemical shift in two different solvents, usually DMSO and chloroform. The difference between the chemical shifts of an OH or NH group in these solvents can be used to demonstrate the hydrogen bond acidity (HBD properties). This approach can be also used for peptides and proteins, to evaluate the tertiary and quaternary structure. NMR is very useful, but unfortunately it is not a high throughput method to measure the IMHBs formation, despite the progress in the instrumentation performance in terms of the amount of compound analyzed and the time for the analysis [67].

The IMHBs formation can be evaluated also using X-ray and neutron diffraction, preferred from X-ray crystallography due to its limitation about the determination of the position of hydrogen atoms due to their low electron density. This makes neutron diffraction a superior and more precise method for studying IMHBs because this technique directly interacts with atomic nuclei. Another important aspect is that crystallography and neutral diffraction data are not available during early drug discovery stage. Despite this, it is very useful for retro prospective studies.

5.2.1 Log P/D$_{oct}$

The Log P_{oct}, or Log D_{oct} when ionizable compound are considered, could be used to evaluate the propensity of compounds to form IMHBs. As discussed in the previous chapters, Log P_{oct} is the lipophilicity descriptor most known and used in early drug discovery. However, the system formed by *n*-octanol and water gives a non-polar environment in which only strong IMHBs could be formed. This means that Log P_{oct} can be used to evaluate this propensity only for high tendency, while it is not the best parameter for IMHBs with low or medium propensity, like molecular chameleons.
In this type of study, it is suggested to use pairwise analysis comparing compounds able to form IMHBs with compounds unable to form IMHBs.

For example, in one of these studies two similar molecules were evaluated, and they demonstrated that compound 3a in *Figure 14*, the control, in which the benzimidazole nitrogen is methylated, is unable to form any IMHBs and it has the Log P_{oct} lower than the compound 3b (0.68 *vs* 1.39) that shows high propensity to form IMHBs. This study reveals the difficulty finding a good control compound because in theory compound 3a, with methylation, should have a higher lipophilicity than reality, thus making it more difficult to compare to 3b [20].

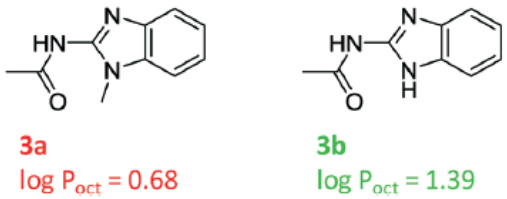
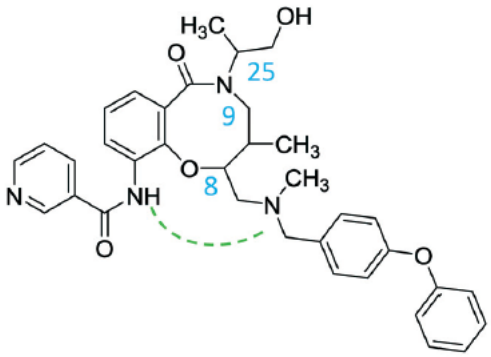


Figure 14. Different Log P_{oct} values between compound 3a unable to form IMHB and compound 3b able to form IMHB [20]

Another application concerning the use of Log D_{oct} to highlight the propensity of a compound to form IMHBs was shown in a study about the physico-chemical properties of a series of 8 beta-lactam stereoisomers (Figure 15), active against *Tripanosoma Cruzi*. In this retrospective study, Log D_{oct} of stereoisomers 1-4 is significantly higher than stereoisomers 5-8. This and other experimental results demonstrated that the (1-4) stereoisomers have a higher propensity to form IMHBs than the other 4 (5-8) stereoisomers that are less lipophilic [20].

As said before, these parameters can evaluate the propensity only for strong IMHBs; the lack of capacity for weak and medium IMHBs is an important limit if there is the interest to have molecular chameleons, compounds that can disguise their physico-chemical profile and their ADME properties on the basis of the environment in which they are.



Cpd	Stereochemistry (C8, C9, C25)	log D_{oct}
1	SSS	4.5
2	RRS	4.3
3	SSR	4.5
4	RRR	4.5
5	SRR	3.5
6	RSS	3.6
7	RSR	3.8
8	SRS	3.9

Figure 15. Different LogD_{oct} values for eight stereoisomeric beta-lactam inhibitors of Tripanosoma Cruzi [20]

5.2.2 ΔLog $P_{oct\text{-}tol}$

One of the best-known physico-chemical descriptors for measuring the propensity of compounds to form IMHBs is ΔLog $P_{oct\text{-}alk}$ [63]. This parameter is defined as the difference between Log P_{oct} and Log P_{alk}. Log P_{alk} is defined as the partition coefficient in alkanes. Interest in this descriptor stems primarily from the study of a research group that demonstrated that ΔLog $P_{oct\text{-}alk}$ can be used in the prediction of brain permeation of drug candidates. Unfortunately, the experimental measurement of this descriptor suffers from serious limitations mostly related to the poor solubility of the compounds in most alkanes [68].

To overcome this limitation, ΔLog $P_{oct\text{-}tol}$ has been introduced. This parameter is more convenient than ΔLog $P_{oct\text{-}alk}$ because compounds are more soluble in toluene than in other alkanes, and the obtained information is the same [63].

When a compound has a single HBD group, the interpretation of ΔLog $P_{oct\text{-}tol}$ is simple. If the value of ΔLog $P_{oct\text{-}tol}$ is close to 0, then the compound has a high propensity to form IMHB. In fact, when the value of ΔLog $P_{oct\text{-}tol}$ is about 0, the value of Log P_{tol} is comparable with the value of Log P_{oct}, and this is only possible if the compound in toluene masks its polarity, that is, it forms IMHBs. Conversely, the higher the value of ΔLog $P_{oct\text{-}tol}$, the lower the propensity of the molecule to form IMHBs.

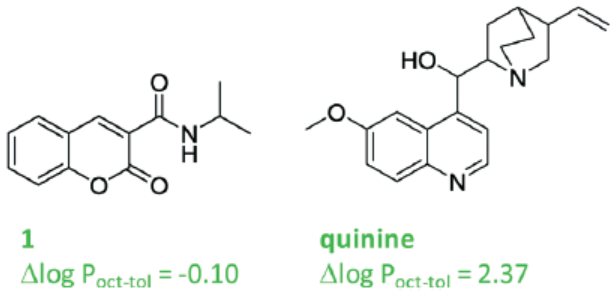


Figure 16. Different Log $P_{oct\text{-}tol}$ values for compound 1 able to form IMHB and quinine unable to form any IMHB [20]

Two examples are shown in Figure 16 where the propensity to form IMHB of compound 1 is revealed by its ΔLog $P_{oct\text{-}tol}$ value of about 0, and the NMR experiments in the original work supported this result [63]. In contrast, quinine has a ΔLog $P_{oct\text{-}tol}$ of about 2.3 and therefore does not form IMHB.

The interpretation of ΔLog $P_{oct\text{-}tol}$ is as complex as the greater the number of HBD groups that are present in the chemical structure because one of them might be involved in the formation of IMHBs. In these situations, some HBDs might be exposed to and not involved in IMHBs formation. In these cases, interpretation of ΔLog $P_{oct\text{-}tol}$ should be performed as described by M. Shalaeva et al. where he explains to evaluate this feature by using a control compound. If the ΔLog $P_{oct\text{-}tol}$ of the sample is

lower than that of the control, the sample has a high propensity to form an IMHBs. Conversely, if ΔLog $P_{oct\text{-}tol}$ of the sample is greater than that of the control, the sample has a low propensity to form IMHB [63].
In general, ΔLog $P_{oct\text{-}tol}$ is a powerful descriptor of the propensity of compounds to form IMHBs. Solubility problems still persist, however, and some experience in its interpretation is required.
For this reason, chromatographic systems that can be easily configured with different types of stationary and mobile phases are preferred to optimally reproduce many types of environments [69].

5.2.3 EPSA and chromatographic methods

To quantify this property, the Chamelogk index was developed. It compares chromatographic retention in polar and nonpolar solvent gradients using a polymeric RP-HPLC system. The difference between extrapolated and experimentally measured retention time in 100% acetonitrile reflects the extent of conformational adaptation. High Chamelogk values (e.g., > 0.6) are indicative of pronounced chameleonic behavior [69]. This index has proven especially useful for ranking and selecting bRo5 drug candidates, including PROTACs and complex peptides. In such molecules, high Chamelogk correlates with better oral bioavailability, making it a valuable predictor in early-stage screening [64].

Another chromatographic descriptor used to assess the exposed polarity of a molecule is the EPSA [16], [18], [49]. It is derived from retention data obtained using Supercritical Fluid Chromatography (SFC), where a polar stationary phase (e.g. Chirex 3014) is paired with a nonpolar mobile phase consisting of supercritical carbon dioxide modified with methanol and 20 mM ammonium formate. This setup establishes a low-polarity environment that tends to promote the folding of molecules capable of internal hydrogen bonding, using methanol with ammonium formate to elute more polar compounds.
By employing a pairwise comparison approach, EPSA can differentiate between molecules that are likely to form intramolecular hydrogen bonds (IMHBs) and those that are not. This is because the presence of IMHBs generally reduces the number of polar groups exposed to the solvent, resulting in a lower EPSA value. Importantly, studies have demonstrated that EPSA is primarily sensitive to the hydrogen bond donor (HBD) characteristics of molecules, with minimal influence from molecular size. Nevertheless, the EPSA method is not without limitations. Notably, molecules containing multiple hydrogen bond acceptor (HBA) groups may exhibit artificially low HBD contributions, potentially leading to underestimation of overall polarity. This effect, although not yet fully understood, restricts the method's reliability in certain chemical contexts [20].

6 Aim of work

The objective of this thesis project, conducted at the research laboratories of Chiesi Farmaceutici S.p.A., within the *Analytics and Early Formulations* (A&EF) Department of the *Global Research & Preclinical Development* division, was to experimentally characterize key physico-chemical properties commonly assessed during the early stages of drug discovery, using chromatographic methods. These properties include lipophilicity, measured as *Chromatographic Hydrophobicity Index* (CHI Log $D_{7.4}$) and polarity, expressed as *Experimental Polar Surface Area* (EPSA), which were evaluated across a selected set of 19 commercial bRo5 drugs. In addition to profiling the selected molecules, CHI Log $D_{7.4}$ and EPSA were also experimentally determined for hundreds of compounds from Chiesi's proprietary chemical library.

Golden standard methods to measure lipophilicity and polarity values, together with acidic dissociation constant (pK_a), were also used for the comparison with chromatographic experimental values of the selected dataset. Specifically, the lipophilicity (Log $D_{7.4}$ and Log P_{oct}) was determined using both *shake-flask partitioning* and *potentiometric titration methods* as experimental golden standard techniques, while *Topological Polar Surface Area* (TPSA) was calculated using computational tools as calculated golden standards technique. The experiments for measuring lipophilicity by titration methods were carried out at the University of Parma using the SiriusT3™ instrument.

The predictive power of these experimentally obtained values was evaluated by comparing them to passive permeability data generated via the Caco-2 cell assay, which is widely considered a golden standard method in permeability studies. Furthermore, EPSA and ΔLog $P_{oct\text{-}tol}$ values were specifically examined to assess their ability in identifying the presence of IMHBs. In this context, NMR spectroscopy was used as a positive control technique to verify IMHBs formation.
The Caco-2 permeability data was made available thanks to the collaboration with Chiesi's *Pharmacokinetics, Biochemistry and Metabolism* (PKBM) Department. Together, chromatographically measured physico-chemical parameters enabled comprehensive profiling of the selected compound series and supported the development of rational correlation models between EPSA, CHI Log $D_{7.4}$, and permeability metrics. At the end, all experimentally acquired data, including both chromatographic parameters and gold standard references (when available), were systematically compared and correlated with the permeability behavior of the investigated compounds; this analysis highlighted pro and cons of each physico-chemical value. The performance of these chromatographic metrics in predicting passive permeability, as well as their utility in

highlighting promising compounds during the early stages of drug discovery within the screening cascade, were critically evaluated and compared to the golden standard assays.

Moreover, the research project I undertook during my international experience at Uppsala University (Sweden), within the *Analytical and Pharmaceutical Chemistry* department (AFK) and in collaboration with AstraZeneca separation scientist, focused on developing a multivariate model to predict chromatographic retention time based on calculated physico-chemical properties of a series of compounds. The methodology, model development, and results of this project will be described in detail in the last chapter.

To avoid repetition and be more concise, the list of the 19 compounds analyzed in this study, for each physico-chemical property, is shown in the Table 1 below:

Table 1. List of the 19 compounds analyzed and the supplier of each compound

Compound	Supplier
ACTH fragment 4-10	Abcr/Merck-Sigma
Angiotensin II	Abcr/Merck-Sigma
Caspofungin diacetate	abcr/Merck/BLDpharma
cetrorelix acetate	Merck-Sigma/Sigma-Aldrich
Desmopressin	Abcr/Merck-Sigma/Bachem/Aaron chemicals
Erythromycin	Sigma-Aldrich
Everolimus	Merck-Sigma/BLDpharma
Goserelin diacetate	Merck-Sigma/abcr
Lanreotide acetate	Abcr/Merck-Sigma/Sigma-Aldrich
Octreotide acetate	Abcr
Oxytocin	Abcr/Merck-Sigma
Rifampicin	Sigma-Aldrich
Rifapentine	Sigma-Aldrich
Rifaximin	Sigma-Aldrich/BDLpharma
Ritonavir	Merck-Sigma
Roxithromycin	Sigma-Aldrich
Saquinavir mesylate	Merck-Sigma/BLDpharma
Teicoplanin A_2-2	Merck-Sigma
Vancomycin	Abcr

7 Experimental section - Acid dissociation constant (pK_a)

7.1 Aim

As previously discussed in the preceding chapters, the determination of pK_a values is one of the most critical steps in the early phases of drug discovery. Knowledge of compound's pK_a is essential for predicting its ionization state under physiological conditions, which directly influences its solubility, permeability, and overall pharmacokinetic profile.

In the context of this PhD project, it is particularly important to assess whether the experimentally determined pK_a values agree with those predicted *in-silico*.

In addition to its predictive utility, pK_a determination is crucial for the proper interpretation of lipophilicity data obtained using the shake-flask method. At physiological pH (7.4), a compound may be present in a neutral or ionized form, significantly affecting its partitioning behavior between *n*-octanol and water phases. This distinction dictates whether the lipophilicity should be expressed as a partition coefficient (Log *P*) or as a distribution coefficient at pH 7.4 (Log $D_{7.4}$). A precise understanding of the ionization profile is therefore fundamental when correlating physico-chemical properties to biological activity.

Log *D* at pH 7.4 is the standard ADME descriptor and most clinically relevant for systemic distribution. However, for compounds with ionizable groups (~95% of drugs), we are aware multi-pH profiling (pH 2, 5.5, 7.4, 8.0) provides superior predictive power for intestinal absorption, hepatic metabolism, renal clearance, and tissue distribution.

The experimental pK_a determination was carried out using pH-metric and UV-metric titration, a well-established analytical technique for this purpose. The automated titrator SiriusT3™ was employed to perform the measurements under standardized conditions. A total of 19 commercially available compounds were selected for the analysis. These compounds were characterized by complex structures and the presence of multiple ionizable functional groups. The selection specifically included bRo5 molecules, such as peptides and macrocyclic compounds, which are often associated with low solubility or lipophilicity, high molecular weight, and flexible conformational behavior.

This study also aimed to evaluate the suitability and performance of the SiriusT3™ system in handling structurally intricate and highly ionizable compounds. The results obtained provided valuable insights into the applicability and limitations of the potentiometric method for the accurate characterization of bRo5 compounds, contributing to a better understanding of their physico-chemical behavior.

7.2 Materials

7.2.1 Materials used to measure the pK_a using SiriusT3^{TM} automatic titrator

To measure the pK_a of the 19 commercial compounds using SiriusT3^{TM} automatic titrator, the following chemical reagents were used: dimethyl sulfoxide (Sigma-Merck), Titrisol unit-dose (Sigma-Merck) used to prepare a solution of hydrochloric acid (HCl) 0.5 N, Titrisol unit-dose (Sigma-Merck) used to prepare a solution of potassium hydroxide (KOH) 0.5 N, potassium dihydrogen phosphate (KH_2PO_4, Sigma-Merck), methanol 80% w/v.
The list of the 19 compounds analyzed is shown in Table 1.

Furthermore, the following instrumentations and accessories were used: SiriusT3^{TM} automatic titrator (Sirius Analytical Instrument Ltd), SiriusT3Control software for data acquisition, SiriusT3Refine for data analysis, analytical balance (Mettler-toledo S.r.l.) characterized by a sensibility of 0.1 mg and a capacity of 100 mg, 1.5 mL Eppendorf vials (Eppendorf), weighting vessels (Sigma-Merck), pipettes P10, P200, P1000 (Gilson), smart spatula (Sigma-Merck), 6 mL vials and Vortex.

7.3 Methods

7.3.1 Methods used for pK_a evaluation

To measure the pK_a using the SiriusT3™ instrument, two different analytical approaches are available: the UV-metric and the pH-metric methods. The UV-metric approach is typically employed for molecules that possess ionizable groups in proximity to chromophores, as changes in the protonation state produce detectable shifts in the UV-absorbance spectra.
In contrast, the pH-metric method, based on potentiometric titration, is used for molecules that lack suitable chromophores, or for those whose spectral changes upon ionization are insufficient for UV detection.
In this project, the pH-metric method was almost exclusively employed, as it is generally more applicable to a broader range of chemical structures, despite requiring slightly larger sample quantities than the UV-metric method, that has been used for only one sample (ritonavir). The reason is that this compound, having only two thiazole groups and being a very weak base, has a pK_a too close to the minimum value measurable by the instrument using pH-metric titration (1.8-2).

For compounds with low aqueous solubility, pK_a values were determined by extrapolating data obtained from titrations performed in methanol-water mixtures with varying methanol percentages. The use of methanol as a cosolvent improves solubility and enables reliable pK_a estimation via *Yasuda-Shedlovsky extrapolation* solvent-correction models.

It is worth noting that sample preparation procedures differ significantly between the two methods. In UV-metric measurements, solutions must be optically clear and within an appropriate absorbance range, while pH-metric titrations require careful control of ionic strength, buffering capacity, and solvent composition to ensure accurate electrode response and proton activity calibration.

7.3.1.1 UV-metric technique

To analyze the samples using the UV-metric technique, the preparation typically is as follows: a 10 mM stock solution was first prepared in DMSO. After, 10 µL of each solution were added in a vial, followed by the addition of 25 µL of K_2HPO_4 buffer solution. In parallel, a blank was prepared in another vial by mixing 10 µL of DMSO and 25 µL of K_2HPO_4 without the sample. The buffer solution is needed to stabilize the pH when the spectrophotometer reads the UV spectra after the addition of titrant while the electrode registers the pH value.
The spectra were registered in the range between 200 to 700 nm, after each pH measurement. The pH changes after the titrant addition are limited to 0.2 pH units. Usually, 20 or 30 pH reads were performed for each titration and all absorbance spectra were collected.

Before started the UV-metric assay, two spectra were registered:

- One dark spectrum, registered when the deuterium lamp was turned off and that should be equal to the noise given by the detector
- The other one, as reference spectrum, in a solution that has a composition as close as possible to that of the sample

These spectra are considered for the calculation of the absorbance of the sample at each wavelength. The instrument measures pK_a values within a pH range of 1.8 to 12. To ensure accurate results with this method, it is advisable to begin at a pH value where the molecule is completely ionized, based on its acidic or basic groups, and then to perform titration with KOH or HCl 0.5 N until reaching pH values at which the molecule is not ionized. For each compound, 3 titrations were performed, and the resulting value is the average of these 3 titrations.

7.3.1.2 pH-metric technique

The pH-metric method was conducted according to the following procedure. Each morning, prior to analysis, the pH electrode was calibrated using acid-base titration. Specifically, a 0.5 N HCl solution was titrated with a 0.5 N KOH solution, which had been previously standardized using potassium hydrogen phthalate as a primary reference standard.

This method is based on the measurement of the potential difference, expressed in millivolts (mV), resulting from the presence of the ionizable compound in solution following the addition of the titrant. This technique can be used to measure pH values from approximately 1.8 to 12. It is generally advisable to begin the titration at a pH where the compound is expected to be fully ionized, to promote adequate solubilization and ensure the reliability of the titration profile. For each assay, 1-2 mg of compound were accurately weighed.

For compounds with limited solubility in water, the assay was performed in the presence of a cosolvent, typically methanol, to enhance solubility. Depending on the solubility profile of the compound, titrations were conducted using various methanol-water mixtures, such as 55%, 45%, or 35% methanol by volume. Following the cosolvent adjustment, the electrode calibration parameters were modified to reflect the impact of methanol on proton activity and electrode response. This adjustment is essential to ensure accurate and comparable pK_a determinations under different solvent conditions. The Yasuda-Shedlovsky linearization was used to determine the aqueous pK_a by extrapolation of the true pK_a in pure water from the measurements performed in water-methanol mixtures.

7.4 Results

The 19 commercial reference compounds selected for pK_a determination were chosen to be representative of different chemical classes, including acidic compounds, basic compounds, and zwitterionic (or ampholytic) species containing both acidic and basic functional groups. This selection was made to ensure a broad representation of ionization behaviors typically encountered in drug-like molecules.

The experimentally determined pK_a values for each compound are reported in Table 2. For comparison and validation purposes, *in-silico* data calculated using ACD/Labs Percepta 2019 are also included. This approach allowed for a critical assessment of the experimental results and their alignment with existing data, supporting the reliability of the methodology employed. All the pK_a were determined using potentiometric approach, using cosolvent for rifapentine and saquinavir, apart from ritonavir for which UV-metric technique has been used due to the presence of two thiazole groups (very weak base) with a pK_a value too close to the minimum value measurable by the instrument using pH-metric titration (1.8-2).

Table 2. pK_a experimental values measured using automatic titrator SiriusT3™ with standard deviation of 19 standards used, their pK_a values taken from the literature, if available, and their calculated values. Rifapentine and saquinavir (*) are analyzed using methanol as cosolvent to promote solvation, while ritonavir (#) was analyzed using UV-metric approach. The red rows are acidic pK_a while the blue rows are basic pK_a

Compound	Exp. pK_a	S.D. (n=3)	Lit. pK_a	ACD pK_a
ACTH fragment 4-10	pK_a1: 3.21	±0.03	n/a	3.32
	pK_a2: 4.26	±0.03	n/a	4.41
	pK_a3: 6.41	±0.01	n/a	6.68
	pK_a4: 7.26	±0.01	n/a	7.20
	pK_a5: >12	n/a	n/a	~13
Angiotensin II	pK_a1: 3.52	±0.01	n/a	3.56
	pK_a2: 4.23	±0.01	n/a	3.71
	pK_a3: 9.88	±0.01	n/a	9.82
	pK_a4: 5.93	±0.01	n/a	6.71
	pK_a5: 7.54	±0.01	n/a	7.54
	pK_a6: >12	n/a	n/a	~13
Caspofungin diacetate	pK_a1: 10.14	±0.06	n/a	9.86
	pK_a2: 3.53	±0.04	n/a	4.40
	pK_a3: 9.10	±0.13	n/a	9.49
	pK_a4: 8.89	±0.13	n/a	9.33
Cetrorelix acetate	pK_a1: 4.01	±0.06	n/a	5.06
	pK_a2: >12	n/a	n/a	~13
	pK_a3: 9.68	±0.02	n/a	9.82
Desmopressin	pK_a1: 9.70	±0.01	n/a	9.90
	pK_a2: >12	n/a	n/a	~13
Erythromycin	pK_a1: 8.86	±0.01	8.9 [70]	8.16
Everolimus	n/a	n/a	n/a	10.40
Goserelin	pK_a1: 4.30	±0.01	n/a	6.70
	pK_a2: >12	n/a	n/a	~13
	pK_a3: 6.02	±0.02	n/a	9.82
Lanreotide acetate	pK_a1: 4.08	±0.03	n/a	6.77
	pK_a2: 5.60	±0.07	n/a	10.43
	pK_a3: 9.28	±0.04	n/a	9.90
Octreotide acetate	pK_a1: 7.00	±0.06	7 [71]	6.77
	pK_a2: 10.44	±0.05	10.5 [71]	10.43
Oxytocin	pK_a1: 4.46	±0.04	n/a	6.14
	pK_a2: 6.27	±0.03	n/a	9.90
Rifampicin	pK_a1: 2.91	±0.02	2.97 [72]	4.96
	pK_a2: 7.60	±0.01	7.50 [72]	7.35
*Rifapentine**	pK_a1: 3.04	±0.17	n/a	4.97
	pK_a2: 7.13	±0.17	n/a	7.71
Rifaximin	pK_a1: 8.85	±0.03	n/a	7.89
	pK_a2: 3.61	±0.05	n/a	5.55
Ritonavir#	pK_a1: 2.31	±0.004	1.9 - 2.5 [73]	2.01 - 2.51

Roxithromycin	pK_a1: 8.89	±0.01	9.2 [74]	8.16
*Saquinavir mesylate**	pK_a1: 6.98	±0.07	7 - 5.5 [75]	6.30 - 2.33
Teicoplanin	pK_a1: 3.08	±0.04	n/a	2.88
	pK_a2: 8.68	±0.11	n/a	8.89
	pK_a3: 9.04	±0.16	n/a	8.94
	pK_a4: 9.57	±0.13	n/a	9.02
	pK_a5: 10.64	±0.08	n/a	9.23
	pK_a6: 7.43	±0.03	n/a	7.11
Vancomycin	pK_a1: 2.33	±0.06	2.9 [76]	2.94
	pK_a2: 9.23	±0.04	9.6 [76]	8.88
	pK_a3: 10.31	±0.03	10.5 [76]	9.34
	pK_a4: 11.67	±0.04	11.7 [76]	10.47
	pK_a5: 7.57	±0.01	7.2 [76]	8.05
	pK_a6: 8.68	±0.02	8.6 [76]	8.38

- **ACTH fragment 4-10**: ACTH (adrenocorticotropic hormone) fragment 4-10 is a short peptide consisting of amino acids 4 to 10 of the full ACTH sequence. It is mainly studied for its neurological and behavioral effects rather than its role in adrenal stimulation. This peptide displays five pK_a values: two acidic and three basics. The acidic pK_a values (3.21 and 4.26) correspond to the C-terminal carboxyl group and to the carboxyl side chain of glutamic acid. The basic pK_a values (6.41 and 7.26) are attributed to the imidazole ring of histidine and the N-terminal amino group. Additionally, the guanidino group of arginine exhibits a $pK_a > 12$, beyond the measurable range of the instrument used.

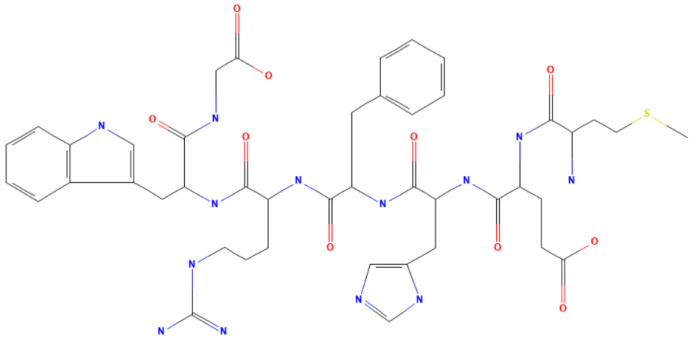


Figure 17. Structure of ACTH fragment 4-10

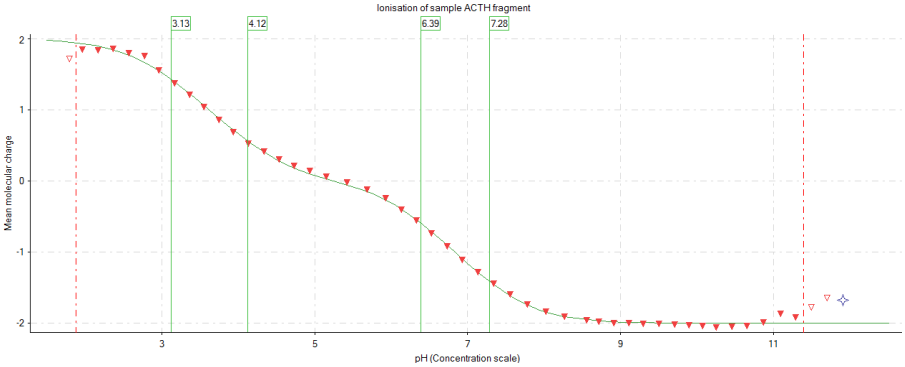


Figure 19. Bjerrum plot that shows the different chemical species as a function of pH of ACTH fragment 4-10

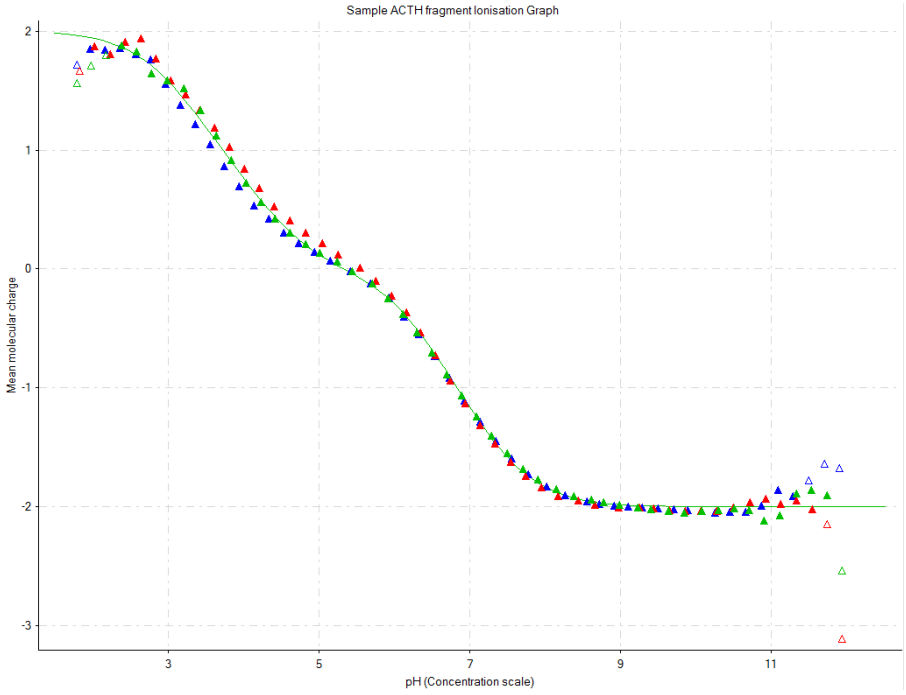


Figure 18. Bjerrum plot of all titrations done for the pK_a evaluation of ACTH fragment 4-10

- **angiotensin II**: is an octapeptide hormone central to the renin-angiotensin-aldosterone system, involved in blood pressure regulation, fluid balance, and vascular tone through vasoconstriction and aldosterone release. It exhibits six pK_a values: three acidic (3.52, 4.23, and 9.88) corresponding to the C-terminal carboxyl group, the carboxyl side chain of aspartate, and the phenol of tyrosine. The basic pK_a values (5.93 and 7.54) are associated with the imidazole group of histidine and the N-terminal amino group. Like other peptides, the guanidino group of arginine has a $pK_a > 12$.

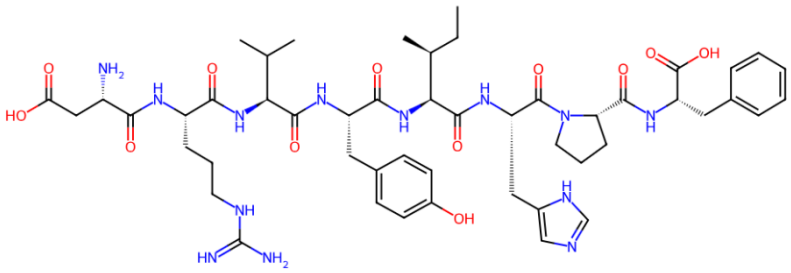


Figure 20. Structure of angiotensin II

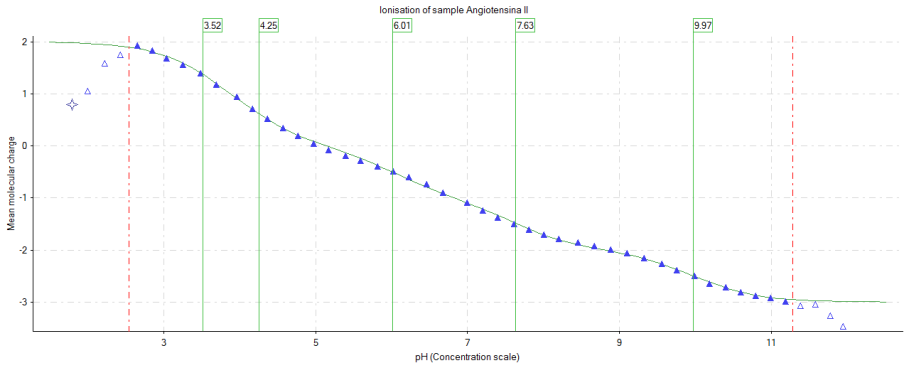


Figure 21. Bjerrum plot that shows the different chemical species as a function of pH of angiotensin II

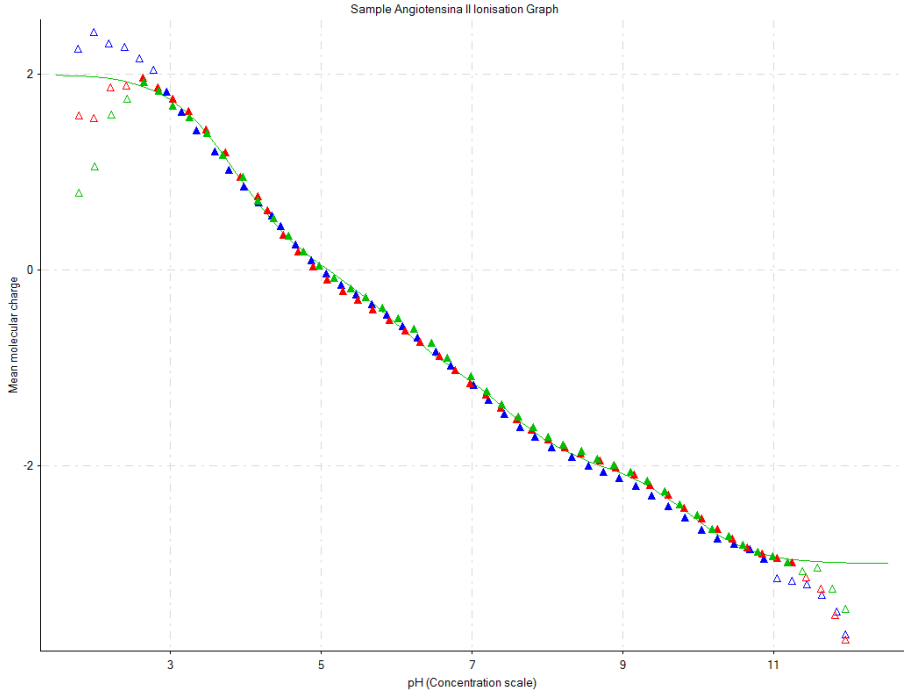


Figure 22. Bjerrum plot of all titrations done for the pK_a evaluation of angiotensin II

- **caspofungin diacetate**: is an antifungal echinocandin that inhibits β-1,3-D-glucan synthase, disrupting fungal cell wall synthesis. It is used for treating invasive aspergillosis and candidiasis. It presents four pK_a values: one acidic (10.14), associated with a tyrosine phenolic group, and three basics (3.53, 8.89, and 9.10) corresponding to primary and secondary amines in side chains.

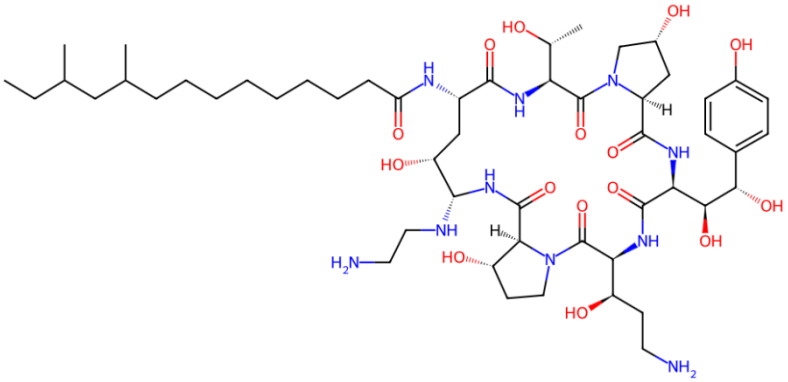


Figure 23. Structure of caspofungin

- **cetrorelix acetate**: is a synthetic decapeptide and potent gonadotropin-releasing hormone (GnRH) antagonist, used in assisted reproduction and certain hormone-sensitive tumors. It shows three pK_a values: one acidic (9.68) linked to the tyrosine phenol, and two basic pK_a values, 4.01 (attributed to a pyridine ring) and one >12 (guanidino group of arginine).

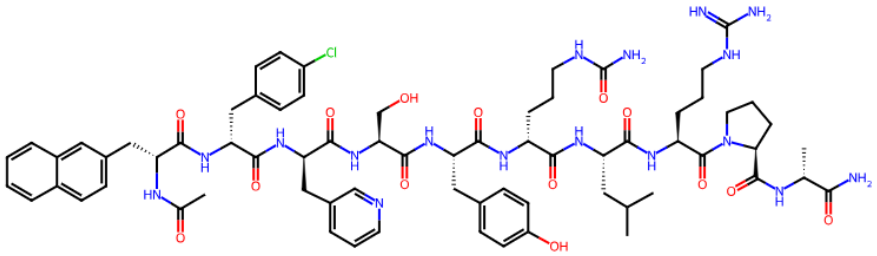


Figure 24. Structure of cetrorelix

- **desmopressin**: it is a synthetic analog of vasopressin, a hormone that regulates water balance in the body. It is primarily used to treat conditions like diabetes, nocturnal enuresis, and bleeding disorder due to its ability to promote water reabsorption in the kidneys and increase clotting factor levels. The main modifications enhance stability and duration of action compared to natural vasopressin. It has two pK_a values: one acidic and one basic. The acidic one (9.70) is associated with a phenol group of tyrosine, and the basic one is associated with guanidine group of arginine, greater than 12 and beyond the instrument measurement range.

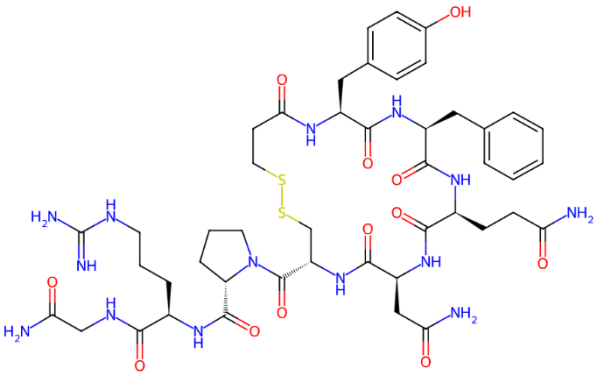


Figure 25. Structure of desmopressin

- **erythromycin**: is a macrolide antibiotic that inhibits bacterial protein synthesis. It has a single basic pK_a (8.86), corresponding to a dimethylamino group in the structure. The pK_a found in literature (8.9) is consistent with the measured pK_a [70].

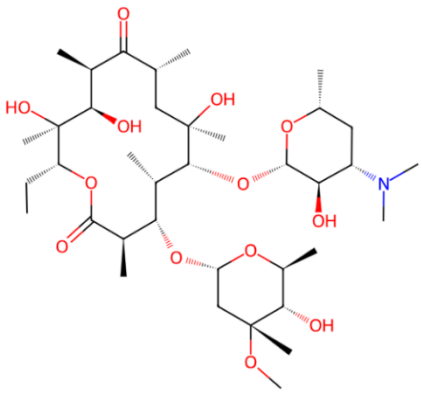


Figure 26. Structure of erythromycin

- **everolimus**: is a synthetic derivative of sirolimus used as an immunosuppressant and antineoplastic agent. It lacks important ionizable functional groups, and it is therefore considered neutral. ACD Percepta reports a pKa value, associated with the hydroxyl group, of 10.40.

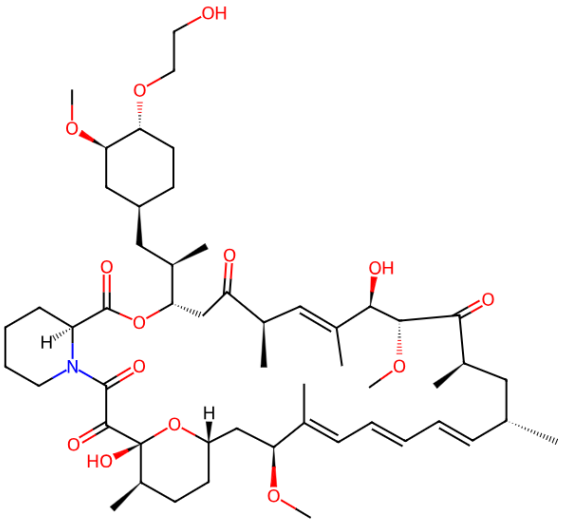


Figure 27. Structure of everolimus

- **goserelin**: is a synthetic GnRH agonist used in hormone-sensitive cancers and endometriosis. It has three pK_a values: one acidic (6.02, tyrosine phenol) and two basics (4.30, imidazole group of histidine; and >12, guanidino group of arginine). The values for tyrosine and histidine are lower than typically expected from *in-silico* pK_a.

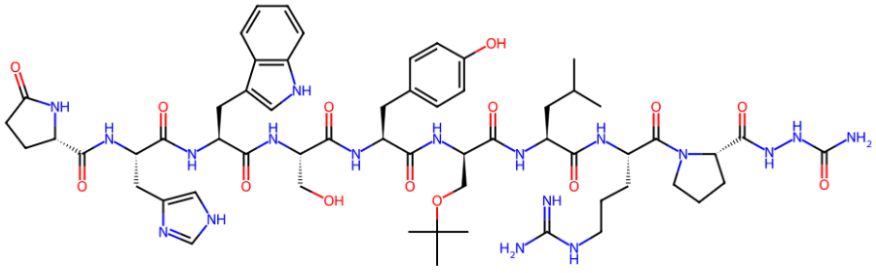


Figure 28. Structure of gosereline

- **lanreotide acetate**: is a cyclic octapeptide and somatostatin analog, used in acromegaly and neuroendocrine tumors. It contains a disulfide bridge. It exhibits three pK_a values: one acidic (9.28, tyrosine phenol) and two basic (4.08 and 5.60), associated with side-chain amino groups and lysine. The basic pK_a of lanreotide is lower than typically expected from *in-silico* data.

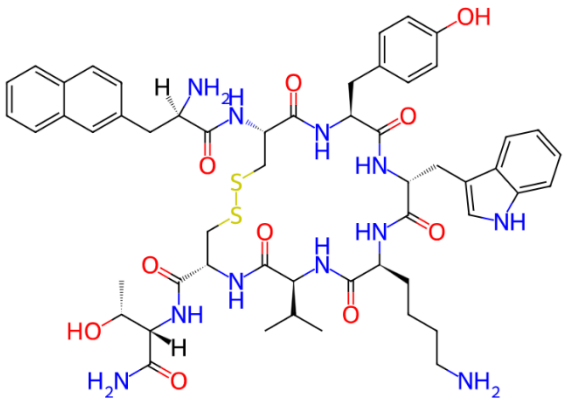


Figure 29. Structure of lanreotide

- **octreotide acetate**: octreotide, structurally similar to lanreotide, is also a somatostatin analog used for the same clinical indications. It displays two basic pK_a values (7.00 and 10.44), both associated with primary amines, including one from lysine. The measured pK_a are aligned with the pK_a found in literature (7.0 and 10.5) [71].

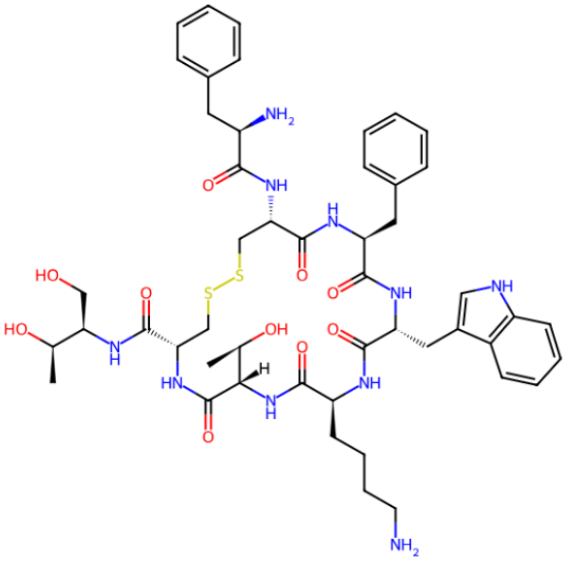


Figure 30. Structure of octreotide

- **oxytocin**: is a cyclic nonapeptide hormone involved in childbirth, lactation, and social bonding. It shows two pK_a values: an acidic one (6.27, tyrosine phenol) and a basic one (4.46, cysteine amino group). Both values are lower than expected from *in-silico* data.

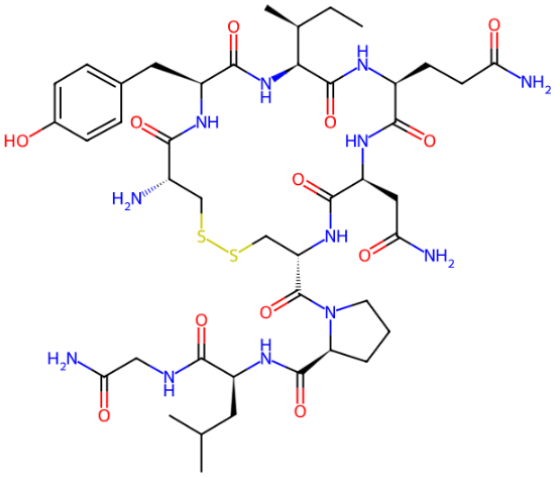


Figure 31. Structure of oxytocin

- **rifampicin**: it is a broad-spectrum antibiotic primarily used to treat tuberculosis, leprosy, and certain bacterial infections including those caused by Gram-positive and negative bacteria. It is a semisynthetic derivative of rifamycin B. It has two pK_a values: 2.91 (phenolic group) and 7.60 (piperazine nitrogen). It has abnormally low pK_a because it is activated by an extended conjugate system and electron-attractor groups that strongly stabilize the phenoxide anion. The measured pK_a are aligned with those found in literature (2.97 and 7.50) [72].

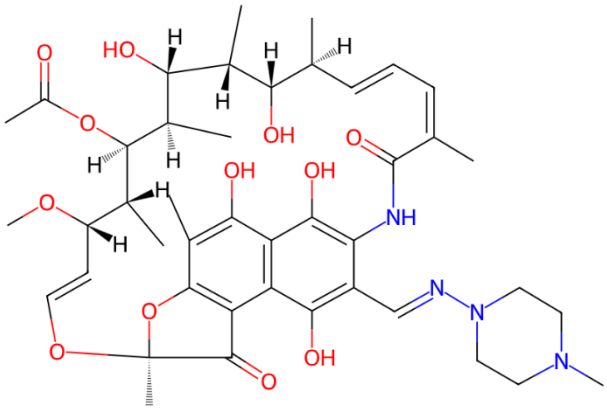


Figure 32. Structure of rifampicin

- **rifapentine**: it is a semisynthetic antibiotic of the rifamycin class, structurally related to rifampin. It is used primarily in the treatment of tuberculosis, in combination therapy. Like rifampicin, it has two pK_a values: 3.18 (phenolic group) and 8.32 (piperazine group). Even in this compound the phenolic one is lower than expected from *in-silico* data. For this molecule methanol as cosolvent has been used to promote the solvation of rifapentine.

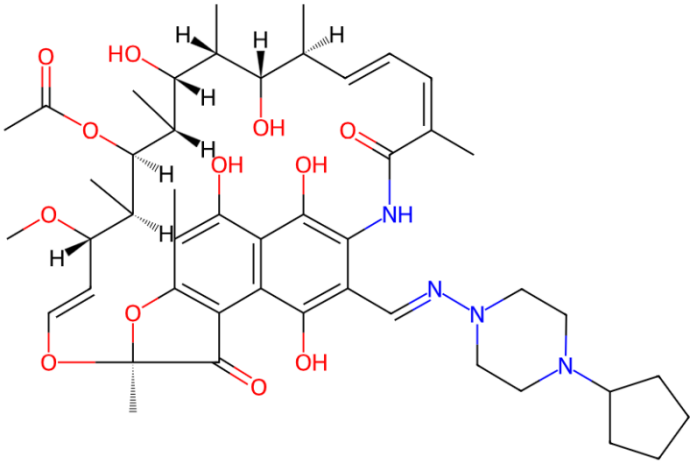


Figure 33. Structure of rifapentine

- **rifaximin**: also this compound is a rifamycin derivate. It is poorly absorbed in the gastrointestinal tract, which makes it ideal for local treatment in the gut (traveler's diarrhea and IBS with diarrhea). It has two pK_a values: 3.61 (phenol) and 8.85 (N of piperidinic ring). In this compound too the phenolic one is lower than expected.

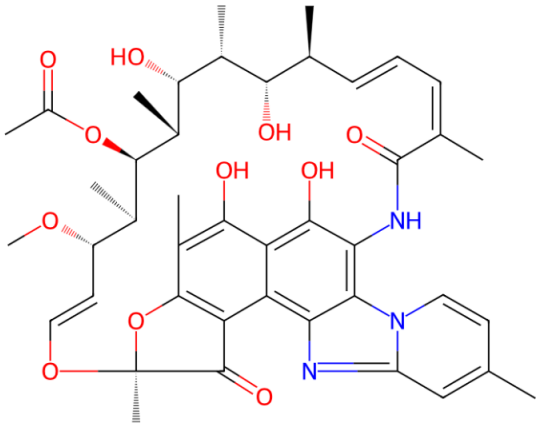


Figure 34. Structure of rifaximin

- **ritonavir**: it is an antiviral compound, used for the treatment and boosting or other antiretrovirals in HIV therapy. It has two thiazole groups (very weak base), with a pK_a value of 2.31, too close to the minimum pH measurable by the instrument (1.8-2). The pK_a of this compound has been measured using UV-metric technique due to the presence of chromophore close to the ionizable group. This value is aligned with that found in literature when two pK_a are associated to ritonavir (1.9 and 2.5) [73].

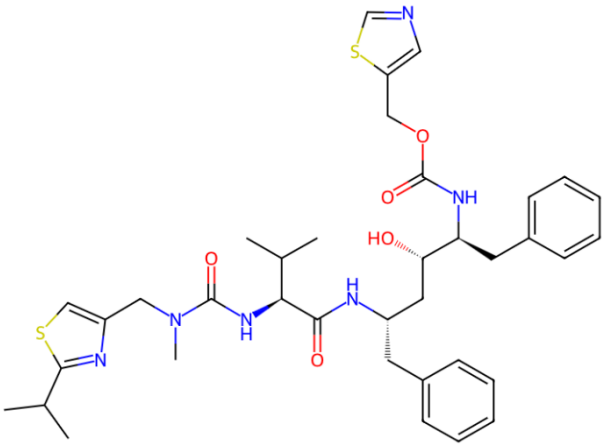


Figure 35. Structure of ritonavir

- **roxithromycin**: it is a macrolide antibiotic that inhibits bacterial protein synthesis. It is primarily bacteriostatic but may be bactericidal at high concentrations. It has one basic pK_a of 8.89 associated with a tertiary ammine group in the structure. This pK_a is consistent with the value found in literature (9.2) [74].

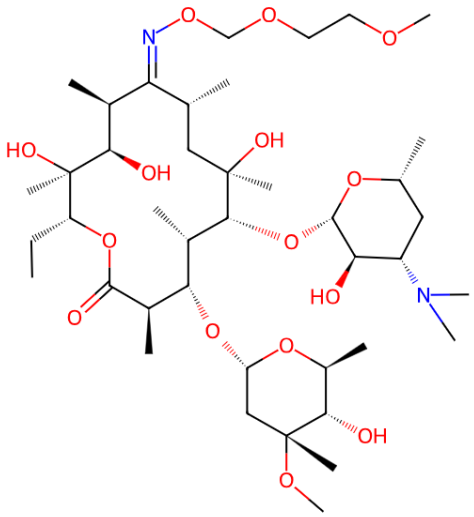


Figure 36. Structure of roxithromycin

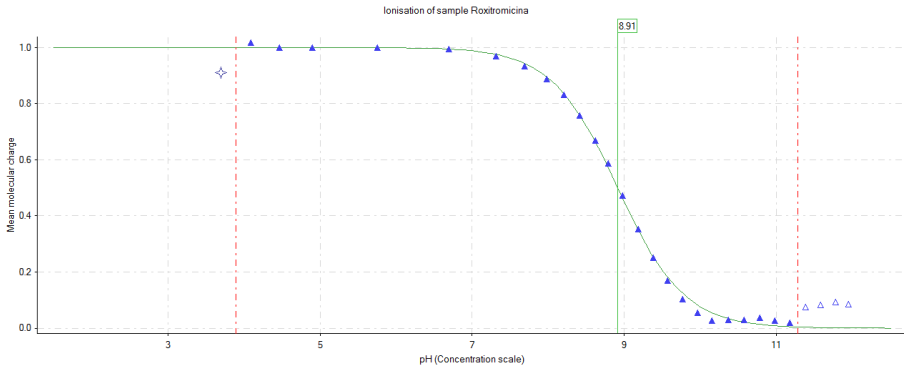


Figure 38. Bjerrum plot that shows the different chemical species as a function of pH of roxithromycin

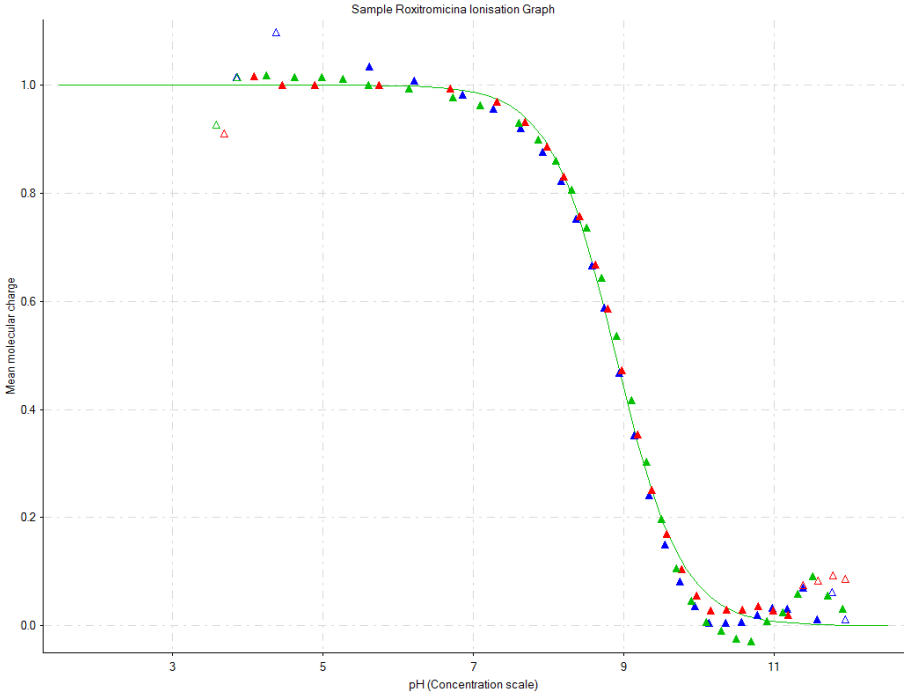


Figure 37. Bjerrum plot of all titrations done for the pK_a evaluation of roxithromycin

- **saquinavir mesylate:** it is an important antiviral, HIV protease inhibitor. It has two basic pK_a values: 6.98 (tertiary amine) and one unmeasured due to proximity to the measurable pH range, associated with an N-quinoline nitrogen. ThepK_a value associated with tertiary amine is in line with that measured with Sirius^{TM} (7). The other pK_a associated with quinoline group in the article found in literature results equal to 5.5 [75]. As for rifapentine, methanol as cosolvent has been used to promote the solvation of saquinavir.

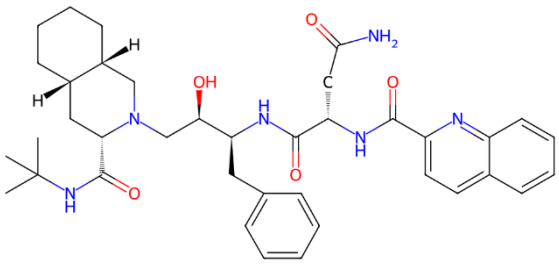


Figure 39. Structure of saquinavir

- **teicoplanin A_2-2**: it is a glycopeptide antibiotic used primarily to treat Gram-positive infection, particularly those caused by methicillin-resistant Staphylococcus aureus (MRSA). It works by inhibiting bacterial cell wall synthesis, specifically targeting peptidoglycan. It has a lot of ionizable groups, five acidic and one basic. The acidic ones are all related to phenolic groups in the structure, 8.68, 9.04, 9.57 and 10.64. The other acidic group is the carboxylic group with a value of 3.08. There is one basic pK_a associated with primary amino group, 7.43. The single basic pK_a, measured at 7.43, is attributed to a primary amino group, likely involved in electrostatic interactions, and possibly influencing the molecule's binding to bacterial targets. The combination of multiple phenolic acidic sites and a single basic amine gives teicoplanin amphoteric properties and allows it to exist in different ionization states depending on the pH. This characteristic is particularly relevant for formulation, solubility, and tissue distribution in clinical use.

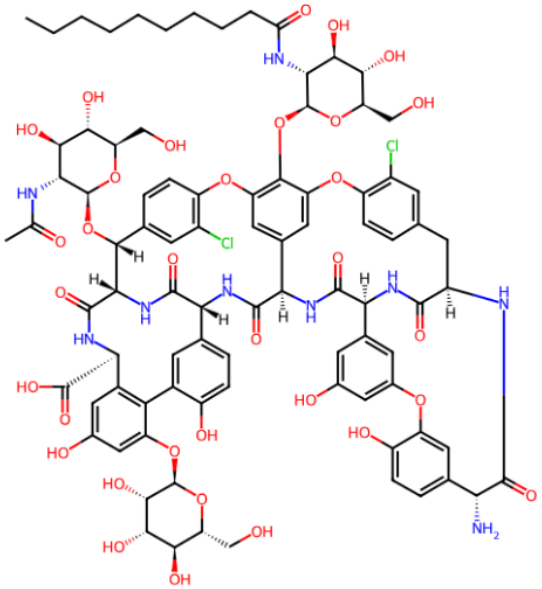


Figure 40. Structure of teicoplanin

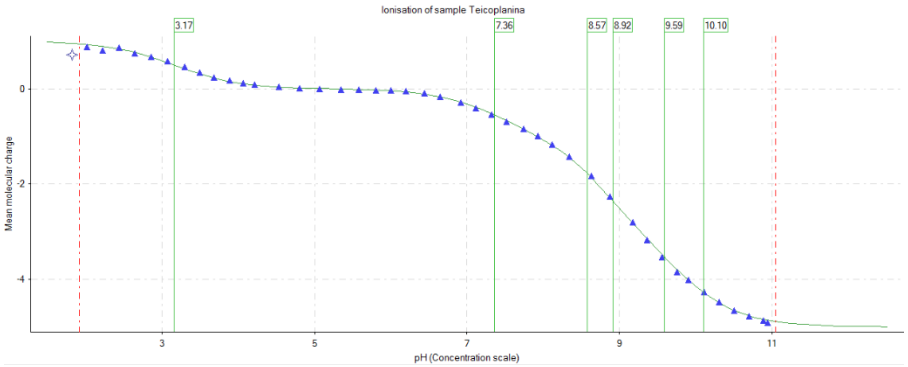


Figure 42. Bjerrum plot that shows the different chemical species as a function of pH of teicoplanin

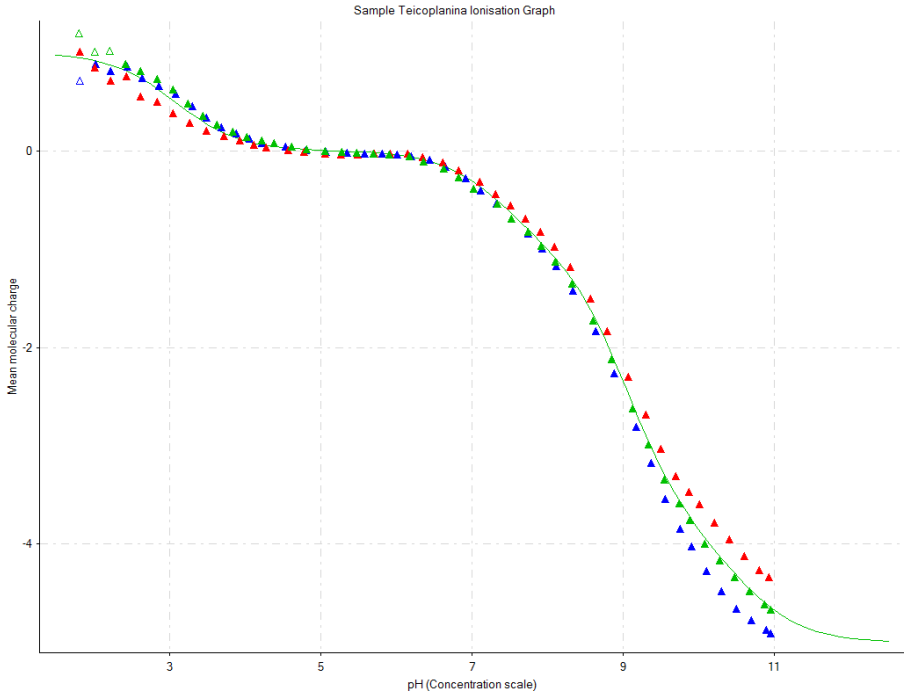


Figure 41. Bjerrum plot of all titrations done for the pK_a evaluation of teicoplanin

- **vancomycin**: it is a glycopeptide antibiotic used primarily to treat serious Gram-positive bacterial infections, particularly those resistant to other antibiotics like methicillin-resistant Staphylococcus aureus (MRSA). It exhibits a complex ionization profile, with six measurable pK_a values: four acidic and two basics. The acidic pK_a values include one carboxylic acid group at 2.33, and three phenolic hydroxyl groups at 9.23, 10.31, and 11.67, reflecting the polyphenolic nature of its structure. These groups are important for hydrogen bonding, solubility, and binding to target sites in bacterial peptidoglycan. The basic pK_a values are attributed to a primary amine group at 7.57 and a secondary amine group at 8.68, which can contribute to the molecule's cationic nature at physiological pH. Vancomycin's amphoteric character enables it to adopt multiple ionization states depending on pH, which supports

aqueous solubility under physiological conditions and promotes electrostatic interactions with the negatively charged components of bacterial cell walls. The broad range of pK_a values across both acidic and basic sites highlights vancomycin's versatile ionization behavior, which is essential for its bioavailability, tissue distribution, and formulation properties, especially in intravenous administration where aqueous solubility is critical.

It is interesting to notice that all the pK_a values found in literature are consistent with those measured using SiriusT3^{TM} (2.9, 9.6, 10.5, 11.7, 7.2, 8.6) [76].

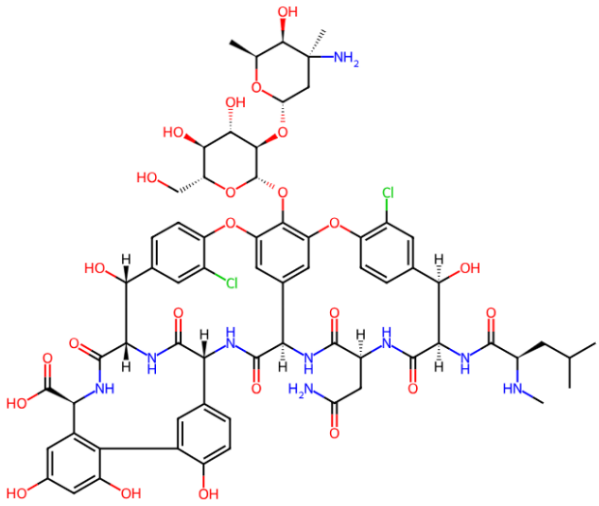


Figure 43. Structure of vancomycin

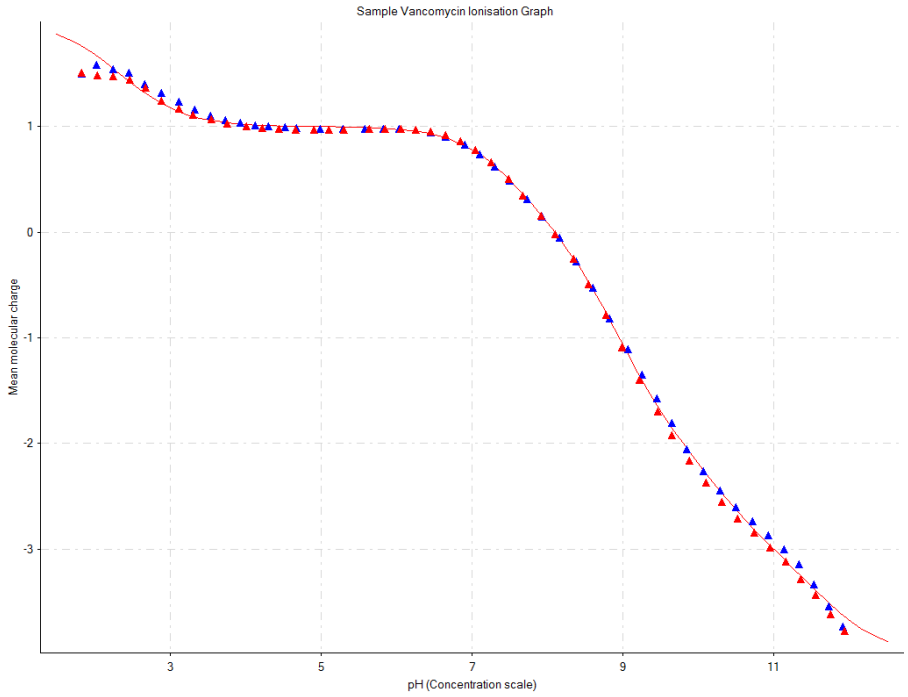


Figure 44. Bjerrum plot of all titrations done for the pK_a evaluation of vancomycin

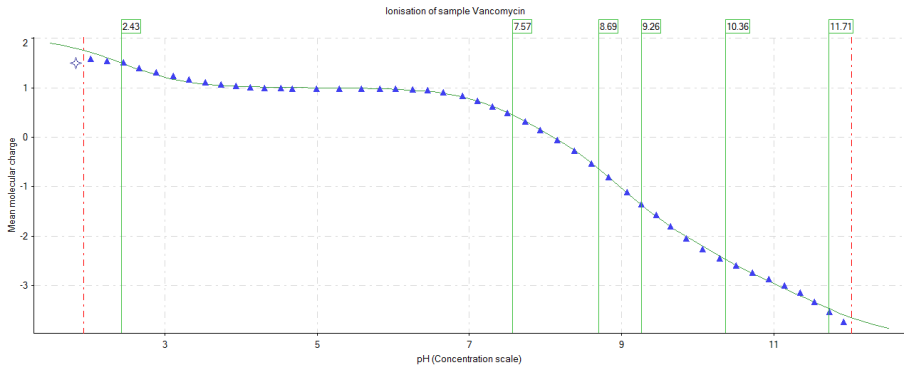


Figure 45. Bjerrum plot that shows the different chemical species as a function of pH of vancomycin

In general, the experimentally determined pK_a values are in good agreement with those predicted using Percepta 2019 software. However, six compounds (i.e.goserelin, octreotide, oxytocin, rifampicin, rifapentine, and rifaximin) showed notable discrepancies between the experimental and predicted values. For the first three compounds (two peptides and one peptidomimetic), all experimentally measured pK_a values differed from the predicted ones, highlighting the difficulty of accurately modeling ionization behavior in structurally complex molecules. In the case of rifampicin, rifapentine, and rifaximin, the divergence was limited to the weakly acidic pK_a values associated with phenolic groups. Specifically for rifampicin, a study by Caron et al. reported experimental pK_a values that are consistent with those obtained in this work, particularly for the phenolic proton, thus supporting the robustness of the present data [72]. Given the structural similarity among rifampicin, rifapentine, and rifaximin, the phenolic pK_a values of the latter two were assumed to be comparable to that of rifampicin. These findings reinforce the idea that predicted pK_a values may not always be reliable for complex or highly functionalized molecules, and they underline the importance of experimental validation when accurate ionization profiles are required for further physico-chemical or pharmacokinetic modeling.

For seven compounds with available literature pK_a data, experimentally determined values are consistent with these references, highlighting the reliability of experimental pH-metric measurements instead of *in-silico* prediction.

7.5 Conclusions

The evaluation of dissociation constants (pK_a) represents one of the earliest and most critical steps in the drug discovery process, due to its fundamental role in understanding the ADME properties of candidate molecules and guiding the selection of appropriate formulation strategies. Accurate determination of pK_a values allows for a more reliable prediction of key pharmacokinetic parameters

such as aqueous solubility, membrane permeability, and oral bioavailability, thereby supporting the selection and rational optimization of compounds with favorable drug-like profiles.

In this PhD project, the pK_a values and full dissociation profiles of a chemically diverse set of compounds were experimentally determined using the potentiometric and UV-metric titration method with the SiriusT3™ instrument. A particularly notable result was the successful characterization of all selected compounds, including those with complex chemical structures and multiple ionizable groups. Although potentiometric titration is conventionally optimized for small molecules with limited ionization sites, this study demonstrated that the SiriusT3™ system also delivers robust and reproducible results for peptides, macrocyclic compounds, and other structurally intricate molecules, provided that solubility and titration parameters are carefully controlled.

The experimentally determined pK_a values showed good overall consistency with those predicted by computational software, but with some exceptions, confirming the validity of *in-silico* models while emphasizing the importance of experimental validation, especially in the case of compounds with unusual functional groups or sterically hindered ionization sites, where predictive algorithms may have limited accuracy. Importantly, acquiring experimentally validated pK_a values also proved to be strategically valuable for the subsequent potentiometric determination of lipophilicity parameters (Log *P* and Log *D*), particularly when carried out using the same SiriusT3™ platform. Since the accuracy of Log *D* measurements across different pH values depends on reliable knowledge of the analyte's ionization profile, integrating pK_a and lipophilicity assessments ensures internal consistency and physico-chemical coherence across experimental datasets. This is especially critical for multi-ionizable compounds, where distribution behavior between aqueous and lipid phases is highly dependent on pH**.**

In summary, the integration of experimental pK_a determination within the broader physico-chemical profiling workflow provides a solid foundation for ADME prediction**,** formulation development, and compound prioritization during the early stages of drug development, while also providing a benchmark for evaluating and improving computational prediction models.

8 Experimental section - Lipophilicity

8.1 Aim

The primary objective of this section is to investigate and compare several experimental approaches used to evaluate lipophilicity, which remains one of the most critical physico-chemical properties assessed during the early phases of drug discovery and development. As briefly introduced in the opening section of this thesis, lipophilicity plays a fundamental role in determining a compound's ADMET profile, affecting parameters such as passive membrane permeability, metabolic stability, plasma protein binding, tissue distribution, clearance, and potential toxicity.

This section specifically focuses on a dataset composed of structurally diverse and highly complex molecules, most of which fall within the bRo5 chemical space. These compounds are characterized by high molecular weight, increased hydrogen bonding capacity, and often display amphiphilic behavior and poor solubility features that pose unique challenges for the accurate determination of lipophilicity using conventional methods.

A secondary aim of this work is to critically assess whether the current gold standard methodologies, such as the *shake-flask* method for Log $D_{7.4}$ and the potentiometric technique for Log P_{oct}, remain reliable and applicable when used on bRo5 compounds. Given the limitations of these traditional approaches in handling poorly soluble or highly amphiphilic molecules, there is a need to explore alternative techniques capable of providing accurate and reproducible lipophilicity data.
In this context, the utility and predictive power of chromatographic methods, particularly the measurement of CHI Log $D_{7.4}$, will be thoroughly evaluated. The experimental design involves comparing chromatographic results with gold standard values and passive permeability data obtained using Caco-2 cell monolayers, a widely accepted *in-vitro* model of intestinal absorption. The comparison aims to assess how well chromatographic hydrophobicity correlates with biologically relevant properties. It will be discussed in the next chapters.

Ultimately, the goal is to identify a high-throughput, reliable, and operationally robust method for assessing lipophilicity in structurally complex, bRo5-type compounds. Overcoming the solubility and amphiphilicity-related limitations that typically hinder standard methods could significantly streamline the compound triaging process during the screening cascade. Furthermore, it would enhance the characterization of this emerging class of drug candidates, whose relevance in modern pharmaceutical pipelines continues to grow in response to increasingly complex therapeutic targets.

In this work, lipophilicity has been assessed using three complementary methods:

- a potentiometric approach for Log P_{oct}
- the shake-flask method for determining Log $D_{7.4}$
- a chromatographic technique for CHI Log $D_{7.4}$

Since ionization state strongly affects lipophilicity, accurate pK_a determination is essential, particularly for potentiometric measurements. However, knowledge of pK_a is also valuable for interpreting data obtained from other methods, as it provides insight into the molecule's ionization state under physiological pH conditions. For compounds lacking ionizable groups, Log $D_{7.4}$ is equivalent to Log P_{oct}, underscoring the interdependence between these parameters and the importance of selecting the appropriate technique.

8.2 Materials

8.2.1 Materials used to measure the lipophilicity using SiriusT3^{TM} automatic titrator

To measure lipophilicity expressed as Log $P_{oct/water}$ and Log $P_{tol/water}$, of the 19 pharmaceutical standard, the following chemical reagents were used: dimethylsulfoxide (Sigma-Merck), Titrisol unit-dose (Sigma-Merck) used to prepare a solution of hydrochloric acid 0.5 N, Titrisol unit-dose (Sigma-Merck) used to prepare a solution of potassium hydroxide 0.5 N, potassium dihydrogen phosphate (Sigma-Merck), methanol 80% w/v, *n*-octanol (Sigma-Merck) and toluene (Sigma-Merck).
The list of the 19 compounds analyzed is shown in Table 1.

The following instruments and accessories were used: SiriusT3^{TM} automatic titrator (Sirius Analytical Instrument Ltd), SiriusT3Control software, SiriusT3Refine for data analysis, analytical balance (Mettler-toledo S.r.l. characterized by a sensibility of 0.1 mg and with a capacity of 100 mg), 1.5 mL Eppendorf vials (Eppendorf), weighting vessels (Sigma-Merck), pipettes P10, P200, P1000 (Gilson), smart spatula (Sigma-Merck), 6 mL vials and Vortex.

8.2.2 Materials used to lipophilicity evaluation through the shake-flask assay

To measure the lipophilicity using the *shake-flask* assay, expressed as Log $D_{7.4}$ or Log $P_{oct/water}$ and Log $P_{tol/wat}$, of the 19 pharmaceutical standard, the following chemical reagents were used: *n*-octanol (Sigma-Aldrich), toluene (Sigma-Aldrich), methanol (Sigma-Aldrich), 3-(*n*-morpholino)propanesulfonic acid sodium salt (MOPS) (Sigma-Aldrich), 3-cyclohexylamino-1-propanesulfonic acid sodium salt (CAPS) (Sigma-Aldrich), water LC-MS grade (Sigma-Aldrich),

acetonitrile LC-MS grade (Sigma-Aldrich), ammonium formate (Sigma-Aldrich), formic acid (Fisher chemical).

The following instruments and accessories were used: analytical balance (Mettler-toledo S.r.l. characterized by a sensibility of 0.1 mg and with a capacity of 100 mg), 1.5 mL Eppendorf vials (Eppendorf), tube revolver (Sigma-Merck), electronic pipettes (Sartorius 5-120 µL, 10-300 µL, 100-5000 µL), smart spatula (Sigma-Merck), 1.5 mL vials, 4 mL vials, ultrasonic bath (Branson 5510), vortex, columns Kinetex C8 100 x 2.1 - 1.7µm (Phenomenex) and Acquity CSH C18 50 x 2.10 - 1.7 µm (Waters), and Acquity UPLC (Waters) coupled to TQabsolute mass spectrometer (Waters).

8.2.3 Material used for the CHI Log $D_{7.4}$ evaluation

For the Chromatographic Hydrophobicity Index (CHI) Log *D* evaluation performed at pH 7.4, 19 pharmaceutical standards were used as 10mM DMSO stock solution.

In accordance with the literature, a set of 10 calibration standards are required [77], [78]: theophylline (Sigma-Aldrich), 5-phenyl-1-tetrazole (Sigma-Aldrich), benzimidazole (Sigma-Aldrich), colchicine (Sigma-Aldrich), 8-phenyltheophylline (Sigma-Aldrich), acetophenone (Sigma-Aldrich), indole (Sigma-Aldrich), propiophenone (Sigma-Aldrich), butyrophenone (Sigma-Aldrich) and valerophenone (Sigma-Aldrich).

In addition to the calibration standards, the following substances were used: ammonium acetate (Sigma-Aldrich), ammonia solution (Sigma-Aldrich), water LC-MS grade (Sigma-Aldrich), acetonitrile LC-MS grade (Sigma-Aldrich), dimethyl sulfoxide (Sigma-Aldrich).

The following instruments and accessories were used: analytical balance (Mettler Toledo; sensitivity 0.001 mg; capacity 52 g, minimum weighing 2 mg), fume hood, electronic pipettes (Sartorius 10-300 µL, 100-5000 µL), Acquity UPLC (Waters) coupled to SQ mass spectrometer detector (Waters), column Luna C-18 30*2 mm - 3µm particle size 100 Å (Phenomenex), ultrasonic bath (Branson 5510), smart spatula (Sigma Aldrich), vial with 200 µL insert, 96-well plate (Phenomenex), 4 mL vial, vortex.

8.3 Methods

8.3.1 Log P_{oct} and Log P_{tol} evaluation using the SiriusT3^{TM} automatic titrator

On the basis of its lipophilicity, a compound could have a different partition in a biphasic system composed normally by *n*-octanol and aqueous buffer solution, immiscible each other.

Using the SiriusT3^{TM} automatic titrator to measure the partition coefficient, expressed as Log P_{oct} or Log P_{tol} on the basis of the lipophilic phase used, 1-2 mg of each compound is weighted in a 4 mL vial and put in the sample place of the instrument. Before starting, the instrument needs some

information, such as the range of the pH of the titration, compound information (MW, amount, pK_a and nature of it, presence of counterion and the expected lipophilicity of the compound). Three titrations were carried out, each of them with a different volumetric ratio between *n*-octanolic and aqueous phase. The volumetric ratios depend on the expected lipophilicity of the compound.
The partition of the compound in the organic solvent shifts the flex of the titration curve (pK_a) to higher pH value for acid compounds and lower for basic compounds. The Log P_{oct} value can be determined from this shift and the volumetric ratio using the Bjerrum plot titration curve analysis. Further refinement of the Log P_{oct} value was performed with Sirius T3^{TM} Refine v.1.1.3.0 software.
The Bjerrum plot analysis and iterative refinement algorithm employed for pK_a determination have been described in detail in the introduction section at p.19, including the mathematical formulation of mean molecular charge calculation, sum of squared residuals minimization, and Goodness of Fit convergence criteria. The information has been taken from SiriusT3^{TM} manual.

8.3.2 Log $D_{7.4}$ evaluation using the shake-flask assay

Distribution coefficient (*D*) was determined by shake-flask method at room temperature (21.0 ± 0.5°C), employing a biphasic system made of *n*-octanol and 3-(morpholin-4-yl)propane-1-sulfonic acid (MOPS) 50 mM for pH 7.4, or 3-cyclohexylamino-1-propanesulfonic acid sodium salt (CAPS) for pH 9.3, with ionic strength (μ) adjusted to 0.15 M with KCl. The organic phase and the aqueous buffer solution were mutually saturated by overnight stirring. 1 mg of solid API was accurately weighed and dissolved in 1 mL of the most suitable phase, according to the expected lipophilicity, in order to obtain a 1 mg/mL solution. If solubilization was incomplete, after sonication and gentle heating, the suspension was centrifuged (1000 g, RT, 5 min) and the supernatant was employed for samples. The solution was split in four centrifuge tubes, and the other phase was added according to three different *n*-octanol/water ratios, i.e. 200/200 (ratio = 1), 150/300 (ratio = 0.5) and 100/400 (ratio = 0.25); each ratio was tested as duplicates (n = 6 samples in total). Biphasic samples were stirred overnight or, at least, for 6/8 hours to reach partition equilibrium. Samples were centrifuged (1000g, RT, 10 min) to ensure phase separation, then the two phases were manually split and diluted with methanol before analysis by LC-UV/MS.
Two chromatographic methods using two different columns (C8 and C18) have been used to analyze the concentration of the compounds in each phase, on the basis of their retention time. The methods are as follows:

- Kinetex C8 column 100mm x 2.1mm 1.7μm, kept at 55°C; Mobile Phase: Eluent B (ACN + 0.1% HCOOH) in eluent A ($HCOONH_4$ 0.025M pH3) from 1% to 30% within 3.00 min, from 30% to 50% in 3.50 min until 6.50 min, from 50% to 80% in 1 min until 7.50 min, held at

80% until 8.00 min, returned to 1% at 8.10 min and maintained at 1% until 10.00 min. Flow rate: 0.500 mL/min. Wavelength: 210-400 nm DAD. Waters acquitiy QSM + Waters PDA + Waters XEVO TQabsolute with acquisition parameters as follows: fullscan centroid as acquisition mode, source temperature 150°C, desolvation temperature 250°C, cone gas flow 250 L/Hr, desolvation gas flow 1000 L/Hr, mass range 150-1500, polarity ES+ (capillary voltage 1.8kV, cone voltage 18V) and polarity ES- (capillary voltage 1.5kV, cone voltage 30V).

- Acquity UPLC CSH C18 column 50mm x 2.1mm 1.7µm, kept at 50°C; Mobile Phase: Eluent B (ACN + 0.1% HCOOH) in Eluent A ($HCOONH_4$ 0.025M pH3) from 20% to 80% within 5.50 min, maintained at 80% for 2 min until 7.50 min, returned to 20% from 7.50 min to 8.00 min and maintained at 20% until 10 min. Flow rate: 0.350 mL/min. Wavelength: 210-400 nm DAD. Waters acquity QSM + Waters PDA + Waters XEVO TQabsolute with acquisition parameters as follows: fullscan centroid as acquisition mode, source temperature 150°C, desolvation temperature 250°C, cone gas flow 250 L/Hr, desolvation gas flow 1000 L/Hr, mass range 150-1500, polarity ES+ (capillary voltage 1.8kV, cone voltage 18V) and polarity ES- (capillary voltage 1.5kV, cone voltage 30V).

Resulting Log $D_{7.4}$ are reported as a means of six independent samples ± SD.

As mentioned above, the shake-flask method requires overnight stirring (12-18 hours at 23°C ± 2°C). Precautions implemented:

- Room temperature (23°C ± 2°C) to avoid thermal degradation
- *n*-octanol/aqueous buffer naturally buffers pH, minimizing hydrolysis
- Nitrogen purging of aqueous phase to prevent oxidation
- Immediate phase separation and HPLC-UV analysis within 2 hours

No chromatographic evidence of degradation was observed during HPLC analysis of shake-flask samples (single peaks and consistent retention times). This confirms chemical stability during 12-18h equilibration at 23°C ± 2°C.

8.3.3 CHI Log $D_{7.4}$ evaluation

For the evaluation of the Chromatographic Hydrophobicity Index Log $D_{7.4}$ (CHI Log *D*), the method developed by Clara Valko et al. was used with a new and shorter column (Luna C18, 50 x 4.6 mm, 5 µm, 100 Å vs Luna C18, 30 x 2 mm, 3 µm, 100 Å - Phenomenex) [33], [34], [39]. In the work of

Bunally et al. and Young et al. another C18 column was used (Gemini NX C18, 50 x 2 mm, 3 µm, 100 Å - Phenomenex) [42], [78].

A 10mM stock solution in dimethyl sulfoxide (DMSO) of each compound was used. An UPLC system (Waters Acquity UPLC) coupled with mass spectrometer (Waters SQ detector) was used as instrument with mass parameters as follows: fullscan centroid as acquisition mode, source temperature 150°C, desolvation temperature 350°C, cone gas flow 50 L/Hr, desolvation gas flow 750 L/Hr, mass range 150-1500, polarity ES+/- (capillary voltage 3kV, cone voltage 30V). The Luna C18 30x2 mm with 3µm particle size 100 Å (Phenomenex) was used as chromatographic column. 10mM ammonium acetate at pH 7.4, adjusted with ammonia solutions, and acetonitrile were used as mobile phases in gradient elution. The flow rate was settled at 0.5 mL/min. The column temperature was kept at 40˚C and the time of the analysis was 2 minutes.

Table 3. Chormatographic gradient used to performed CHI analysis for chromatographic lipophiliicity evaluation

Time (min)	Ammonium acetate 10mM (%)	Acetonitrile (%)
0.00	100	0
1.30	10	90
1.53	10	90
1.60	100	0
2	100	0

8.4 Results

This section presents and discusses the lipophilicity profiles of the compounds included in the dataset, assessed using three different experimental methodologies: pH-metric Log P_{oct}, shake-flask Log $D_{7.4}$, and chromatographic CHI Log $D_{7.4}$. The primary goal is to evaluate the degree of correlation among these techniques, to identify their methodological strengths and limitations, and to determine the applicability of each method to structurally complex compounds, including those categorized as bRo5. In the dataset three endogenous peptides, one lipo- and two glycopeptides, seven peptidomimetic compounds and six macrocycles are present. Natural peptides are generally characterized by hydrophilicity and high solubility, while macrocycles and synthetic peptides by usually medium or high lipophilicity and low solubility.

8.4.1 pH-metric Log P, CLog P and XLog P: correlation and discrepancies

Table 4. Log P_{oct} measured by pH-metric technique with corresponding standard deviation, their values taken from the literature when available, CLogP and XLogP calculated using Percepta 2019 algorithm. Compound with "" such as desmopressin, everolimus, ritonavir and saquinavir were evaluated using the shake-flask due to their pK_a. They are neutral at pH of 7.4 and 9.3 for saquinavir. For ritonavir and saquinavir the shake-flask was chosen also due to some issues during the pH-metric analysis. The literature values marked with "#" are those with experimental value found in DrugBank with no specific reference provided. Compounds highlighted in italic type are those characterized by an emulsion formation after the experiment*

Compound	Exp. Log P_{oct}	S.D.	Lit. Log P_{oct}	CLog *P*	XLog *P*
ACTH fragment	0.56	±0.16	NA	0.32	-3.46
Angiotensin II	-7.95	±0.42	NA	2.34	-2.86
Caspofungin diacetate	*1.23*	*±0.04*	*0 #*	*-5.05*	*-2.19*
Cetrorelix acetate	*2.35*	*±0.12*	*NA*	*3.2*	*0.68*
Desmopressin*	-3.2	±0.06	-3.5 [79]	-5.32	-2.48
Erythromycin	2.58	±0.01	3.06 [70]	2.83	2.55
Everolimus*	2.66	±0.38	NA	3.34	7.64
Goserelin	-0.98	±0.01	-2#	-0.45	-1.04
Lanreotide acetate	*0.42*	*±0.26*	*NA*	*2.89*	*2.7*
Octreotide acetate	*2.55*	*±0.01*	*NA*	*0.77*	*0.65*
Oxytocin	-6.33	±0.73	NA	-4.26	-1.64
Rifampicin	0.39	±0.03	0.49 [72]	1.09	4.29
Rifapentine	-0.96	±0.11	4#	2.58	5.76
Rifaximin	2.79	±0.04	2.6#	3.22	7.06
Ritonavir*	3.57	±0.18	4.4 [73]	5.28	6.68
Roxithromycin	3.13	±0.01	2.61 [74]	3.73	2.86
Saquinavir mesilate*	3.85	±0.03	4.1 [75]	4.44	3.6
Teicoplanin	*-5.51*	*±0.85*	*NA*	*0.4*	*-2.51*
Vancomycin	*-0.72*	*±0.13*	*-3.1#*	*-1.44*	*-3.15*

The compounds analyzed exhibit a broad range of lipophilicity values. The Log P_{oct} values, measured using the pH-metric method with the SiriusT3™ instrument, ranged from -7.95 for angiotensin II to 3.13 for roxithromycin. Everolimus lipophilicity was assessed using the shake-flask assay, due to the absence of ionizable groups. For desmopressin (-3.2) and ritonavir (3.57), the shake-flask method at pH 7.4 was also selected, as both compounds are neutral at this pH. In addition, Desmopressin's high hydrophilicity and Ritonavir's solubility limitations made the pH-metric method unsuitable. The same rationale was applied to Saquinavir (3.85), which also presents solubility challenges. For all these compounds, the same method was used to determine their Log P_{tol} values.

One of the major limitations of the pH-metric approach in lipophilicity determination lies in its reduced reliability when applied to highly hydrophilic and amphiphilic compounds. While the technique is well-suited for molecules with moderate to high lipophilicity (typically within a Log *P* range of 1 to 3), its accuracy significantly decreases when applied to highly polar, multi-ionizable molecules and low solubility compounds. Such compounds often fail to partition effectively into the

n-octanol phase, and in many cases, it is difficult or impossible to identify a pH range in which the compound exists predominantly in its neutral form. As a result, the characteristic shift in the ionization curve, which is critical for the potentiometric determination of Log *P*, cannot be accurately tracked during titration in the presence of *n*-octanol. This limitation is clearly illustrated by the results obtained for highly polar peptides such as angiotensin II, oxytocin, and teicoplanin, which exhibited extremely low Log P_{oct} values of -7.95, -6.33, and -5.51, respectively. Moreover, the high standard deviations associated with these measurements (±0.42, ±0.73, and ±0.85, respectively) further reflect the poor reproducibility and difficulty in fitting the partitioning model under these experimental conditions. Teicoplanin, with cetrorelix, caspofungin, octreotide, lanreotide and vancomycine, also show amphiphilic features and this is reflected in the formation of an emulsion at the end of the experiment, especially due the strong stirring during the pH-metric assay.
These findings highlight the inherent challenges of applying potentiometric methods to highly ionizable, hydrophilic/amphiphilic molecules and underscore the need to interpret Log *P* values cautiously in such cases.

Interestingly, despite being highly polar and containing multiple ionizable groups, compounds such as ACTH fragment 4-10, yielded more reliable Log P_{oct} values, with lower standard deviations compared to other hydrophilic peptides. These compounds may possess favorable conformational or physico-chemical properties, such as hydrogen bonding, reduced molecular flexibility, or enhanced solubility in the biphasic system, which facilitate a more stable titration profile. These factors could contribute to the greater reproducibility and robustness of the potentiometric measurement of Log P_{oct} in these specific cases, bearing in mind, however, that it was not possible to monitor a clear shift in the ionization curve for this highly hydrophilic compound.

Beyond the four extremely hydrophilic compounds discussed previously, two additional peptidomimetics, one macrocycle and one glycopeptide in the dataset exhibit high hydrophilicity (Log $P_{oct} < 0$), with values ranging from -3.2 to -0.72: desmopressin (-3.2), goserelin (-0.98), rifapentine (-0.96) and vancomycin (-0.72). Three compounds, rifampicin (0.39, a macrocycle), lanreotide (0.42, a cyclic octapeptide) and the ACTH fragment 4-10 (0.56, a peptide), can be classified as having low lipophilicity, with Log P_{oct} values between 0 and 1. Five compounds, including one cyclic lipopeptide, three macrocycles, one synthetic decapeptide and one cyclic octapeptide, exhibit moderate lipophilicity (Log P_{oct} between 1 and 3). These are: caspofungin (1.23), everolimus (2.66), erythromycin (2.58), rifaximin (2.79), cetrorelix (2.35) and octreotide (2.55). The remaining three compounds, roxithromycin (3.13), ritonavir (3.57) and saquinavir (3.85) can be classified as having

high lipophilicity (Log P_{oct} > 3). This group comprises one macrocyclic compound and two peptidomimetics.

About the alignment with our experimental data and the lipophilicity data available in literature, six compounds have good correspondence: desmopressin, erythromycin, rifampicin, ritonavir, roxithromycin and saquinavir. Five compounds have an experimental value taken from the *DrugBank* but no source reference are provided, and they are caspofungin, goserelin, rifapentine, rifaximin and vancomycin. Although *DrugBank* values are provided without direct citations, alternative literature sources for experimental pK_a and Log *D* were searched without results. Rifaximin is the only compound between these with the value taken from the experimental properties section of the *DrugBank* aligned with our experimental value. Goserelin and vancomycin have an important discrepancy between the values even if the lipophilicity range is the same (high hydrophilicity). Caspofungin and rifapentine differ significantly in both numeric value (1.23 *vs* 0 and -0.96 *vs* 4, respectively) and lipophilicity range, with rifapentine showing a notable discrepancy.

During the experimental phase, some assays could not be completed under ideal conditions due to the challenging partitioning behavior of certain compounds. Highly hydrophilic molecules exhibited minimal transfer to the organic phase (*n*-octanol), while highly lipophilic compounds showed poor solubility in the aqueous phase. These solubility constraints often limit the formation of a stable biphasic equilibrium. Additionally, amphiphilic compounds, containing both hydrophilic and hydrophobic regions, frequently demonstrated complex and unstable behavior at the phase interface. This was particularly evident in the assays involving caspofungin, cetrorelix, lanreotide, ocreotide, teicoplanin and vancomycin where visible accumulation at the aqueous-organic interface was observed, likely due to their dual-affinity structures. In some cases, mechanical agitation exacerbated the issue, resulting in precipitation or emulsion formation and causing turbidity in the titration mixtures, which interfered with potentiometric measurements. These observations underscore the limitations of traditional biphasic methods when applied to compounds with extreme hydrophilicity or amphiphilic nature, as physical instabilities can compromise both the reliability and reproducibility of lipophilicity measurements. Consequently, the data obtained under such conditions should be interpreted with caution and in general it’s better to don’t use it. In the context of this work the data were used and shown in the graphs only to highlight the unreliability of the technique and the differences compared to the other techniques.

The relationship between the pH-metric Log P_{oct} values and the CLog *P* values calculated using the Percepta 2019 algorithm appeared to be highly linear, showing a strong correlation across fifteen compounds (R^2 = 0.72). The observed consistency between experimental and computational data demonstrates strong agreement, thereby reinforcing the predictive reliability of the algorithm under appropriate conditions. However, four compounds (i.e. angiotensin II, caspofungin, oxytocin and teicoplanin) showed substantial deviation from this trend and did not exhibit the linear correlation shown in Figure 46, with an overall R^2 value of 0.26. Octreotide and rifapentine deviated to a lesser extent compared to other compounds.

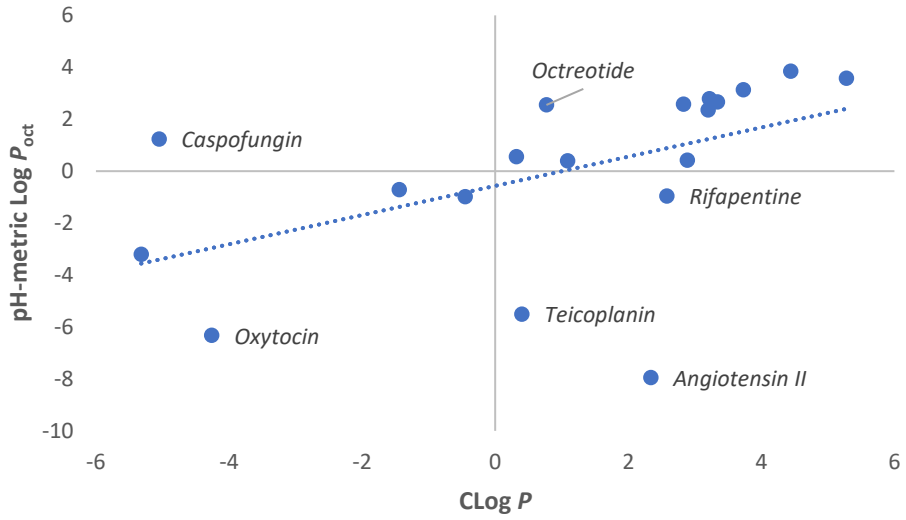


Figure 46. Relationship between CLog P vs pH-metric Log P_{oct}
*R^2 = 0.26; y = 0.56*CLog P - 0.57 ±3.10*

Angiotensin II and oxytocin, are small peptides characterized by very low and negative pH-metric lipophilicity values, accompanied by high standard deviations. In these cases, no appreciable shift in the ionization curve was observed during the titration in the presence of *n*-octanol, likely due to their extremely low solubility in the organic phase, which prevented meaningful partitioning and accurate measurement. It is interesting to note that, although the numerical values differ, oxytocin consistently falls within the same lipophilicity range, classified as very low lipophilicity, across different measurement methods. In contrast, angiotensin II shows greater variability, not only in the absolute lipophilicity values obtained, but also in the lipophilicity classification range. For example, the predicted value from ACD/Labs Percepta software places angiotensin II in the moderate lipophilicity range (2.34), which differs substantially from the experimental data (-7.95).

The other two outliers, caspofungin and teicoplanin, exhibited the highest precipitation phenomena and phase interface accumulation, leading to the formation of unstable emulsions. These issues are

most likely related to their amphiphilic nature, as both molecules contain distinct hydrophilic and lipophilic domains that complicate their behavior in biphasic systems. Taken together, these observations emphasize the methodological challenges associated with the experimental determination of lipophilicity for compounds that are either highly hydrophilic or amphiphilic, particularly when using potentiometric approaches.

Octreotide and rifapentine are the other two compounds that deviate from the trend and interestingly octreotide has a higher pH-metric lipophilicity than the calculated one (2.55 *vs* 0.77), while the rifapentine has a higher calculated lipophilicity than the pH-metric one (2.58 *vs* -0.96). Octreotide showed some emulsion formation phenomena.

In terms of classification range, seven compounds have big discrepancy between pH-metric and calculated lipophilicity: angiotensin II (-7.95 *vs* 2.34), caspofungin (1.23 *vs* -5.05), lanreotide (0.42 *vs* 2.89), octreotide (2.55 *vs* 0.77), rifampicin (0.39 *vs* 1.09), rifapentine (-0.96 *vs* 2.58) and teicoplanin (-5.51 *vs* 0.4).

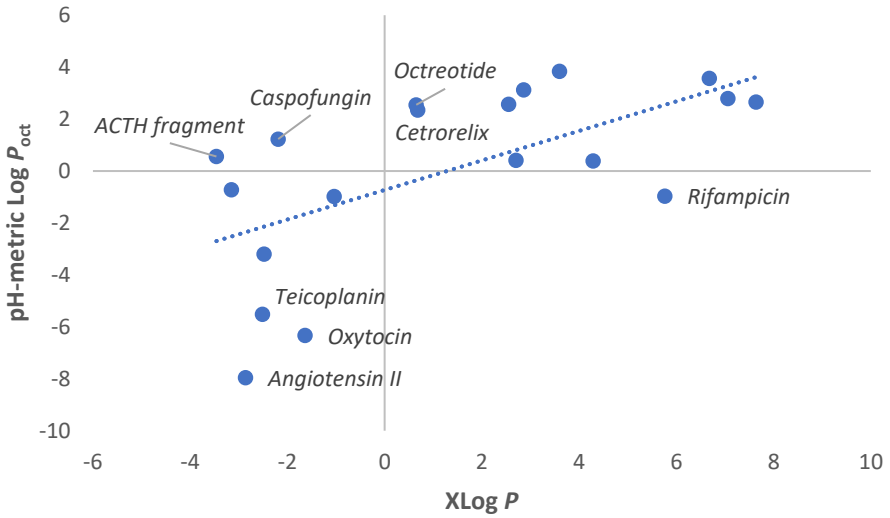


Figure 47. Relationship between XLog P vs pH-metric Log P_{oct}
*R^2 = 0.38; y = 0.57*XLog P - 0.73 ±2.83*

When comparing the pH-metric Log P_{oct} values with another calculated lipophilicity descriptor, XLog *P*, the correlation (R^2 = 0.38) appears to be slightly stronger (Figure 47) than with CLog *P* (Figure 46). However, the same outlier compounds, angiotensin II, oxytocin, caspofungin, and teicoplanin, are still present, further confirming the challenges associated with the experimental determination of lipophilicity for highly hydrophilic and amphiphilic molecules. In this case, only

caspofungin deviates from the predicted range, while teicoplanin and angiotensin II seem to be ranged in the same lipophilicity range even if there is a discrepancy in the numeric value.
In this correlation plot, ACTH fragment 4-10 also appears to deviate slightly from the expected trend, further highlighting the challenges associated with the experimental or *in-silico* characterization of physico-chemical properties for this class of compounds. Such deviations emphasize the inherent complexity of peptides and other multifunctional molecules, which may not behave predictably in biphasic systems due to conformational flexibility, multiple ionizable groups, and limited solubility.

In terms of lipophilicity range, another interesting divergence is observed among seven compounds: ACTH fragment (-3.46 *vs* 0.56), caspofungin (-2.19 *vs* 1.23), cetrorelix (0.68 *vs* 2.35), lanreotide (2.7 *vs* 0.42), octreotide (0.65 *vs* 2.55), rifampicin (4.29 *vs* 0.39), rifapentine (5.76 *vs* -0.96). All the compounds exhibit not only quantitative discrepancies but also differences in classification range. These results suggest that computational descriptors such as XLog *P* may not fully capture the partitioning behavior of large, conformationally flexible molecules, emphasizing the importance of experimental validation in such cases, paying attention to the compounds with unexpected permeability profiles.

Agreement between experimental and predicted lipophilicity range was observed for only twelve compounds within the dataset. For the remaining compounds, the reliability of *in-silico* predictions proved to be limited, particularly for molecules with complex or atypical properties. These findings underscore that both computational models and potentiometric methods can struggle under certain physico-chemical conditions, especially in the presence of extreme hydrophilicity, amphiphilicity, or interfacial instability. Such limitations are especially relevant for bRo5 compounds, whose behavior in partitioning assays can be influenced by low solubility, multiple ionizable groups, or intramolecular interactions, all of which are challenging to capture accurately through *in-silico* or classical experimental approaches.

Although XLog *P* shows a slightly better overall correlation with the pH-metric Log *P* values, this may be due to the presence of significant outliers in the correlation between CLog *P* and experimental lipophilicity. However, the linearity and lower data scatter in the CLog *P* correlation suggest that, despite its limitations, CLog *P* may provide more consistent predictions for bRo5 compounds, particularly those with few ionizable groups and amphiphilic characteristics. This highlights both the potential and the limitations of CLog *P* in modeling the lipophilicity of structurally complex molecules.

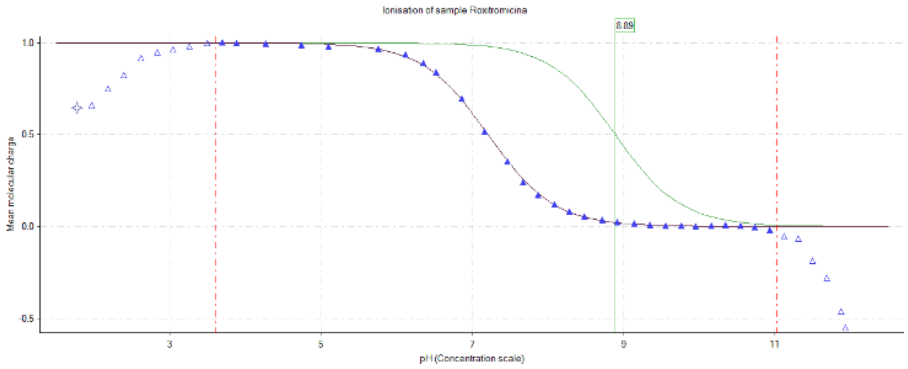


Figure 48. Ionization profile (Bjerrum plot) of roxithromycin in water (blue line) and with the presence of n-octanol (green line)

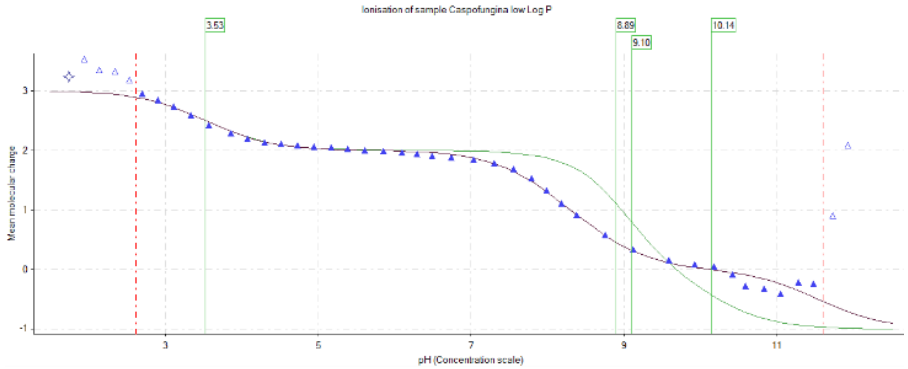


Figure 49. Ionization profile (Bjerrum plot) of caspofungin in water (blue line) and with the presence of n-octanol (green line)

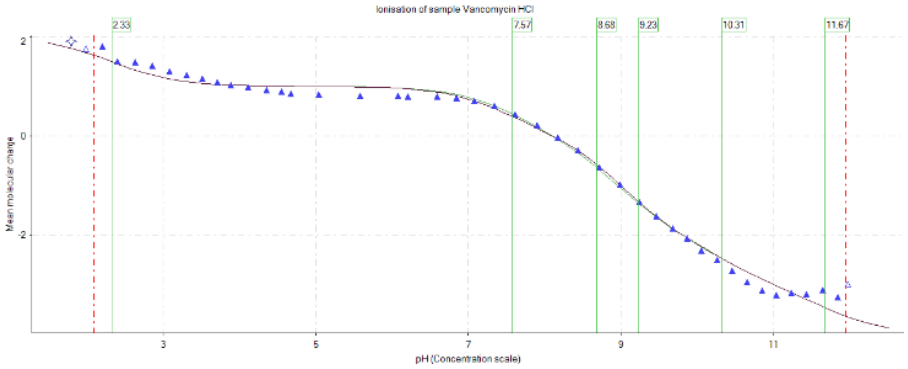


Figure 50. Ionization profile (Bjerrum plot) of vancomycin in water (blue line) and with the presence of n-octanol (green line)

8.4.2 Shake-flask Log D and CLog $D_{7.4}$: correlations and discrepancies

Table 5. Log $D_{7.4}$ measured by shake-flask technique for 19 pharmaceutical compounds with corresponding standard deviation and CLogD calculated using ACD/Labs Percepta 2019. The compounds with "" are prevalently neutral at pH of 7.4 due their pK_a or lack of ionizable group such as everolimus. n/a is used for compounds characterized by an interface accumulation after shake-flask experiment*

Compound	Log $D_{7.4}$ (n=3)	S.D.	CLog $D_{7.4}$
ACTH fragment	-3.33	±0.16	-2.56
Angiotensin II	-3.17	±0.01	-0.41
Caspofungin diacetate	n/a	n/a	-8.7
Cetrorelix acetate	-0.1	±0.08	1.19
Desmopressin*	-3.03	±0.08	-7.32
Erythromycin	0.56	±0.23	2.08
Everolimus*	2.66	±0.38	3.34
Goserelin	-1.50	±0.16	-2.59
Lanreotide acetate	0.39	±0.12	0.04
Ocreotide acetate	-0.36	±0.11	-2.08
Oxytocin	-1.58	±0.15	-4.28
Rifampicin	1.41	±0.14	-1.43
Rifapentin	2.28	±0.05	0.03
Rifaximin	2.60	±0.21	0.05
Ritonavir*	3.57	±0.18	5.28
Roxithromycin	1.26	±0.20	2.91
Saquinavir mesilate	3.87	±0.15	4.41
Teicoplanin	n/a	n/a	-2.62
Vancomycin	-2.72	±0.02	-4.39

Regarding the shake-flask results, the Log $D_{7.4}$ values measured for the dataset ranged from -3.33 for the ACTH fragment to 3.87 for Saquinavir. A total of eight compounds exhibited very high hydrophilicity (Log $D_{7.4} < 0$), and all of them were peptides (natural, synthetic or glycopeptides). These include ACTH fragment (-3.33), angiotensin II (-3.17), desmopressin (-3.03), oxytocin (-1.58), goserelin (-1.50), octreotide (-0.36), cetrorelix (-0.10) and vancomycin (-2.72). Two compounds showed low lipophilicity (Log $D_{7.4}$ between 0 and 1): the macrocyclic antibiotic erythromycin (0.39) and the synthetic cyclic octapeptide lanreotide (0.56). Another five macrocycles, roxithromycin (1.26), rifampicin (1.41), rifapentine (2.28), rifaximin (2.60), and everolimus (2.66) fell within the range of moderate lipophilicity (Log $D_{7.4}$ between 1 and 3). The remaining two compounds exhibited high lipophilicity (Log $D_{7.4} > 3$): the peptidomimetics ritonavir (3.57) and saquinavir (3.87).

Similar to pH-metric assays, evaluating the lipophilicity of amphiphilic compounds such as caspofungin and teicoplanin remains particularly challenging due to their complex physico-chemical properties and their tendency to accumulate at the interface between the aqueous and organic phases. Both the pH-metric technique and the shake-flask method are poorly suited for measuring the

lipophilicity of such compounds, highlighting the critical need for alternative analytical approaches capable of overcoming these limitations. The shake-flask due to its more gently stirring is better than pH-metric Log *P* using SiriusT3^{TM} in the lipophilicity measurements of amphiphilic compounds.

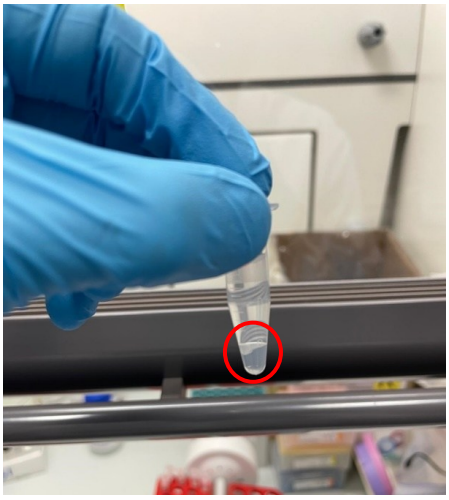


Figure 51. Interface accumulation of Teicoplanin

Figure 52. Interface accumulation of Caspofungin

It is important to highlight that, for highly hydrophilic compounds, inverted phase ratios were also employed to verify the reliability of the shake-flask measurements. Specifically, tests were conducted using various *n*-octanol/water volume ratios such as 1:1, 1:2, and 1:4, where the aqueous phase predominates. These compounds typically exhibit very low affinity for the organic phase, and if partitioning into *n*-octanol is not sufficiently promoted, the resulting Log *D* values may be misleading or underestimated. To overcome this limitation, inverted ratios (i.e., an excess of *n*-octanol), such as 2:1, 4:1, and 10:1, were used to enhance detection and ensure more accurate assessment of lipophilicity.

This strategy also improves the reliability of UV or MS quantification, as the concentration of the compound in the *n*-octanol phase may otherwise fall below the limit of quantification (LOQ), making it difficult to detect and measure accurately.

The correlation between CLog *D* and shake-flask Log $D_{7.4}$ (Figure 53) appears to be stronger than the previous correlations involving different Log *P* values, with a R^2 value of 0.64, indicating a moderate agreement between the experimental and predicted data. However, notable outliers are again observed, particularly among peptide-based compounds. Angiotensin II and desmopressin show significant discrepancies between the calculated and experimental values: -3.17 *vs*. -0.41 for angiotensin II and -3.03 *vs*. -7.32 for desmopressin. Although the numerical values differ, it is

important to highlight that the overall classification remains consistent, with both compounds falling within the highly hydrophilic range.

A more interesting divergence is observed among macrocyclic compounds, such as erythromycin, rifampicin, rifapentine, rifaximin, and roxithromycin. Unlike the peptides, where differences exist but the lipophilicity class remains unchanged, these macrocycles exhibit not only quantitative discrepancies but also differences in classification range. Specifically, erythromycin and roxithromycin display higher predicted CLog *D* values (2.08 and 2.91) compared to their experimental results (0.56 and 1.26), whereas rifampicin, rifapentine, and rifaximin show lower CLog *D* values (-1.41, 0.03 and 0.05) than those estimated experimentally (1.41, 2.28 and 2.60). These results suggest that computational descriptors such as CLog *D* may not fully capture the partitioning behavior of large, conformationally flexible, and macrocyclic molecules, emphasizing the importance of experimental validation in such cases.

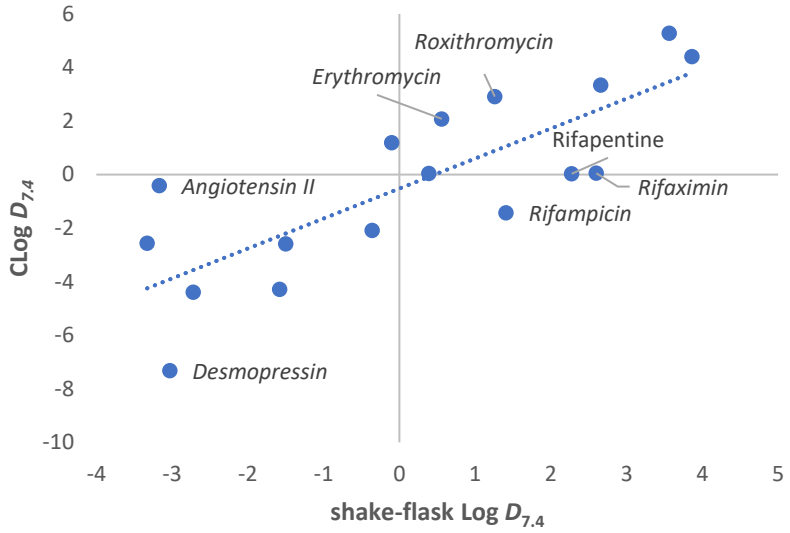


Figure 53. Relationship between CLog $D_{7.4}$ vs pH-metric Log $D_{7.4}$
*R^2 = 0.64; y = 1.12*Log $D_{7.4}$ - 0.52 ±2.09*

8.4.3 CHI Log $D_{7.4}$ as chromatographic help in lipophilicity measurements

Table 6. CHI and corresponding CHI Log $D_{7.4}$ measured by chromatography using the Valko's equation, with corresponding standard deviation

Compound	CHI	CHI $D_{7.4}$	S.D. (n=3)
ACTH fragment	33.59	0.36	±0.02
Angiotensin II	31.66	0.24	±0.01
Caspofungin diacetate	57.71	1.65	±0.01
Cetrorelix acetate	55.78	1.61	±0.10
Desmopressin	32.62	0.32	±0.04
Erythromycin	59.64	2.08	±0.02
Everolimus	118.51	4.98	±0.08
Goserelin	43.24	0.95	±0.08
Lanreotide acetate	49.99	1.31	±0.11
Ocreotide acetate	48.06	1.20	±0.11
Oxytocin	37.45	0.55	±0.00
Rifampicin	63.50	2.03	±0.09
Rifapentine	79.91	2.82	±0.04
Rifaximin	88.59	3.27	±0.08
Ritonavir	91.48	3.47	±0.00
Roxithromycin	72.19	2.73	±0.06
Saquinavir mesilate	94.38	3.66	±0.04
Teicoplanin	42.27	0.84	±0.04
Vancomycin	22.01	-0.26	±0.04

The Chromatographic Hydrophobicity Index (CHI) is a robust, experimentally derived measure of lipophilicity based on gradient reversed-phase HPLC. It approximates the volume percentage of organic modifier (e.g., acetonitrile) required to elute a compound when it reaches a specific retention factor, typically $\log k = 0$ (k=1). This corresponds to a point where the compound is equally distributed between the stationary and mobile phases, analogous to an equilibrium condition used in partition-based methods.

To calculate CHI values accurately, Valkó et al. [80], [81] introduced a calibration procedure using a standard test mixture of 10 reference compounds. Each of these compounds has an assigned CHI value, and their gradient retention times are measured under consistent chromatographic conditions (e.g., 0-100% acetonitrile over ~2 minutes, C18 stationary phase, pH 7.4 buffer solution). The most commonly used calibration compounds include theophylline, phenyltetrazole, benzimidazole, colchicine, phenyltheophylline, acetophenone, indole, propiophenone, butyrophenone, and valerophenone, covering a CHI values range from ~18 to ~96.

To convert CHI values into more interpretable lipophilicity scales, particularly for medicinal chemists familiar with Log *P* or Log *D*, Valkó et al. developed CHI transformation equations based on 98 known drugs [34], [39]:

$$\text{CHI Log } D_{7.4} = 0.054*\text{CHI} - 1.467$$

Regarding the measured CHI Log $D_{7.4}$ values for the dataset used in this PhD project, the results ranged from -0.26 for vancomycin to 4.98 for everolimus. All compounds were successfully evaluated using this chromatographic approach, with no significant technical issues observed during the assays. In contrast to the difficulties encountered in *shake-flask* and potentiometric assays, particularly with amphiphilic, such as caspofungin and teicoplanin, high hydrophilic or poorly soluble compounds, the chromatographic method demonstrated excellent operational robustness and reproducibility.

A notable observation is the broad lipophilicity range covered by the dataset, which highlights the versatility of the CHI-based method in capturing both highly hydrophilic and highly lipophilic compounds. Importantly, a strong linear correlation was observed between CHI Log $D_{7.4}$ and *shake-flask* Log $D_{7.4}$ values, with a $R^2 = 0.86$ (Figure 54). This indicates that the chromatographic approach is reliable and predictive even for structurally complex, bRo5-type molecules, which are typically challenging for conventional Log *D* assays due to aggregation, ionization, phase separation issues, time consuming and interlaboratory variability.

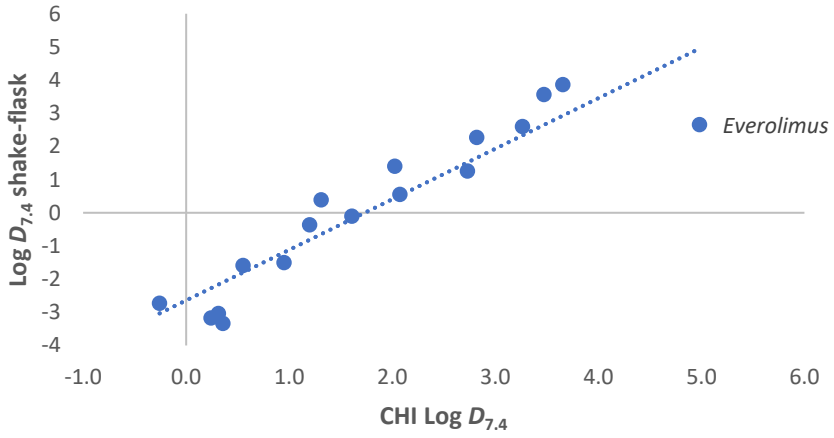


Figure 54.Correlation plot between CHI Log D7.4 and shake-flask LogD7.4
$R^2 = 0.86$; $y = 1.52*(\text{CHI Log } D_{7.4}) - 2.6 \pm 0.57$

Everolimus appears to be an outlier, deviating from the general correlation trend observed between lipophilic values obtained by *shake-flask* and chromatographic method. As it lacks relevant ionizable functional groups, it can be considered a neutral compound across the physiological pH range.

Consequently, its Log *D* at pH 7.4 should be equivalent to its Log *P*, as the compound remains fully uncharged under these conditions. Interestingly, the Log P_{oct} value measured is inconsistent with chromatographic value (~5) with a difference of ~2.5 of Log *P* units. This discrepancy may be attributed to the methodological limitations of the shake-flask approach when applied to highly lipophilic, poorly water-soluble compounds or due the strong interactions of neutral/high lipophilic compounds with the stationary phase (reverse phase).

8.5 Conclusions

The lipophilicity of 19 bRo5 drug compounds was evaluated using three different experimental approaches: pH-metric, *shake-flask*, and chromatographic (CHI Log $D_{7.4}$) methods. In comparison, both the pH-metric and *shake-flask* techniques demonstrated some limitations when applied to highly hydrophilic and amphiphilic compounds, especially potentiometric technique. While these methods can provide accurate estimations of Log *P* for neutral species, based on the pH of the measurement, they can fail or under/overestimate lipophilicity in compounds with pronounced polarity or amphiphilic behavior. In contrast, CHI Log $D_{7.4}$ could offers a more realistic representation of the true partitioning behavior and surface interaction of such molecules under biologically relevant conditions.

Moreover, the chromatographic approach enables lipophilicity profiling across multiple pH values, providing dynamic insight into how hydrophobicity changes with ionization. Importantly, the method maintains high reproducibility and robustness even in the presence of complex structural features, making it suitable for characterizing the behavior of bRo5 compounds in different physiological environments. Furthermore, there is an important linear correlation between lipophilicity measured using *shake-flask* and chromatographic method, that makes CHI Log $D_{7.4}$ reliable for permeability prediction use.

From a drug discovery perspective, the ability to rapidly and reliably assess lipophilicity across diverse chemical classes makes CHI Log $D_{7.4}$ a valuable tool in early-stage screening. Its compatibility with high-throughput workflows and superior performance with structurally complex molecules positions as an effective complement, or even an alternative, to traditional *shake-flask* and potentiometric techniques. In conclusion, the data generated in this project strongly support the implementation of CHI Log $D_{7.4}$ as a robust**,** scalable**,** and biologically relevant method for lipophilicity determination, particularly for compounds that challenge conventional assay systems. It showed a very strong correlation to *shake-flask* Log $D_{7.4}$ ($R^2 = 0.86$).

On the other hand, computationally predicted lipophilicity values remain important in early-stage drug discovery screening. However, with the increasing presence of bRo5 compounds in the pharmaceutical landscape, relying solely on *in-silico* models may represent an oversimplification of the physico-chemical complexity of these molecules, particularly when attempting to predict their ADMET properties.

Based on the experimental data generated in this PhD project, CLog *P* values calculated using Percepta 2019 software aligned well with pH-metric Log P_{oct} values for 12 out of 19 compounds. However, the remaining seven compounds showed large discrepancies, especially for highly hydrophilic and amphiphilic molecules, leading to a poor correlation (R^2 = 0.26). A similar observation was made when comparing Log P_{oct} with XLog *P* values computed using a different algorithm. In that case, the correlation appeared slightly better (R^2 = 0.38), but the data remained widely scattered, likely due to the reduced impact of outliers compared to the previous model. It is important to note that in each correlation, seven compounds showed discrepancy between the lipophilicity classification range.

Regarding the comparison between *shake-flask* Log $D_{7.4}$ and CLog *D*, a better correlation was observed (R^2 = 0.64). However, despite the improved fit, a detailed analysis revealed several discrepancies in the classification of compounds by lipophilicity range. This once again underscores the difficulty in accurately predicting lipophilicity for bRo5 compounds using only computational models. These types of molecules often exhibit conformational flexibility and maybe are capable of forming intramolecular interaction such as IMHBs, both of which can significantly alter their partitioning behavior in experimental systems. Additionally, for amphiphilic compounds, issues related to interfacial accumulation between aqueous and organic phases and emulsion formation persist, further complicating experimental measurements.

9 Experimental section - Polarity

9.1 Aim

As discussed in the chapter on polarity and the importance of this physico-chemical property in membrane permeability, absorption, and bioavailability, developing a new high-throughput method to measure this feature is a key factor in the drug discovery process. Such methods are essential for identifying the most promising molecules, especially as the number of bRo5 and other structurally complex compounds increases in the pharmaceutical market. In response to this trend, accurately characterizing molecular properties has become more and more important to guide the optimization of solubility, membrane permeability, potency, and other ADME-related parameters. The EPSA method, originally developed by G. Goetz et al. [16], [18], [49], provides an experimental parameter that could be correlated to the propensity of NCEs to permeate through biological membranes. It also enables the creation of a database of experimental PSAs to be compared with PSAs calculated through computational approaches, allowing for iterative refinement of predictions and the building of increasingly accurate correlation models.

From an experimental point of view, SFC conditions may offer potential advantages for assessing PSA, as CO_2, a non-polar eluent, is mixed with methanol and ammonium formate, a more polar solvent. This combination allows strong interaction with a polar stationary phase when methanol concentration is low, and as the methanol level increases, it enables the elution of highly polar molecules that are strongly bonded to the stationary phase.

Furthermore, this method can be used to visualize the propensity of molecules to form IMHBs since the non-polar environment created by CO_2 favors their formation.

Following the fine-tuning and verification phase of the SFC method, through the evaluation of analytical parameters such as the precision and accuracy of experimental measurements done during the previous work for my master's degree and supported by numerous literature sources, experimental determination of EPSA was carried out for 19 pharmaceutical standards and hundreds Chiesi's compounds.

Finally, the experimentally determined EPSA values were compared to the calculated TPSA values to assess whether they convey consistent information regarding molecular polarity. In addition, both EPSA and TPSA were correlated with permeability data obtained from the Caco-2 assay to evaluate their ability to predict the passive permeability of complex compounds. For TPSA, the commonly used cut-off value in screening cascades was applied, while for EPSA, the literature-reported threshold was assessed and compared with the cut-off value identified during this PhD project.

9.2 Materials

9.2.1 Materials used for the EPSA measurements

For the EPSA evaluation, 19 pharmaceutical standards were used. The list of the 19 compounds analyzed is shown in Table 1.

Aligned to the literature, a set of 9 calibration standards are needed [14] to perform the analysis: lidocaine (Sigma-Merck), antipirine (Sigma-Merck), chlorpromazine chloride (Sigma-Merck), desipramine chloride (Sigma-Merck), pindolol (Sigma-Merck), diclofenac sodium (Sigma-Merck), 3-nitrobenzoic acid (Sigma-Merck), bumetanide (Sigma-Merck) and furosemide (Sigma-Merck).

In addition to the calibration standards, the following substances were used: methanol (Merck-Sigma), ammonium formate (Merck-Sigma), CO_2 (SOL), and dimethyl sulfoxide (Merck-Sigma).

The following instruments and accessories were used: analytical balance (Mettler Toledo; sensitivity 0.001 mg; capacity 52 g, minimum weighing 2 mg), fume hood, electronic pipettes (Sartorius 5-120 µL, 10-300 µL, 100-5000 µL), magnets, SFC-4000 (Jasco Europe S.r.l.), Advion Expression mass spectrometer (Advion Interchim Scientific), Chromnav Control Center software for SFC control and data analysis, Advion CMS Control for mass spectrometer control, Advion QuantExpress software for mass spectrometry, Chirex 3014 (S)-valine and (R)-1-(alphanaphthyl)ethylamine 250*4.6 mm and 50*4.6mm – 5µm as chromatographic column (Phenomenex), ultrasonic bath (Branson 5510), smart spatula (Sigma Aldrich), 200 µL insert vials, 1.5 mL vials, 4 mL vials, Vortex.

9.3 Methods

9.3.1 Supercritical fluid chromatography (SFC)

This technique, that has been very underrated during these years despite it has a lot of advantages, is seeing a major increase in its use thanks to chiral separation and metabolomics analysis [82]. The SFC uses supercritical CO_2 as the main phase, which is a very safe and cheap solvent. To reach the supercritical phase we need to be over determinate condition of pressure and temperature, 31.3°C and 73.8 bar, which are easy condition to reach. Furthermore, supercritical fluids possess gas-like viscosity, which results in higher diffusivity compared to liquids, thereby enhancing both throughput and resolution. Additionally, their high density, comparable to that of liquids, increases the solubility of analytes. This is why one of the most use concerns the enhancement of time efficiency in chiral separation. Moreover, CO_2 is an apolar solvent similar to the *n*-hexane used in normal phase chromatography but, it can be used with organic modifiers such as methanol, isopropanol and acetonitrile to increase the flexibility in the polarity, using modifiers not miscible with the *n*-xenane. It is also possible to add acids or bases to the organic modifiers. This means that we can used polar

and apolar stationary phases with the same mobile phase, and it allows us to analyze different class of compounds with different features, and this is why the use of this technique is increasing in the metabolomics field [82].

9.3.2 Experimental Polar Surface Area evaluation (EPSA)

The method developed by Goetz et al. [16], [18], [49] before, and the shorter one by Wang et al. [14] more recently, for the experimental polarity characterization of the compounds using SFC as analytical technique is the following.

The 10 mM stock solution in dimethyl sulfoxide (DMSO) of each compound was analyzed. For compounds presenting solubility issues in DMSO, such as teicoplanin and vancomycin, stock solutions 10mM were prepared in water and subsequently diluted 1:10 in DMSO prior to analysis. To minimize degradation due to the potential instability of the compounds and due to the freeze-thaw cycles, the stock solutions have been listed and stored in refrigerator at 4°C. A vortex was used to facilitate solubilization. For poorly soluble molecules an ultrasonic bath was used. From the stock solution, 10 µL were transferred in a vial with insert, then 70 µL of DMSO were added and the diluted sample was analyzed by SFC. Alternatively, the compounds were analyzed directly from the 96-well plate where they were stored, always at the same concentration (10 µL of 10mM solution in 70 µL of DMSO).

The instrument used is a Supercritical Fluid Chromatograph provided by Jasco S.r.l which uses supercritical CO_2 as the main mobile phase. In addition, an organic modifier, methanol with 20mM of ammonium formate was used to elute the most polar compounds characterized by a higher affinity to the chromatographic column. The analysis was performed in gradient mode with a flow rate of 5mL/min. The gradient conditions are the following:

Table 7. Chromatographic gradient used to measure EPSA using SFC, provided by Goetz et al.

Time (min)	CO_2 (%)	MeOH + Amm. Form. 20mM (%)
0.00	95	5
11.00	40	60
15.90	40	60
16.00	95	5
20.00	95	5

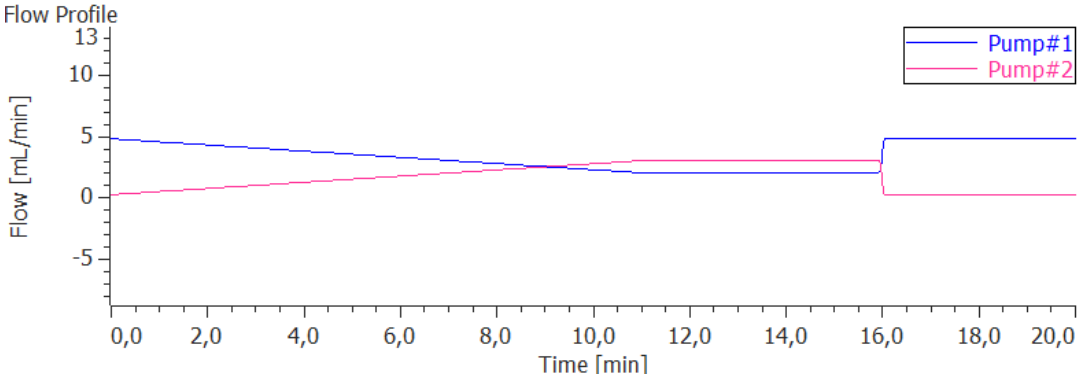


Figure 55. Flow profile of original EPSA method

Initially the percentage of CO_2 was 95%, whereas the percentage of the organic modifier was 5%. From 0 to 11 minutes the organic modifier increases up to 60%, while CO_2 decreases. These conditions are held for 4 minutes until 15 minutes time. From 15 minutes to 16 minutes initial conditions are restored and then maintained for 4 minutes until the analysis ends at 20 minutes.

The injection volume is 5µL. The Chirex 3014 250*4.6 mm with particle size of 5 µm and characterized by stationary phase composed by (S)-Valin and (R)-1-(alphanaphthyl)ethylamine was used as chromatographic column. This one was chosen due to the good balance between liphophilic and hydrophilic features that characterized the stationary phase, and due to the capability of this column to separate a wide range of polarity [18]. The column oven is set to a temperature of 40°C. There are two lamps, a deuterium lamp and a halogen lamp which cover a wide range of wavelength, from 200 to 650 nm. The Photodiode Array Detector (PDA) is used as a detector and it can cover a wide wavelength range, from 200 to 900 nm. To maintain the supercritical state of the CO_2, as stated before, it is important to set specific conditions of temperature and pressure. The latter is controlled by the Back Pressure Regulator (BPR) and for analysis it was set at 120 Bar.

According with Wang et al. [14] a new high throughput method to evaluate the EPSA has been developed. In comparison to the "long method," the shorter approach exhibits only minimal differences. The first one is the addition of a new calibration standard, lidocaine, to the previous list. This molecule was added to cover a wider range of retention time, since antipyrine, the first standard that elutes in original method, has a greater retention time due to the different column size and condition. At the t_0 the percentage of CO_2 is 95%, whereas the percentage of the organic modifier is 5% as well. From the beginning to the 2.75 minutes there is an increase in the percentage of organic modifier up to 65%. These conditions are maintained for 1.25 minutes. From 4 minutes to 5.5 minutes the organic modifier is held to 98% and from 5.5 minutes until the end of the run at 6 minutes, conditions return to their initial conditions. The flow is 4mL/min, which is 1mL/min lower, thus

saving the amount of mobile phase used. The BPR was set at 207 Bar. The column used is the same but shorter, Chirex 3014 (S)-valine and (R)-1 (alphanaphthyl)ethylamine 50*4.6mm (Phenomenex). The gradient is as follows:

Table 8. Chromatographic gradient used to measure EPSA using SFC provided by Wang et al.

Time (min)	CO_2 (%)	MeOH + Amm. Form. 20mM (%)
0.00	95	5
2.75	35	65
4.00	35	65
5.5	2	98
6.00	95	5

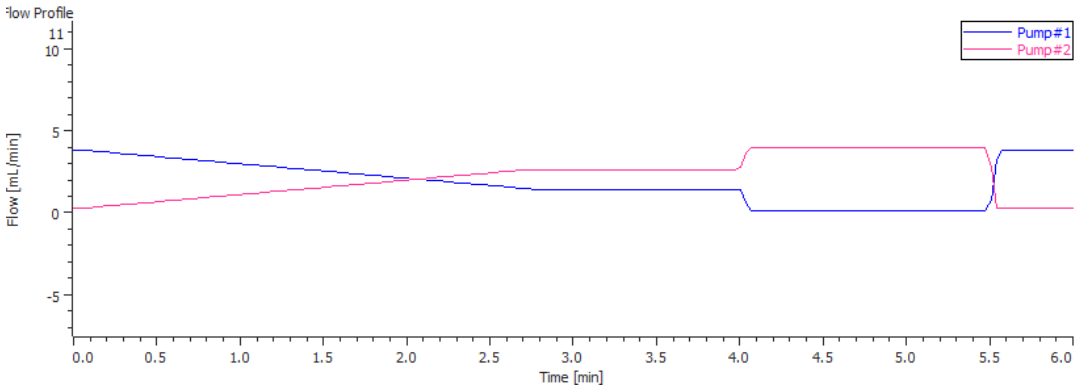


Figure 56. Flow profile of shorter EPSA method developed by Wang et al.

This newer, higher-throughput method was introduced after this project began. To verify its reproducibility, the 19 reference standards were re-analyzed using the shorter method, and the absolute and percentage differences between the new and original EPSA values for 19 standard compounds were evaluated.

9.3.3 Topological Polar Surface Area calculation (TPSA)

To calculate the Topological Polar Surface Area (TPSA) of the compounds analyzed in this study, the software ACD/Labs Percepta 2019 (Advanced Chemistry Development, Inc.) was used. TPSA is a widely adopted molecular descriptor that estimates the surface area occupied by polar atoms, primarily oxygen, nitrogen, and their attached hydrogens, based on the two-dimensional structure of

the molecule. Unlike traditional PSA calculations that require 3D structure generation and surface modeling, TPSA is derived using a fragment-based approach, which offers both speed and computational efficiency, as reported in the work of Ertl et al. [54].

Specifically, the algorithm implemented in ACD/Labs Percepta uses a library of predefined polar fragments containing nitrogen and oxygen (sometimes it is also possible to include phosphorus and sulfur in the calculation), each associated with a tabulated surface area contribution. In this study it is reported that the PSA with inclusion of sulfur atoms provides better correlation with human jejunum permeability than just O- and N-based PSA [54].

The software scans the 2D molecular structure to identify the presence and frequency of these fragments and then calculates the TPSA by summing the individual contributions of all matched fragments. This method enables rapid and consistent polarity estimation, even across large chemical libraries, and has been validated through its correlation with experimentally observed permeability, absorption, and other ADME-related properties.

The use of a fragment-based approach also ensures high reproducibility and minimizes dependence on conformational variability, making it particularly well-suited for high-throughput screening and early-stage compound triaging. However, as discussed in the previous chapter, this simplification may become a limitation when dealing with bRo5 compounds, which often exhibit substantial conformational flexibility. In such cases, TPSA may overestimate the effective polarity of a molecule, as it fails to account for intramolecular interactions such as IMHBs that can shield polar groups from solvent exposure. These dynamic structural features, which are especially relevant in large and flexible molecules, highlight the need for complementary methods, such as experimental EPSA, that can capture the three-dimensional and environment-dependent nature of molecular polarity.

9.4 Results

9.4.1 EPSA measured using long and short method and in-house validation

Table 9. Difference between EPSA measured using long (L.M. EPSA) and short (S.M. EPSA) method. Compounds highlighted with "", teicoplanin and vancomycin, are characterized by an EPSA value out of the measurable range, reported as > 230, using the method developed by Goetz et al. "n/a": not available. Var % : variation percentage*

Compound	L.M. EPSA	S.D. (n=3)	S.M. EPSA	S.D. (n=3)	Var %
ACTH fragment	164	±1.50	163	±1.53	0.61
Angiotensin II	247	±3.51	240	±4.36	2.83
Caspofungin diacetate	152	±2.88	153	±0.58	0.66
Cetrorelix acetate	143	±3.09	143	±3.06	0
Desmopressin	153	±3.61	156	±0.58	1.96
Erythromycin	79	±1.14	83	±0.58	5.1
Everolimus	83	±0.21	83	±0.58	0
Goserelin	152	±1.53	156	±0.58	2.63
Lanreotide acetate	158	±2	160	±1.15	1.27
Ocreotide acetate	147	±0.58	148	±0.01	0.68
Oxytocin	155	±0.58	161	±1.53	3.87
Rifampicin	121	±1.05	121	±0.58	0
Rifapentine	129	±1	129	±0.58	0
Rifaximin	100	±1.26	102	±0.50	2
Ritonavir	81	±0.33	82	±0.01	1.23
Roxithromycin	78	±1.15	82	±0.58	5.12
Saquinavir mesilate	93	±0.94	96	±0.01	3.23
*Teicoplanin**	>230	n/a	>230	n/a	n/a
*Vancomycin**	>230	n/a	280	±3.51	n/a

Table 9 presents the results of a comparative analysis between the original Goetz method and the short gradient approach, conducted using the dataset comprising nineteen compounds performed to verify the applicability and consistency of the short SFC-MS EPSA method, used for EPSA evaluation of part of Chiesi's compounds, developed by Wang et al. (2024) within the context of this work. Since formal validation has already been carried out in literature, the aim here was to confirm that both methods yield consistent EPSA values across the compounds investigated in this study.

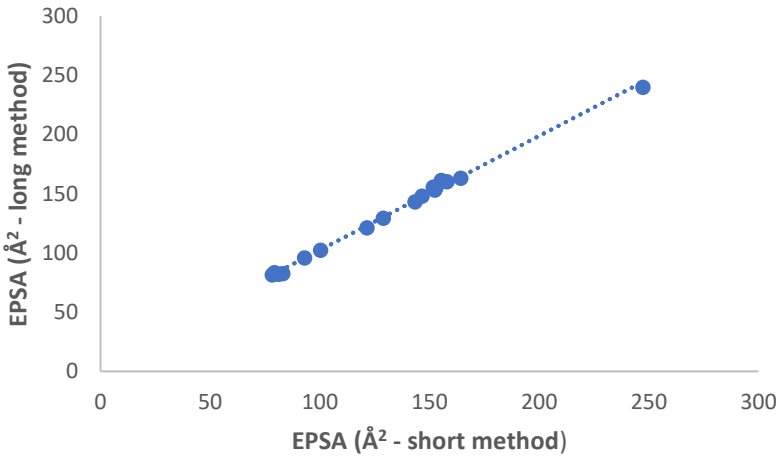


Figure 57. Relationship between EPSA values measured using short method and long method
$R^2 = 0.99$; $y = 0.96*x + 5.98 \pm 2.62$

The correlation (Figure 57) between the EPSA values measured using the original method developed by Goetz et el. and the shorter one developed by Wang et al. has a $R^2 = 0.99$, indicating strong agreement between the two methods and that they yield the same EPSA values.
Also, the linearity obtained using the calibration standard mix (Figure 58) gives the same R^2 value > 0.99 in each method. Lidocaine has been introduced in the calibration set of the short method due to the different condition of the shorter one. The lower void volume and flow, and the shorter column increase the volume of mobile phase to elute antipyrine (first standard in the original calibration set) and the retention time of it, so to extend the calibration range lidocaine, with an EPSA value of 47, was introduced.

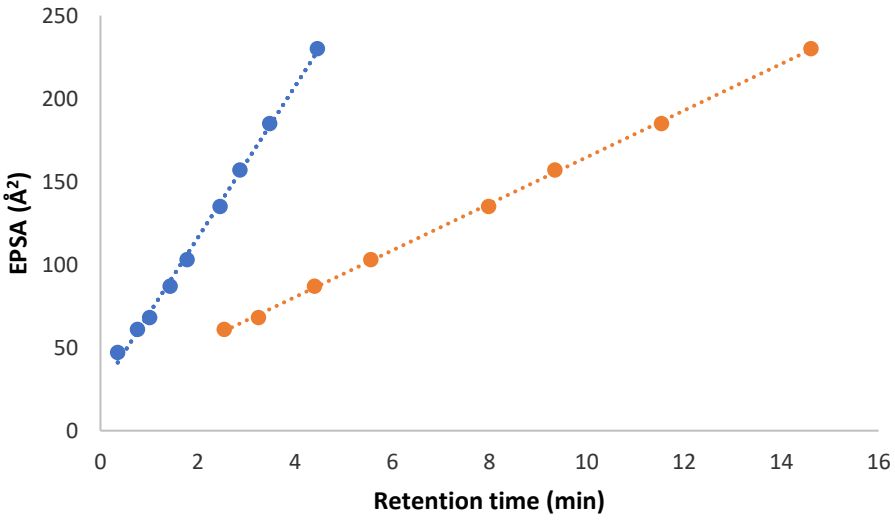


Figure 58.Calibration curves built using the HT-EPSA method (blue) and the original method (orange)
HT-EPSA method: $R^2 = 0.998$, $y = 45.61*Rt + 24.92 \pm 3.30$; Original method: $R^2 = 0.999$, $y = 14.05 + 24.27 \pm 1.41$

Regarding the EPSA values measured using both methods (Table 9), fifteen out of seventeen compounds show a percentage deviation in the EPSA value below 5%, and two of them slightly above 5%, erythromycin (5.1%) and roxithromycin (5.12%). In terms of absolute difference, fifteen out of seventeen compounds show a difference of less than 5 EPSA units, while only angiotensin II (7 units) and oxytocin (6 units) exceed this threshold. In general, no compounds differ in the EPSA range. Interestingly, vancomycin was not detectable under the conditions of the long EPSA method (Goetz et al.), whereas it was clearly observed using the shorter method (Wang et al.). This may be attributed to the faster gradient and higher elution strength early in the run, which reduces the time vancomycin remains bound to the stationary phase and enhances its detection by MS. The shorter method also results in less peak broadening and higher concentration at the detector, improving signal intensity for large polar molecules with low ionization efficiency.

9.4.2 Correlation between chromatographic and computational values

Table 10. EPSA values with their standard deviation and Topological Polar Surface Area of the 19 pharmaceutical standard

Compound	EPSA	S.D. (n=5)	TPSA
ACTH fragment	163	±01.53	384
Angiotensin II	240	±4.36	409
Caspofungin diacetate	153	±0.58	412
Cetrorelix acetate	143	±3.06	498
Desmopressin	156	±0.58	438
Erythromycin	83	±0.58	194
Everolimus	83	±0.58	205
Goserelin	156	±0.58	496
Lanreotide acetate	160	±1.15	355
Ocreotide acetate	148	±0.01	332
Oxytocin	161	±1.53	400
Rifampicin	121	±0.58	220
Rifapentine	129	±0.58	220
Rifaximin	102	±0.50	198
Ritonavir	82	±0.01	146
Roxithromycin	82	±0.58	217
Saquinavir mesilate	96	±0.01	167
Teicoplanin	>230	n/a	662
Vancomycin	280	±3.51	531

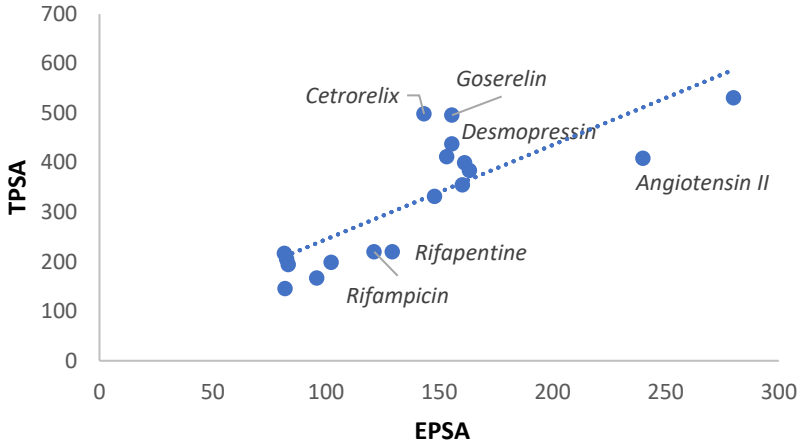


Figure 59. Relationship between EPSA vs TPSA
*R^2 = 0.64; y = 1.90*EPSA + 54.89 ±79.31*

The correlation between experimentally measured and computationally calculated polarity values was evaluated (Figure 59). The calculated values were obtained using ACD/Labs Percepta 2019, a software that estimates polarity based on a fragment-based approach, considering the presence of nitrogen, oxygen, and their directly bonded hydrogen atoms, as previously mentioned. As shown in Figure 59, a discreet correlation was observed, with a coefficient of determination (R^2) of 0.64.

However, some compounds deviated from the expected trend. Specifically, angiotensin II**,** rifampicin, and rifapentine showed higher EPSA values than predicted by the correlation, while cetrorelix, goserelin, and desmopressin exhibited higher TPSA values than expected. These discrepancies may be attributed to structural features such as conformational flexibility or intramolecular hydrogen bonding, which affect experimental and calculated polarity differently, to the limitations of calculated properties in terms of simplification of the polarity features of complex compounds, and to the intrinsic difference between these two polarity values also in terms of numerical scale value and measurement methodology.

It is also important to note that EPSA values could not be determined for teicoplanin. On rare occasions, highly polar peptide-based compounds may present challenges in detection. Some compounds elute over multiple gradient runs but are so diluted that neither UV nor MS detectors are able to detect them.

This highlights a limitation of the EPSA method in analyzing large, polar molecules, particularly peptides, that may fall outside the detection capabilities of standard SFC-MS setups. These are also the compounds with the highest TPSA value in the dataset.

Polarity is commonly evaluated within defined thresholds, where compounds with values below certain limits are considered more likely to be permeable and suitable for oral administration. These thresholds and their implications will be discussed in more detail in the chapter on permeability, which explores the correlation between passive permeability and the physico-chemical properties assessed in this PhD work. Briefly, for TPSA, values in the range of 60-90 Å² are generally associated with permeable compounds, while values above 90 Å² are considered potentially limiting for absorption and oral bioavailability. In the specific context of beyond Rule of 5 (bRo5) compounds, the acceptable TPSA threshold can extend up to 120-140 Å², to account for molecules that might still display permeability through alternative mechanisms or due to dynamic conformational effects. For EPSA, according to literature data, values below 100 Å² are typically associated with a high likelihood of passive permeability and absorption. Interestingly, while all the compounds analyzed in this study show TPSA values above 140 Å², some of them exhibit EPSA values below 100 Å². This suggests that, based solely on TPSA, none of the compounds would be predicted to cross biological membranes efficiently. In contrast, EPSA offers a more favorable assessment, as it may account for permeability-enhancing features not captured by static TPSA calculations.

9.5 Conclusions

In the context of polarity evaluation through both experimental chromatographic and computational approaches, this study assessed the reproducibility between two EPSA measurement methods and the correlation between EPSA and calculated TPSA values.

The project initially employed the method developed by Goetz et al. [16], [18], [63], and later transitioned to the shorter, high-throughput protocol introduced by Wang et al. [14], in order to streamline EPSA analysis within a screening cascade. The results confirmed a high degree of reproducibility between the two methods: EPSA values obtained with the shorter protocol closely matched those from the original, longer method. To verify its reliability, interday variability and absolute differences were evaluated across a representative set of compounds, showing that the faster protocol maintains accuracy despite significantly reducing analysis time, by approximately 14 minutes per sample. This efficiency has allowed the implementation of the Wang method into the Chiesi's internal screening cascade, enhancing throughput without compromising data quality.

Regarding the correlation between experimental EPSA and calculated TPSA, a discreet linear relationship ($R^2 = 0.76$) was observed across most compounds. However, six molecules were identified as outliers, deviating from the general correlation trend. This discrepancy may arise from

multiple factors: in complex and flexible molecules, conformational changes induced by the apolar chromatographic environment may lead to reduced effective polarity, particularly through the formation of IMHBs. Conversely, in some cases, strong interactions with the stationary phase may increase retention and result in EPSA values that exceed the measurable experimental polarity range. These findings confirm that while TPSA provides a useful computational polarity estimation, EPSA offers a finer experimental measure, especially for structurally complex or bRo5 compounds. The integration of both descriptors can thus provide a more comprehensive polarity profile, improving permeability prediction and compound triaging in early drug discovery. It is interestingly to notice that on the basis of the TPSA values, no molecules should be permeable through biological membrane because all of them have TPSA > 140 $Å^2$.

10 Experimental section - Intramolecular Hydrogen Bond

10.1 Aim

Molecular chameleonicity, defined as the propensity of a compound to form dynamic intramolecular hydrogen bonds (IMHBs), is gaining increasing importance in the field of drug discovery. With the growing interest in bRo5 molecules, such as peptides, macrocycles, and PROTACs, this property provides medicinal chemists with a valuable tool to identify promising candidates with potential oral bioavailability. Molecular chameleons are compounds that can form IMHBs between hydrogen bond donors (HBDs) and acceptors (HBAs) depending on the surrounding environment [10], [11], [20], [49]. By doing so, these compounds reduce their overall exposed polarity, which can enhance membrane permeability and support oral administration.

In the context of this PhD project, the formation of IMHBs in bRo5 compounds was investigated using two main approaches. The first approach involved measuring the difference in partition coefficients (ΔLog $P_{oct\text{-}tol}$), determined by a pH-metric method across two biphasic systems: one consisting of *n*-octanol/water buffer solution and the other of toluene/water buffer. The rationale is that while compounds may partition well into *n*-octanol (a relatively polar organic phase that can donate and accept hydrogen bonds), they generally partition poorly into toluene (a more apolar solvent) unless they can form IMHBs that mask their polar groups. As described by Caron et al. [20], [63], ΔLog $P_{oct\text{-}tol}$ can thus serve as an indirect indicator of IMHB formation.
The second approach utilized Experimental Polar Surface Area (EPSA) measurements. EPSA, as discussed earlier, is an experimental parameter developed by Goetz et al. at Pfizer that quantifies the exposed polarity of a molecule under apolar conditions using supercritical fluid chromatography (SFC) [16], [18], [49]. In an apolar environment, molecules capable of forming IMHBs are more likely to adopt folded conformations, resulting in shorter retention times and consequently lower EPSA values. In some studies, pairwise analysis between compounds (such as cyclic peptides) with and without the ability to form IMHBs has demonstrated that those capable of forming IMHBs exhibit lower EPSA values. Unfortunately, due to the high cost and limited availability of the compounds in our dataset, it was not possible to synthesize or obtain suitable analogues for pairwise comparison in this project.

To verify the presence of IMHBs, a set of 13 compounds was selected from a dataset of 19 pharmaceutical standards, based on their EPSA, their ΔLog $P_{oct\text{-}tol}$ and their permeability values. By combining ΔLog $P_{oct–tol}$ and EPSA data, we aimed to identify compounds with a high likelihood of forming IMHBs. Selected candidates from this analysis were further examined by solution-state NMR

spectroscopy, basing on a well-established experimental concept: when a compound is dissolved in two different solvents, deuterated dimethyl sulfoxide (DMSO-d_6), a strong hydrogen bond acceptor, and deuterated chloroform ($CDCl_3$), a weaker one, any significant change in the chemical shift of hydrogen bond donors (typically NH or OH groups) suggests that those groups are solvent-exposed and interacting with the external medium. In contrast, if the chemical shift remains consistent across both solvents, it indicates that the donor group is internally satisfied by forming an IMHB, which shields it from solvent interaction. This method has been widely discussed and validated in the literature and provides a robust tool for confirming the presence of IMHBs in complex molecules [20], [67].

10.2 Materials

10.2.1 Materials used to measure the lipophilicity using SiriusT3^{TM} automatic titrator

For the materials used to measure the lipophilicity expressed as Log $P_{oct/wat}$ and Log $P_{tol/wat}$, using pH-metric technique, see the chapter in the experimental section about lipophilicity (p. 77).

10.2.2 Materials used for lipophilicity evaluation through the Shake-Flask assay

For the materials used to measure the lipophilicity expressed as Log $P_{oct/wat}$ and Log $P_{tol/wat}$, through the shake-flask assay, see the chapter in the experimental section about the lipophilicity (p. 77).

10.2.3 Material used for EPSA evaluation using Supercritical Fluid Chromatography

For the material used to measure the polarity expressed as EPSA, through SFC, see the chapter in the experimental section about polarity (p. 97).

10.2.4 Materials used to verify the presence of IMHBs using NMR spectroscopy

For the NMR experiments, 13 pharmaceutical standards were used. The list of the compounds used is the following: caspofungin diacetate (abcr/Merck/BLDpharma), erythromycin (Sigma-Aldrich), everolimus (Merck-Sigma/BLDpharma), lanreotide acetate (abcr/Merck-Sigma/Sigma-Aldrich), octreotide acetate (abcr), rifampicin (Sigma-Aldrich), rifapentine (Sigma-Aldrich), rifaximin (Sigma-Aldrich/BDLpharma), ritonavir (Merck-Sigma), roxithromycin (Sigma-Aldrich), saquinavir mesylate (Merck-Sigma/BLDpharma), teicoplanin (Merck-Sigma), vancomycin (abcr).
In addition, the following substances were used: deuterated dimethyl sulfoxide (DMSO-d_6) (≥99.9 atom % D) (Sigma-Aldrich), deuterated chloroform ($CDCl_3$) (≥99.8 atom % D) (Sigma-Aldrich).

The following instruments and accessories were used: analytical balance (Mettler Toledo; sensitivity 0.001 mg; capacity 52 g, minimum weighing 2 mg), fume hood, glass Pasteur pipettes, standard 5 mm precision NMR tubes (Wilmad® or equivalent) with polyethylene caps, Bruker Avance III HD 600 MHz spectrometer, equipped with a CryoProbe™ for enhanced sensitivity, ACD software suite (ACD Labs) for data analysis.

10.3 Methods

10.3.1 Log P_{oct} and Log P_{tol} evaluation using the SiriusT3^{TM} automatic titrator

For the method used to measure the lipophilicity expressed as Log $P_{oct/wat}$ and Log $P_{tol/wat}$, using pH-metric technique, see the chapter in the experimental section about the lipophilicity (p. 78).

10.3.2 Log P_{oct} and Log P_{tol} evaluation through the shake-flask assay

For the method used to measure the lipophilicity expressed as Log $P_{oct/wat}$ and Log $P_{tol/wat}$, through the shake-flask assay, see the chapter in the experimental section about the lipophilicity (p. 79).

10.3.3 EPSA evaluation using Supercritical Fluid Chromatography

For the method used to measure the polarity expressed as EPSA, through SFC, see the chapter in the experimental section about polarity (p. 98).

10.3.4 Method used to verify the presence of IMHBs using NMR spectroscopy

To evaluate the potential formation of intramolecular hydrogen bonds (IMHBs), each compound was analyzed in both DMSO-d_6 and $CDCl_3$, which differ significantly in hydrogen bond accepting capacity. DMSO-d_6, as a strong hydrogen bond acceptor, tends to interact with exposed HBD groups, leading to downfield shifts in the corresponding NMR signals. In contrast, $CDCl_3$, being a weak acceptor, does not significantly shift such signals unless the HBD is internally involved in hydrogen bonding. Therefore, the determination of the solute hydrogen bond acidity descriptor *A*, calculated as $0.0065 + 0.133*\Delta\delta$ (where $\Delta\delta = \delta DMSO - \delta CDCl_3$ for each exchangeable proton), offers insights into the presence of IMHBs [67]. All samples were prepared in standard 5 mm NMR tubes, with a typical concentration in the range of 1-10 mg/mL, depending on compound solubility/availability and spectral quality. A value of *A* close to 0 indicates a strong IMHB, instead a value > 0.20 indicates that IMHB is not present.

^{1}H NMR spectra were recorded using the standard zg30 pulse program, with acquisition parameters adjusted based on sample quality: 16-64 scans, relaxation delay (D1) of 1-2 seconds, and acquisition time of 2-3 seconds. Chemical shifts were internally referenced to the residual solvent peaks (δ = 2.50 ppm in DMSO-d_6, and δ = 7.26 ppm in $CDCl_3$). Spectra were processed using manual phase and baseline correction, and particular focus was given to NH and/or OH proton signals, which were monitored in both solvents. Spectra were collected at a controlled temperature of 298 K (25 °C).

2D spectra have been useful for the identification of NH and OH groups, implicated in the possible formation of IMHBs. The following 2D spectra (with the relative standard pulse program) have been acquired: COSY (cosygpmfppqf), TOCSY (dipsi2gpphzs), ROESY (roesyadjsph), HSQC (hsqcedetgpsp.3), HMBC (hmbcetgpl3nd), with acquisition parameters adjusted based on sample quality: 4-8 scans.

10.4 Results

Table 11. The measured values of Log P_{oct}, Log P_{tol} using pH-metric technique, the resulting ΔLog $P_{oct-tol}$ and the EPSA values of the dataset selected for the NMR analysis, highlighted in green. Compound with "" such as desmopressin, everolimus, ritonavir and saquinavir were evaluated using the shake-flask due their pK_a. They are neutral at pH of 7.4 and 9.3 for saquinavir. For ritonavir and saquinavir the shake-flask was chosen also due to some issues during the pH-metric analysis. Compounds highlighted in italic type are those characterized by an emulsion formation after the pH-metric lipophilicity evaluation*

Compound	Log P_{oct}	Log P_{tol}	ΔLog $P_{oct-tol}$	EPSA
ACTH fragment	0.56	-0.63	1.19	163
Angiotensin II	-7.95	-6.25	-1.7	240
Caspofungin diacetate	*1.23*	*0.38*	*0.85*	*153*
Cetrorelix acetate	*2.35*	*1.85*	*0.5*	*143*
Desmopressin*	-3.2	-3.24	0.04	156
Erythromycin	2.58	1.44	1.14	83
Everolimus*	2.66	3.38	-0.72	83
Goserelin	-0.98	-0.7	-0.28	156
Lanreotide acetate	*0.42*	*2.53*	*-2.11*	*160*
Ocreotide acetate	*2.55*	*1.07*	*-1.42*	*148*
Oxytocin	-6.33	-0.96	-5.37	161
Rifampicin	0.39	0.82	-0.43	121
Rifapentine	-0.96	2.05	-3.01	129
Rifaximin	2.79	1.41	1.38	102
Ritonavir*	3.57	2.2	1.37	82
Roxithromycin	3.13	2.73	0.4	82
Saquinavir mesilate*	3.85	3	0.85	96
Teicoplanin	*-5.51*	*-6.68*	*1.17*	*>230*
Vancomycin	*-0.72*	*-0.67*	*-0.05*	*280*

Compounds were selected based on criteria intended to highlight those with high or low likelihood of forming IMHBs. We selected molecules showing EPSA values below 140 and/or $\Delta\text{Log}\,P_{\text{oct-tol}}$ values close to 0 or between 0 and 1.5 were interpreted as having a greater potential to form IMHBs, since this pattern suggests reduced exposed polarity and balanced distribution between polar and apolar environments. On the other hand, compounds with EPSA values above 140 and $\Delta\text{Log}\,P_{\text{oct-tol}}$ values greater than 1.5 or displaying very low partitioning in both solvents were considered unlikely to form IMHBs and were used as negative controls.

Among the compounds identified as likely IMHB formers were erythromycin, everolimus, rifampicin, rifapentine, rifaximin, ritonavir, roxithromycin and saquinavir. Notably, saquinavir was also used as a positive control, as its ability to form IMHBs has been previously discussed in the literature [11], [64]. These compounds exhibited $\Delta\text{Log}\,P_{\text{oct-tol}}$ values ranging from approximately -2.05 to 1.4, and EPSA values generally below 130, consistent with the hypothesis of polarity masking through IMHB formation.
In contrast, compounds such as caspofungin, lanreotide, octreotide, teicoplanin and vancomycin were selected as negative controls, displaying EPSA values well above 140, some even exceeding 230, and showing poor or highly asymmetric partition behavior between *n*-octanol and toluene.

It is worth noting that for several of the negative control compounds, the measurement of lipophilicity presented significant experimental difficulties due to the formation of emulsions during the potentiometric titration process. Although no system warnings were issued by the instrument (SiriusT3™), this phenomenon may has affected the precision of the partition coefficient values. Nonetheless, the experimental results obtained allowed for a classification of compounds with differing IMHB propensities and supported the use of $\Delta\text{Log}\,P_{\text{oct-tol}}$ and EPSA as complementary predictors, with NMR analysis serving as the final validation tool to confirm the presence of intramolecular hydrogen bonding where suggested.

In recent years, the use of $\Delta\text{Log}\,P_{\text{oct-tol}}$ as a descriptor for assessing the propensity of molecules to form IMHBs has gained increasing interest in drug discovery. In particular, the studies by Shalaeva et al. and Caron et al. have extensively described both the theoretical rationale and practical implications of this parameter. Specifically, a small $\Delta\text{Log}\,P_{\text{oct-tol}}$ value close to zero suggests that the compound exhibits similar partitioning in both systems, which is typically interpreted as evidence that polar hydrogen bonding groups are internally satisfied through IMHBs and therefore less available for interaction with the solvents [63]. However, despite this conceptual clarity, no strict

numerical cutoff has been formally defined to distinguish between IMHB-forming and non-IMHB-forming compounds. The interpretation of $\Delta\text{Log}\,P_{\text{oct-tol}}$ often relies on a pairwise comparison approach, where the compound under investigation is evaluated against a control compound of similar structure that is incapable of forming IMHBs. When the $\Delta\text{Log}\,P_{\text{oct-tol}}$ of the control is significantly higher than that of the test compound, this suggests that the test compound preferentially distributes in toluene, indicative of folded conformations stabilized by IMHBs. On the contrary, if the $\Delta\text{Log}\,P_{\text{oct-tol}}$ of the test compound is higher than that of the control, the molecule likely lacks IMHBs and maintains a more extended conformation in apolar environments [63].

Caron et al. confirmed these observations and emphasized that $\Delta\text{Log}\,P_{\text{oct-tol}}$ is a clean and high-throughput descriptor for assessing exposed hydrogen bond donor (HBD) capacity. In particular, they showed that compounds with a single HBD group and $\Delta\text{Log}\,P_{\text{oct-tol}}$ near zero are likely to form IMHBs, whereas higher $\Delta\text{Log}\,P_{\text{oct-tol}}$ values correspond to molecules with exposed HBD groups [20]. However, interpretation becomes more complex when multiple HBDs are present, as only a subset may be engaged in IMHBs. In such cases, $\Delta\text{Log}\,P_{\text{oct-tol}}$ reflects an average effect, and matched molecular pairs or controls are especially critical for accurate interpretation [20]. Due to the important limitation of this type of assay concerning solubility of compounds in toluene, Caron et al. performed the analysis using computational software that predicts the partition *n*-octanol and toluene.

Taking together, these works establish $\Delta\text{Log}\,P_{\text{oct-tol}}$ as a powerful experimental and computational tool for the prospective identification of IMHBs, although its predictive power is best realized when used in relative comparisons rather than as an absolute threshold. The descriptor is particularly valuable in early-stage drug discovery to support the design of compounds with improved permeability and oral bioavailability through strategic masking of polarity via intramolecular hydrogen bonding. Due to the lack of analogues for each compound, the use of EPSA value can help to distinguish compounds able to form IMHBs.

According to Goetz et al., the EPSA method enables the indirect detection of intramolecular hydrogen bonding by measuring the chromatographic retention of compounds under supercritical fluid conditions. While the study does not define a strict cutoff EPSA value to determine whether a compound forms an IMHB, it introduces the concept of pairwise analysis. In this framework, a difference in EPSA (ΔEPSA) greater than +30 units between a compound and its non-IMHB-forming analogue is considered significant and supportive of intramolecular hydrogen bonding [49]. Generally, compounds capable of forming IMHBs exhibit lower EPSA values due to polarity masking effects, often falling below 120, though interpretation is context-dependent and must account for chemical topology and matched controls.

Following the NMR experiments, it was possible to obtain interpretable results for 6 compounds out of the 13 initially selected. Unfortunately, NMR analysis requires a relatively high amount of material, with typical concentrations of at least 0.1mg/mL to ensure sufficient signal-to-noise ratio. For 5 compounds, it was not possible to reach the required concentration due to limited solubility, particularly in $CDCl_3$. While this might initially seem like a limitation, it can be considered informative in itself. In fact, poor solubility in apolar solvents such as chloroform may suggest that the compound retains a highly polar, solvent-exposed conformation, and is therefore less likely to form IMHBs. If IMHBs were present, they might have contributed to the internal masking of polar groups, improving solubility in nonpolar environments. Thus, the lack of solubility could be interpreted as a negative indicator for chameleonic behavior and supports the hypothesis that these specific compounds do not form IMHBs under the tested conditions. Regarding the remaining compounds, everolimus was found to be chemically unstable in $CDCl_3$, and the presence of significant impurities in the sample prevented reliable spectral interpretation. As for rifaximin, its NMR spectrum in DMSO-d_6 was not suitable for analysis due to the extensive presence of chemical exchanges, which resulted in severe signal broadening and overlap, making the identification of diagnostic peaks impossible.

Table 12. The results from NMR analysis for the evaluations of IMHBs with ppm measured in DMSO and CDCl3, the difference of measured ppm (Δδ), the solute hydrogen bond acidity (A) and where present the strength of the IMHB, on the basis of "A", with the groups implicated in the formation of IMHB. Structures available from p. 58 in the results chapter of dissociation constant

Compound	ppm δ_{DMSO}	ppm δ_{CDCl_3}	Δδ	A	IMHB
Saquinavir	8.90	9.13	-0.23	-0.02	Strong (32-NH)
	4.76	3.73	1.03	0.14	Weak (14-OH)
	7.81	6.96	0.85	0.11	Weak (19-NH)
Erythromycin	3.87	3.91	-0.04	0.001	Strong (21-OH)
Rifampicin	12.52	12.02	0.50	0.07	Strong (40-OH)
	4.17	3.64	0.53	0.08	Strong (44-OH)
Rifapentine	12.52	12	0.52	0.08	Strong (40-OH)
	4.15	3.61	0.54	0.08	Strong (44-OH)
Roxithromycin	3.93	4.34	-0.41	-0.05	Strong (21-OH)
	3.93	3.39	0.54	0.08	Strong (54-OH)
	4.05	3.15	0.9	0.13	Weak (19-OH)
Ritonavir	7.68	6.57	1.11	0.15	Very weak (11-NH)

The results, summarized in Table 12, reveal that five out of six compounds exhibit clear evidence of IMHB formation.

Saquinavir (ΔLog $P_{oct–tol}$ = 0.85; EPSA = 96) shows one strong IMHB involving the 32-NH, and two weaker interactions associated with hydroxyl groups at positions 14 and 19, in agreement with literature reports and as previously discussed in earlier chapters. Erythromycin (ΔLog $P_{oct–tol}$ = 1.14; EPSA = 83) presents one strong intramolecular hydrogen bond. Rifampicin (ΔLog $P_{oct–tol}$ = -0.43; EPSA = 121) displays two strong IMHBs involving hydroxyl groups at positions 40 and 44, while its analogue, rifapentine (ΔLog $P_{oct–tol}$ = –2.05; EPSA = 129), also exhibits two strong IMHBs at the same positions. Roxithromycin (ΔLog $P_{oct–tol}$ = 0.4; EPSA = 82) forms two strong IMHBs at positions 21 and 54, and one weaker interaction at position 19. Finally, ritonavir (ΔLog $P_{oct–tol}$ = 1.37; EPSA = 82) shows evidence of a single weak IMHB involving the 11-NH proton.

These findings support the predictive value of both ΔLog $P_{oct–tol}$ and EPSA in identifying compounds with the capacity to adopt folded conformations stabilized by internal hydrogen bonds.

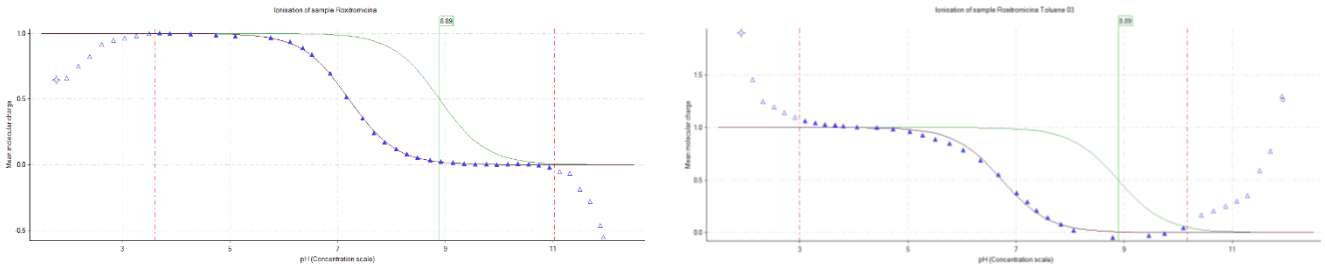


Figure 60. Bjerrum plot of roxithromycin in two different biphasic system, n-octanol-water (right plot) and toluene-water (left plot). The partitioning is visible in both the biphasic systems

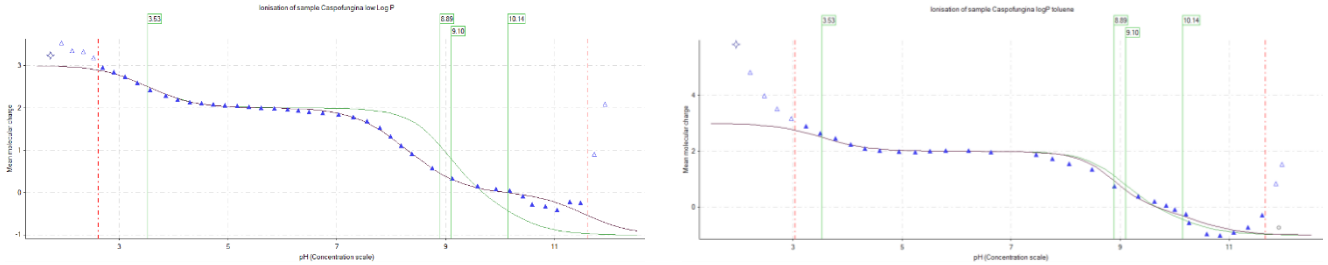


Figure 61. Bjerrum plot of caspofungin in two different biphasic system, n-octanol-water (right plot) and toluene-water (left plot). The partitioning is more important in n-octanol than in toluene

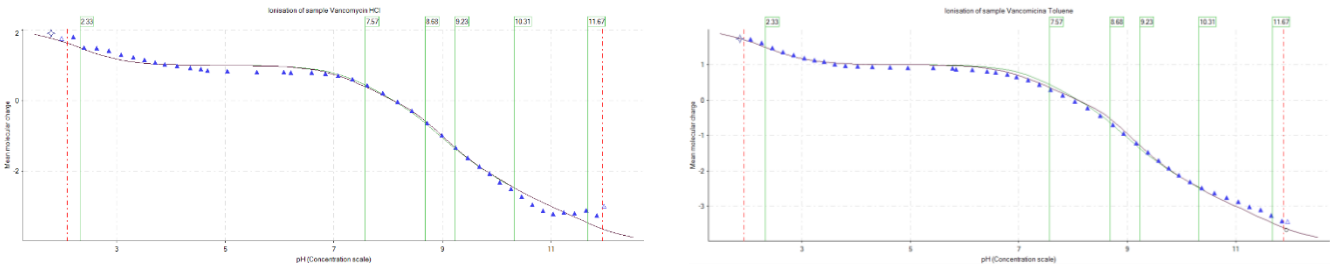


Figure 62. Bjerrum plot of vancomycin in two different biphasic system, n-octanol-water (right plot) and toluene-water (left plot). It was not possible to highlight difference in ionization profile due the high hydrophilicity of the compound

10.5 Conclusions

In conclusion, the results obtained in this study support the predictive value and utility of both $\Delta\text{Log}\,P_{\text{oct–tol}}$ and EPSA in identifying compounds with a strong propensity to form intramolecular hydrogen bonds (IMHBs). Specifically, compounds showing $\Delta\text{Log}\,P_{\text{oct–tol}}$ values close to zero, or that exhibit clear partitioning into the toluene phase, and those with EPSA values below 130 A^2, were consistently associated with the presence of IMHBs, even in the absence of reference analogs for pairwise analysis. The solubility issues of the other compounds in the selected dataset could be a clue to the missing of IMHBs, as discussed in the previous chapters. The NMR technique proved to be a robust and reliable confirmatory tool for detecting IMHBs at the atomic level. However, it is not a high-throughput method like chromatographic (EPSA) or lipophilicity-based ($\Delta\text{Log}\,P_{\text{oct–tol}}$) assays, and its application is often limited by compounds solubility and the requirement for relatively high concentrations (5-10 mM) to ensure sufficient spectral resolution. Unfortunately, we were unable to acquire NMR data for most of the compounds presumed to be unable to form IMHBs. The only compound fully analyzed in this category was ritonavir, which showed a single weak intramolecular hydrogen bond, with a calculated A value of 0.15, falling within the intermediate range. The solubility limitations observed for the remaining compounds in the dataset, particularly in apolar solvents like $CDCl_3$, may further support the absence of IMHBs, as such behavior is consistent with a high degree of solvent-exposed polarity, as discussed in the previous chapters.

11 Experimental section - Caco-2 permeability prediction using physico-chemical properties

11.1 Aim

Permeability refers to the ability of a NCE to cross cell membranes, thereby enabling its absorption from the site of administration into the systemic circulation. The percentage of the drug that reaches the circulatory system after administration is referred to bioavailability. Permeability is a crucial biological process that influences drug absorption and, consequently, its bioavailability. The goal of pharmaceutical optimization is to develop a NCE with optimal physico-chemical properties that ensure sufficient absorption, allowing the drug to effectively interact with its biological target. All the physico-chemical properties described in the previous chapters, such as pK_a, lipophilicity, polar surface area (PSA) and the presence of intramolecular hydrogen bonds (IMHBs) can significantly affect permeability, absorption, and bioavailability.

In this chapter, the permeability data of 19 pharmaceutical standards, measured using Caco-2 cell culture assay with the help of the *Pharmacokinetics, Biochemistry, and Biometabolism* (PKBM) department of Chiesi Farmaceutici, will be evaluated and correlated with their measured physico-chemical properties, in order to assess whether chromatographic parameters can improve physico-chemical screening during the drug discovery phase. Furthermore, a larger dataset from the Chiesi's compounds library will be analyzed with the aim of developing a holistic model capable of describing and predicting passive permeability.

11.1.1 Permeability in the Biopharmaceutics Classification System (BCS)

Permeability is a characteristic that together with solubility is part of the *Biopharmaceutics Classification System* (BCS) [83]. This classification is mainly intended for orally administered and gastrointestinal absorbed drugs. In this classification, compounds are divided into four classes:

- *Class I*: compounds with good permeability and good solubility. This is the best class, and the one sought after in drug discovery projects
- *Class II*: compounds with good permeability but low solubility. The bioavailability of these compounds depends on their degree of solvation
- *Class III*: compounds with low permeability but high solubility. Bioavailability depends on the rate of permeation even if the drug is solvated well

- *Class IV*: these are compounds with poor bioavailability because they are poorly permeable and soluble. They are the compounds with the worst characteristics.

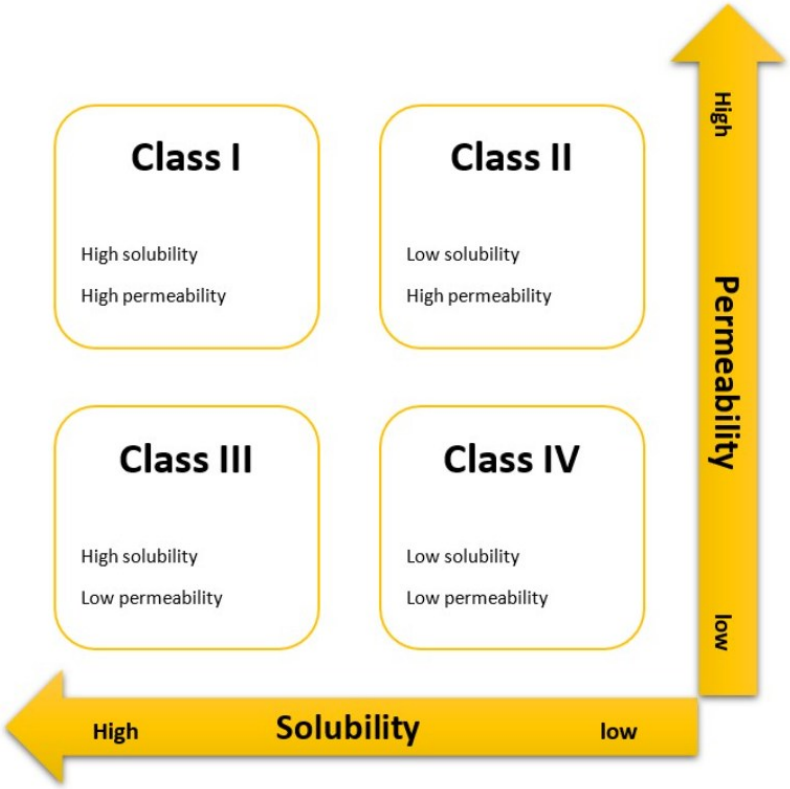


Figure 63. Biopharmaceutics Classification System (BCS)[83]

A drug is considered highly soluble when its highest single dose strength is soluble in approximately 250 mL of aqueous media across a pH range of 1.0 to 7.5. This 250 mL volume reflects typical conditions used in bioequivalence studies, where drugs are administered to fasted human volunteers with a glass of water. Permeability classification is based either indirectly on the extent of drug absorption in humans or directly on measured mass transfer rates across the human intestinal membrane. Alternatively, in vitro or non-human models may be used to predict human absorption. A drug substance is classified as highly permeable when 90% or more of the administered dose is absorbed in humans, as determined by mass balance studies or by comparison with an intravenous reference dose [84].

In the preclinical stage, various types of *in-vitro* assays are employed to evaluate permeability and efflux. These tests are particularly important in the development of orally administered drugs, as gastrointestinal permeability remains a key factor. The results obtained from in vitro assays are often correlated with *in-vivo* studies; however, they serve primarily as predictive tools and may not always fully replicate *in-vivo* outcomes. One of the most well-established and widely used *in-vitro* models

for assessing permeability is the human colorectal adenocarcinoma cell line (Caco-2), which was used in this project to evaluate compounds permeability. In recent decades, a variety of alternative artificial techniques, like PAMPA or Permeapad®, have emerged and gained popularity. Some of these approaches utilize artificial membranes, while others rely on computational methods to predict passive permeability [85].

11.1.2 Caco-2 permeability studies

One of the most important assays for evaluating intestinal permeability is the Caco-2 cell culture assay, which employs a monolayer of cells derived from a human colorectal adenocarcinoma cell line. This model enables the measurement of bidirectional transport: from the apical to the basolateral side (A>B) and from the basolateral to the apical side (B>A). The latter direction is particularly useful for identifying compounds that are substrates of P-glycoprotein (P-gp) or other active transports, which are actively secreted back into the intestinal lumen. The assay typically includes control compounds with known high and low permeability, as well as known P-gp substrates and inhibitors. These controls help determine whether the test compound is a P-gp substrate and whether it is subject to efflux transport (B>A).

Caco-2 cells exhibit many morphological and functional characteristics of human enterocytes. Although they originate from a human colorectal adenocarcinoma, Caco-2 cells have been shown to express a range of efflux transporters, microvillar transporters, hydrolases, and phase II conjugation enzymes similar to those found in small intestinal enterocytes. They also express key digestive enzymes, such as membrane-bound peptidases and disaccharidases. Importantly, during a 21-day culture period, Caco-2 cells differentiate into polarized monolayers featuring apical microvilli (brush border) and well-formed tight junctions, mimicking the structure of the intestinal epithelium.
Tight junctions play a critical role as a barrier to the paracellular absorption of orally administered drugs, while the microvillar region of enterocytes facilitates transcellular drug uptake. Although the *in-vivo* intestinal epithelium consists of various specialized cell types (e.g., endocrine, exocrine), enterocytes are the predominant absorptive cells. As a result, cell-based systems like Caco-2 monolayers have been shown to be reliable predictors of transcellular drug permeability.

As discussed in previous chapters, oral administration remains the preferred route for drug delivery. Following oral intake, a drug may be absorbed and reach systemic circulation through various transport mechanisms, both active and passive, across the intestinal epithelial cells. A major objective

in drug development is to design compounds that can be efficiently absorbed through this epithelial barrier.

Lipophilic, small, and uncharged molecules typically permeate the biological membrane with ease, favoring absorption via the transcellular pathway. In contrast, hydrophilic, large, or charged molecules, as well as most peptides, are generally unable to diffuse across cell membranes and rely predominantly on the paracellular (intercellular) route for absorption. In addition to passive diffusion, active transport mechanisms also play a crucial role, requiring energy input to move substances against their concentration gradients. These include both active transport and efflux. Essential nutrients such as amino acids, glucose, and vitamins are actively transported by specific membrane transporters. When drugs rely on these transporter-mediated pathways, their permeability in the Caco-2 model often appears lower than what is observed in vivo. This discrepancy is largely attributed to the fact that Caco-2 cells express lower levels of certain transporters compared to native enterocytes in the human intestine [86].

The presence of multiple transport pathways highlights the complexity of intestinal absorption. Consequently, in vitro studies of epithelial transport have become essential tools for improving the accuracy of absorption predictions during early-stage drug development. For these reasons, the Caco-2 assay continues to be one of the most widely used and reliable models for studying passive permeability in the preclinical phase of drug development.

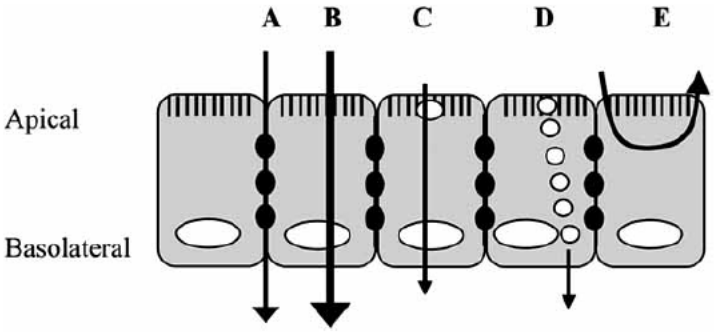


Figure 64. Representation of possible transport mechanisms through the intestinal epithelium. Following oral administration, the molecule can permeate via paracellular mechanisms between cells (A), transcellular mechanisms through the cell (B), carrier-mediated active mechanisms (C), transcytosis mechanisms via vesicles (D). There is the possibility of efflux via P-gp as a protection mechanism against xenobiotics (E) [86]

11.2 Materials

11.3 Materials used to measure the passive permeability (P_{app}) through Caco-2 cell culture permeability assay

To measure the passive permeability, expressed by P_{app} (nm/s), using the Caco-2 cell culture permeability assay, of the 19 pharmaceutical standard, the following chemical reagents were used: dimethyl sulfoxide (Sigma-Merck), elacridar (Sigma-Merck), Hank's Balance Salt Solution (Sigma-Merck), acetonitrile (Sigma-Merck), warfarin (Sigma-Merck), sulpiride (Sigma-Merck), metoprolol (Sigma-Merck), talinolol (Sigma-Merck), Caco-2 cells monolayer (MedTech Barcelona).

The following instrument and accessories were used: 1.5 mL Eppendorf vials (Eppendorf), pipettes (Sartorius 5-120 µL, 10-300 µL, 100-5000 µL), 1.5 mL vials, 4 mL vials, UPLC (Waters) coupled to mass spectrometer (Waters), 96-well plate, centrifuge, EVOM instrumentation (Endhom, WPI, Germany), robotic system STAR LAB (Hamilton).

11.4 Methods

11.5 Caco-2 cells culture permeability assay

11.5.1 Apical to basolateral assay

The assay was performed using mono stratification culture cells in 96-well plates. Cells were allowed to differentiate between 18-23 days. Transepithelial Electrical Resistance (TEER) is the technique used to verify the integrity of the cellular monolayer with the EVOM instrumentation (Endhom, WPI, Germany). The robotic system Hamilton STAR LAB was used to perform the analysis and an HPLC/MS system for the quantification of compounds.

The permeation from the apical-to-basolateral side (A > B) is measured by the apparent permeability coefficient (P_{app}). During the day of the experiment, the culture was removed by washing the cells three times using a buffer solution HBBS (Hank's Balance Salt Solution) at the temperature of 37°C. Cells were pre-incubated for 20 minutes within a buffer solution at 37°C in a 95% humidity and 5% of CO_2 atmosphere. After 20 minutes, the buffer solution was removed and in the donor compartment 80 µL of solution of compound or positive control were added, while in the basolateral compartment 235 µL of the buffer solution were added. 5 µL of the solution from donor compartment were collected and added to 145 µL of buffer solution in 96-well plate to measure the compound concentration at T_0.

Afterwards, cells were incubated for 120 minutes with an atmosphere of 95% humidity and 5% of CO_2. The incubation was done in triplicates. At the end (T_{120}), the samples from donor and receiver compartment were taken and transferred to the plate for the LC/MS quantification. 5 µL of the

solution taken from the donor compartment were added to 145 µL of buffer solution (dilution of 1:30), and 150 µL of the receiver solution were transferred in the same 96-well plate for the analysis. Before the analysis, 300 µL of acetonitrile with warfarin 10 nM as internal standard were added in each sample (T_0 and T_{120}). The plates were centrifuged (2300g for 10 minutes) and 150 µL of supernatant were used for the LC/MS quantification.

As controls, compound with known permeability were used. Sulpiride and metoprolol are compounds characterized with low permeability (<10 nm/sec) and high permeability (>100 nm/sec) respectively. Furthermore, a P-glycoprotein inhibitor (elacridar) and a substrate (talinolol) could be used to measure the real passive permeability of the compounds.

11.5.2 Basolateral to apical assay

Secretion was assessed by calculating the apparent permeability coefficient (P_{app}) of the compound in the basolateral-to-apical (B > A) direction using the Caco-2 cell monolayer model. On the day of the experiment, the culture medium was aspirated, and the cell monolayers were washed three times with pre-warmed (37 °C) Hank's Balanced Salt Solution (HBSS) to remove residual medium. The monolayers were then pre-incubated in HBSS for at least 20 minutes at 37 °C, under an atmosphere of 5% CO_2 and 95% relative humidity. Following pre-incubation, 240 µL of donor solution, containing either the test compound or a positive control at 10 µM, with or without the P-gp/BCRP inhibitor Elacridar, was added to the basolateral (B) compartment of each transwell insert. Simultaneously, 75 µL of HBSS (receiver solution) was added to the apical (A) compartment. To determine the initial concentration (T_0) of the compound, 5 µL aliquots were collected immediately after donor addition and transferred to a 96-well plate containing 145 µL of HBSS (resulting in a 1:30 dilution). The incubation was initiated upon addition of the compound and maintained at 37 °C, 95% humidity and 5% CO_2, for 120 minutes. All assays were conducted in triplicate.
At the end of the incubation (T_{120}), samples were collected from both compartments:

- Donor (basolateral): 5 µL transferred into 145 µL of HBSS (1:30 dilution)
- Receiver (apical): 50 µL transferred into 100 µL of HBSS (1:3 dilution)

To each sample (both T_0 and T_{120}), 300 µL of acetonitrile containing 10 nM of the internal standard (warfarin) was added. Samples were centrifuged at 2300g for 10 minutes, and 150 µL of supernatant was transferred to a fresh 96-well plate for analysis. Quantification was performed by LC-MS/MS.

Positive control experiments were run in parallel using talinolol as the reference compound (stock solution 10 mM in 100% DMSO, final assay concentration 10 μM with 0.1% DMSO).

11.5.3 Permeability data analysis

Once assayed, the apparent permeability (P_{app}) is calculated in nm/s, according to the following equation:

$$P_{app}\ (cm/sec) = (V_r/C_0) * (1/A) * (C_r(t)/t)$$

$$P_{app}\ (nm/sec) = P_{app}\ (cm/sec) * 1000000$$

- V_r = volume of solution in mL of receiving compartment
- C_0 = initial concentration of the analyzed compound in the donor compartment
- A = membrane surface area (0.14 cm^2)
- $C_r(t)$ = concentration measured in the receiving compartment at time T_{120}
- t = time (seconds)

When the transport of tested compound is studied in both directions, an efflux ratio is expressed which is determined by:

$$\text{Efflux ratio (ER)} = P_{app}\,(B > A) / P_{app}\,(A > B)$$

This value provides useful indications as to whether the compound under analysis may be subject to active efflux. When the ER value is greater than 2.5, it suggests that the compound can be actively secreted by efflux transporters such as the mentioned P-glycoprotein (P-gp), breast cancer resistance protein (BCRP) or other transporters.

Recovery was determined by measuring concentration of the test compound in the donor chamber at the end of experiment, then adding the receiving chamber concentration; this sum divided by the initial donor concentration yields the % recovery.

$$\text{Recovery \%} = (C_r(t) * V_r) + (C_d(t) * Vd / C_0(t) * V_d) *100$$

Where:

- $C_r(t)$, is the measured concentration of the receiver chamber at time t (expressed as area ratio)

- $C_d(t)$, is the measured concentration of the donor chamber at time t (expressed as area ratio)
- $C_0(t)$, is the measured concentration of the donor chamber at time zero
- V_r, is the volume of the receive chamber (mL)
- V_d, is the volume of the donor chamber (mL)

The following classification is normally used to subdivide compounds according to their permeability, low medium or high:

- *Low* < 10 nm/s
- *Moderate* = 10 – 100 nm/s
- *High* > 100 nm/s

11.6 Results

Table 13. Permeability values measured using the Caco-2 assay in both apical-to-basolateral (A > B) and basolateral-to-apical (B > A) directions, reported as the average of two interday replicates (AVG). Measurements were performed in the absence (Control) and presence (GF) of the P-glycoprotein (P-gp) and BCRP inhibitor GF120918 (Elacridar). The table also includes the efflux ratio (EF) and recovery percentage (Recovery %) for each compound

Compound	WO/W	AVG A>B (nm/s)	AVG B>A (nm/s)	EF	Recovery %
ACTH fragment	Control	BLOQ	BLOQ	n/a	n/a
	GF	BLOQ	BLOQ	n/a	
Angiotensin II	Control	BLOQ	BLOQ	n/a	n/a
	GF	BLOQ	BLOQ	n/a	
Caspofungin diacetate	Control	0.9	0.7	0.7	72.3
	GF	1.9	0.8	0.6	
Cetrorelix acetate	Control	0.1	0.1	1.7	88.4
	GF	0.1	0.25	4	
Desmopressin	Control	0.7	1.8	2.4	100
	GF	0.7	0.7	0.9	
Erythromycin	Control	0.7	71.6	97.8	101.4
	GF	6.8	18	3.1	
Everolimus	Control	36.7	28.3	0.7	70
	GF	35.2	13.3	0.3	
Goserelin	Control	0.2	0.2	1.1	100
	GF	5.7	1.5	1.1	
Lanreotide acetate	Control	0.3	0.3	0.9	96.4
	GF	0.8	0.25	0.1	
Octreotide acetate	Control	0.5	0.7	1.9	100
	GF	1.9	1.2	0.6	
Oxytocin	Control	0.2	0.5	2.2	70.6

	GF	0.7	0.4	0.5	
Rifampicin	Control	25.4	139.9	5.5	100
	GF	32.1	83.7	1.8	
Rifapentin	Control	120.2	106.3	0.8	82.6
	GF	109.5	93.8	0.7	
Rifaximin	Control	1.2	204.3	168.9	82.9
	GF	7.9	93.8	11.1	
Ritonavir	Control	41.4	193.6	4.6	80.1
	GF	107.5	81.2	0.5	
Roxithromycin	Control	1.9	126.6	65.8	100
	GF	31.8	40.2	1.1	
Saquinavir mesilate	Control	3.6	151.2	41.6	70
	GF	61.7	68	0.9	
Teicoplanin	Control	0.4	0.2	0.4	96.5
	GF	3	0.9	0.2	
Vancomycin	Control	0.3	0.5	1.66	74
	GF	0.5	0.4	0.8	

In the context of the permeability studies performed using Caco-2 cell cultures, both in the presence and absence of a P-glycoprotein (P-gp) inhibitor (GF120918 - Elacridar), two compounds, ACTH fragment, angiotensin II, were found to be below the limit of quantification (BLOQ). This is likely due to their inherently low permeability. Additionally, these compounds may exhibit a tendency to adhere to the surface of the assay plates, possibly due to interactions with the plastic material of the 96-well plates used in the assay, or instability under assay conditions so the recovery values were not determined (n/a). A threshold value of 0.1 nm/s² was used to define low permeability of these compounds.

The apical-to-basolateral (A > B) transport rate was used to calculate $P_{app,}$ Seven compounds showed low permeability even in the presence of the P-gp inhibitor: cetrorelix (0.1), desmopressin (0.7), oxytocin (0.7), lanreotide (0.8), caspofungin (1.9), octreotide (1.9), teicoplanin (3.0), goserelin (5.7), erythromycin (6.8) and rifaximin (7.9). These include six peptides (synthetic and natural), two macrocycles, one cyclic glycopeptide and one cyclic lipopeptide.

In contrast, roxithromycin (31.8), rifampicin (32.2), everolimus (35.2), saquinavir (61.7), ritonavir (107.5) and rifapentine (109.5), four macrocycles and two peptidomimetics, showed moderate permeability, with saquinavir, ritonavir and rifapentine exceeding 50 nm/s, indicating moderate-to-high permeability and ritonavir and rifapentine exceeding 100 nm/s, indicating high permeability.

Significant changes in Efflux Ratio (EF) were observed when comparing assays with and without the P-gp inhibitor, especially for oxytocin (2.2 *vs* 0.5), desmopressin (EF 2.4 *vs* 0.9), ritonavir (EF 4.6 *vs* 0.5), rifampicin (EF 5.5 *vs* 1.8), saquinavir (EF 41.6 *vs* 0.9), roxithromycin (EF 65.8 *vs* 1.1), erythromycin (EF 97.8 *vs* 3.1) and rifaximin (EF 168.9 *vs* 11.1) indicating strong involvement of P-gp-mediated efflux. For some compounds, such as erythromycin (EF 3.1), and rifaximin (EF 11.1), the EF value remained above 2 even in the presence of the inhibitor, suggesting the involvement of other active transport mechanisms, as reported in literature [87], [88]. If the other transports implicated in the efflux are inhibited, the permeability data increases a bit, highlighting how the permeability data in the context of this PhD work, for these two compounds, are influenced by the presence of other efflux transporters [87], [88].

Recovery rates were all above 70%, ranged from 70% for saquinavir to 100% for six compounds: desmopressin, erythromycin, goserelin, octreotide, rifampicin, and roxithromycin, supporting the reliability of the experimental assay.

11.6.1 Polarity in the permeability prediction of bR05 compounds

As discussed in previous chapters, polarity is one of the most important physico-chemical properties influencing the passive permeability of compounds. In the context of this PhD project, the predictive ability of TPSA and EPSA for permeability was evaluated. It is important to note that the dataset is not fully representative of all bRo5 compounds; however, efforts were made to include a range as broad as possible within this chemical space, taking into account also the cost that this type of molecule has. The dataset comprises three endogenous peptides, seven synthetic peptidomimetics, six macrocycles, with two cyclic glycopeptides and one cyclic lipopeptide.

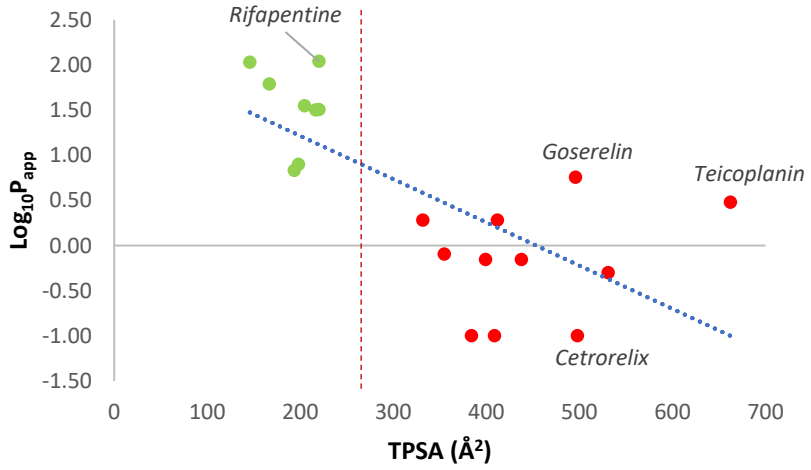


Figure 65. Correlation between TPSA and $Log_{10}P_{app}$ for the 19 bRo5 pharmaceutical standards. Green dots are compounds orally delivered while the red dots are compounds not orally delivered $R^2 = 0.48$; $y = -0.005*TPSA + 2.17 \pm 0.75$

The correlation between TPSA and $Log_{10}P_{app}$ in Figure 65, appears to be stronger than initially expected, with an R^2 value of 0.48. A clear permeability cutoff is observable above 200-250 Å², which is higher than the commonly reported threshold of 120-140 Å² typically used to differentiate permeable from non-permeable compounds. It is important to note that for ACTH fragment, and angiotensin II, the P_{app} value was set at 0.1 nm/s because all assay results were below the limit of quantification (BLOQ). This indicates their extremely low permeability and challenges for assay performance with these types of compounds. Four additional outliers in the correlation plot are cetrorelix, teicoplanin, goserelin, and rifapentine. While cetrorelix, teicoplanin and goserelin show low permeability consistent with their very high TPSA values (498, 662 and 496 Å², respectively), rifapentine displays unexpectedly high permeability despite having a TPSA of 220 Å². This could mean that TPSA alone does not explain everything: rifapentine has high lipophilicity and a rigid structure so it can pass by transcellular diffusion. Also, the other macrocycles such as rifampicin and roxithromycin show higher TPSA value than the cutoff normally considered but all these macrocycles are characterized by high lipophilicity (consideration about the propensity to form IMHBs will be discussed in the next chapter).

Overall, most compounds follow the general trend of the correlation, showing a clear distinction between low- and high-permeability profiles, and interestingly between orally delivered drugs (green dots) that are erythromycin, everolimus, rifampicin, rifapentine, rifaximin, ritonavir, roxithromycin and saquinavir and not orally delivered drugs (red dots). However, the TPSA threshold that separates these classes appears to shift upward in our dataset compared to traditional models, suggesting that for this particular chemical space (e.g., bRo5-like structures), higher TPSA values may still allow for relevant passive permeability, especially when the compounds show moderate-high lipophilicity. This concept is clearly explained in the article of Doak et al., where for bRo5 the presence of numerous HBA and HBD increases the overall polarity but could promote the formation of IMHBs that decrease the exposed polarity of these compounds, promoting the passage through biological membrane [60].

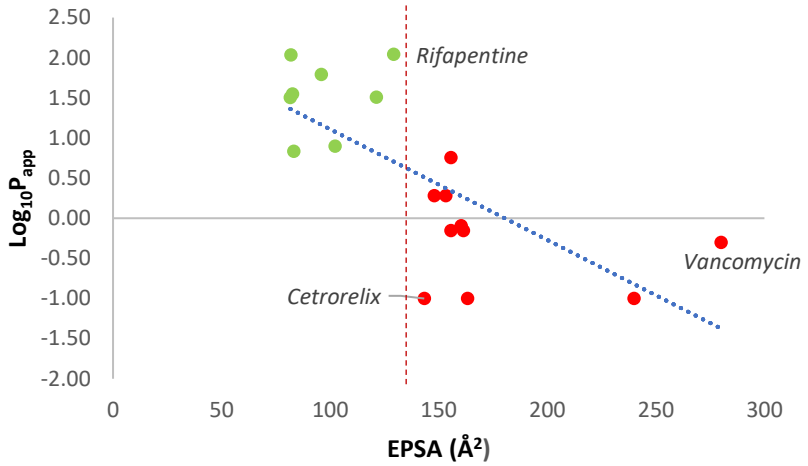


Figure 66. Correlation between TPSA and $Log_{10}P_{app}$ for the 19 bRo5 pharmaceutical standards. Green dots are compounds orally delivered while the red dots are compounds not orally delivered $R^2 = 0.51$; $y = -0.013*EPSA + 2.49 \pm 0.76$

The correlation between EPSA and $Log_{10}P_{app}$ in Figure 66 also appears stronger than initially expected, with an R^2 value of 0.51, comparable to that observed for TPSA. In this correlation plot as well, a clear permeability cutoff is visible above 130 Å², which is slightly higher than the commonly reported threshold of 100 Å² typically used to distinguish permeable from non-permeable compounds. As with the previous analysis, the P_{app} values for ACTH fragment, and angiotensin II were fixed at 0.1 nm/s due to all assay results being below the limit of quantification (BLOQ), reflecting their extremely low permeability. Teicoplanin, rifapentine and cetrorelix again appear as outliers. Vancomycin and cetrorelix show low permeability consistent with their very high EPSA value (280 and 144 Å²), whereas rifapentine exhibits unexpectedly high permeability despite having an EPSA of 129 Å². Overall, the majority of compounds follow the general correlation trend, revealing a clear separation between compounds with low and high permeability. However, the EPSA threshold that distinguishes these two classes appears to be shifted upward in this dataset compared to conventional models, suggesting that within this specific chemical space (e.g., bRo5-like compounds), higher EPSA values may still be compatible with relevant passive permeability, especially in case of moderate-high permeability. In addition, the distinction between orally delivered compounds (green dots) and not orally delivered compounds (red dots) is as evident as for TPSA.

11.6.2 Permeability prediction of bR05 compounds using in silico lipophilicity

In addition to polarity, lipophilicity is another key physico-chemical property that significantly influences drug permeability. Achieving a proper balance between polarity and lipophilicity is crucial

for optimizing both permeability and bioavailability, ultimately enabling the compound to reach its desired target.

Within the framework of this PhD project, the lipophilicity of 19 pharmaceutical standards was evaluated and correlated with passive permeability, as measured in Caco-2 cell monolayers in the presence of a P-gp/BCRP inhibitor. Specifically, three computational lipophilicity values (XLog *P*, CLog *P*, and CLog $D_{7.4}$) calculated using the ACD/ Labs Percepta 2019 software and three experimental values obtained through different methodologies such as Log P_{oct} (pH-metric assay), Log $D_{7.4}$ (*shake-flask assay*) and CHI Log $D_{7.4}$ (chromatographic assay) were compared with passive permeability to assess their individual predictive power because they represent different facets of lipophilicity.

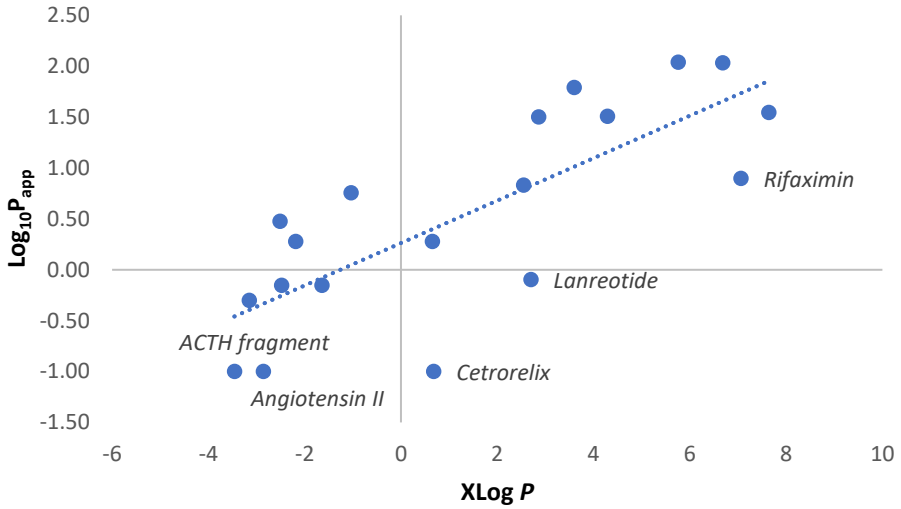


Figure 67.Correlation between XLog P and $Log_{10}P_{app}$ for the 19 bRo5 pharmaceutical standards
*R^2 = 0.61; y = 0.21*XLog P + 0.26 ±0.65*

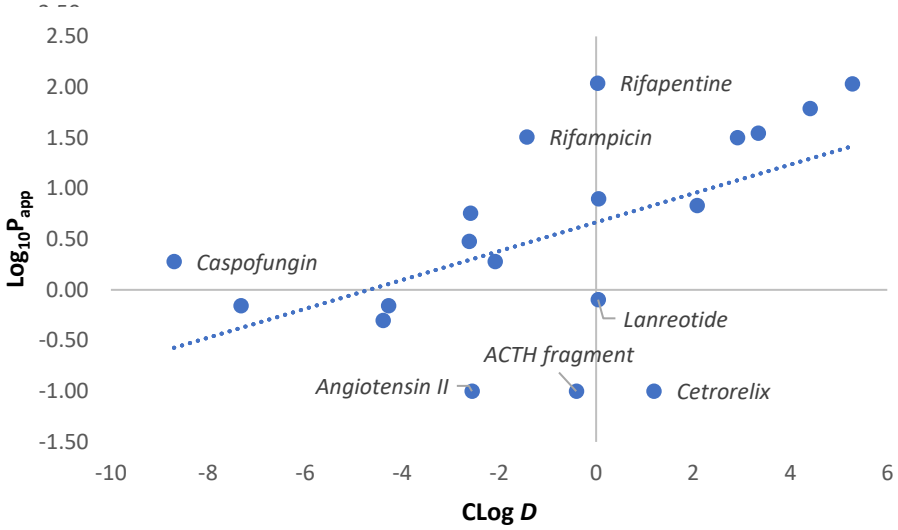


Figure 69. Correlation between CLog $D_{7.4}$ and $Log_{10}P_{app}$ for the 19 bRo5 pharmaceutical standards
*R^2 = 0.27; y = 0.1*CLog $D_{7.4}$ + 0.67 ±0.89*

The *in-silico* descriptors used to estimate the lipophilicity, and indirectly, the permeability, of the 19 bRo5 pharmaceutical standards in this study are XLog *P*, CLog *P*, and CLog $D_{7.4}$. Although these parameters are all computational estimations of lipophilicity, they differ in their underlying calculation methods.

XLog *P* is calculated using an atom-based approach, summing the contribution of individual atoms in the molecule. While straightforward, this method is generally less reliable for complex or flexible structures, such as those commonly found in the bRo5 chemical space.

CLog *P*, in contrast, uses a fragment-based approach, incorporating contributions from molecular substructures (fragments) and applying correction factors based on known intramolecular interactions. This typically makes it more accurate than XLog *P*, although its reliability depends on the quality and scope of the fragment library used.

CLog $D_{7.4}$ (at pH 7.4) extends the CLog *P* approach by also integrating the compound's pK_a values, thereby considering the ionization state of the molecule at physiological pH. As a result, CLog $D_{7.4}$ should provide a more accurate estimation of lipophilicity under biological conditions, even if its reliability is strongly related to the accuracy of the predicted or assigned pK_a values.

Interestingly, in our dataset, XLog *P* appears to correlate more strongly with passive permeability than the other two lipophilicity descriptors, showing an R^2 value of 0.61 (Figure 67), compared to 0.20 for CLog *P* (Figure 68) and 0.27 for CLog $D_{7.4}$ (Figure 69). Among the three, CLog *P* performed the worst in terms of predictive power. As previously mentioned, the permeability values for ACTH fragment and angiotensin II were consistently below the limit of quantification (BLOQ), likely due to their extremely poor ability to permeate biological membranes. For statistical purposes, a fixed

P_{app} value of 0.1 nm/s was assigned to these two compounds. As evident in the correlation plots, these molecules act as strong outliers, particularly affecting the correlation for CLog *P* and CLog $D_{7.4}$, and significantly lowering the R^2 values.
When these two outliers are excluded from the analysis, the correlation improves, especially for the descriptors previously showing weak performance. The adjusted R^2 values decrease to 0.53 for XLog *P* (minimal change), and increase minimally to 0.30 for CLog *P*, and 0.35 for CLog $D_{7.4}$. Although XLog *P* still demonstrates the highest correlation, the performance gap between the descriptors becomes less pronounced.

These results suggest that, at least within this specific set of bRo5 compounds, XLog *P* surprisingly outperforms the theoretically more advanced descriptors like CLog $D_{7.4}$. This may be partly explained by the fact that the Caco-2 permeability assay is conducted under a fixed pH of 7.4, which may limit the benefit of simulating pH-dependent distribution, particularly if pK_a predictions are not highly accurate. Computational pK_a estimations are often associated with a certain degree of error, especially when applied to large, flexible structures with multiple ionizable functional groups, as is often the case for bRo5 compounds. Since the calculation of CLog $D_{7.4}$ relies heavily on the accuracy of predicted pK_a values, even small deviations can significantly affect the resulting lipophilicity estimate at physiological pH. This likely contributes to the reduced predictive power of CLog $D_{7.4}$ compared to simpler descriptors such as XLog *P*, which do not take ionization into account. Indeed, lanreotide, rifampicin and rifapentine are the compounds with the highest discrepancy between the experimental and calculated pK_a, and it could be reflected in the correlation between CLog $D_{7.4}$ and passive permeability data.
Following the correlation analysis between the different types of *in-silico* lipophilicity descriptors and permeability, the relationship between experimental lipophilicity measurements and permeability was also evaluated. In the context of this PhD project, three experimental lipophilicity values were determined: Log P_{oct} using the pH-metric method, Log $D_{7.4}$ using the shake-flask technique, and CHI Log $D_{7.4}$ using a chromatographic approach.

11.6.3 Permeability prediction of bR05 compounds using experimental lipophilicity

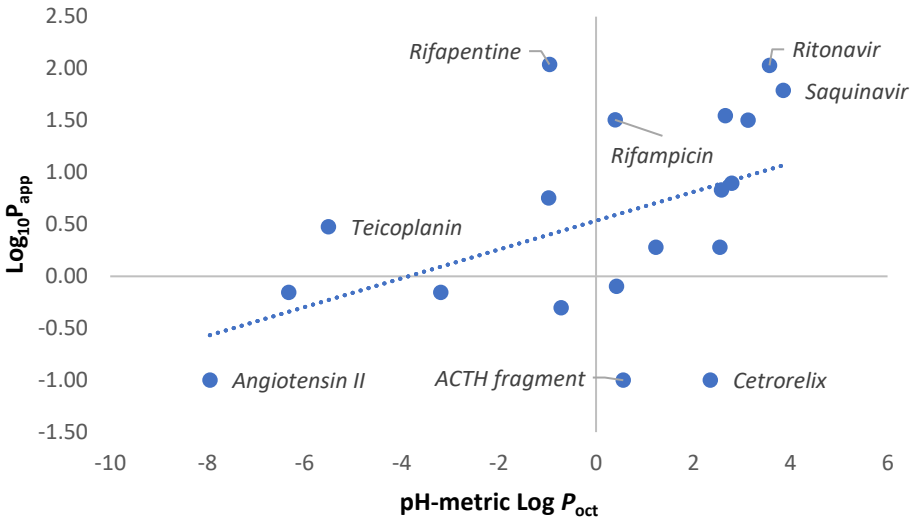


Figure 70. Correlation between pH-metric Log Poct and Log10P_{app} for the 19 bRo5 pharmaceutical standards
*R^2 = 0.23; y = 0.14*pH-metric Log P_{oct} + 0.54 ±0.92*

Log P_{oct}, which represents the partition coefficient in a biphasic system composed of *n*-octanol and an aqueous buffer solution, describes the lipophilicity of the neutral form of a compound. In this study, it was determined using a pH-metric method with the SiriusT3™ instrument. Two major outliers were identified in the dataset: ACTH fragment and cetrorelix, with Log P_{oct} values of 0.56 and 2.35, respectively. As mentioned earlier, some compounds formed unusual emulsions at the interface during the assay, likely due to their amphiphilic nature and the intense stirring applied during the experiment, although the instrument's sensor did not detect any precipitation or solubility issues. Another critical factor is the use of fixed permeability values for compounds such as ACTH fragment and angiotensin II, due to their concentrations being below the quantification limit in the Caco-2 assay.

Teicoplanin was another compound that showed interfacial emulsion formation, which likely affected the accuracy of the Log P_{oct} value. Despite these outliers, the remaining compounds showed a general trend of increasing permeability with increasing Log P_{oct}. The correlation coefficient (Figure 70) was relatively low (R^2 = 0.23).

Although the pH-metric assay is highly effective for small molecules and moderately to highly lipophilic compounds (Log *P* 1-3), it shows significant limitations when applied to bRo5 or structurally complex molecules, which often contain multiple ionizable groups and amphiphilic regions. For such compounds, this technique may not be the most reliable for accurately characterizing lipophilicity.

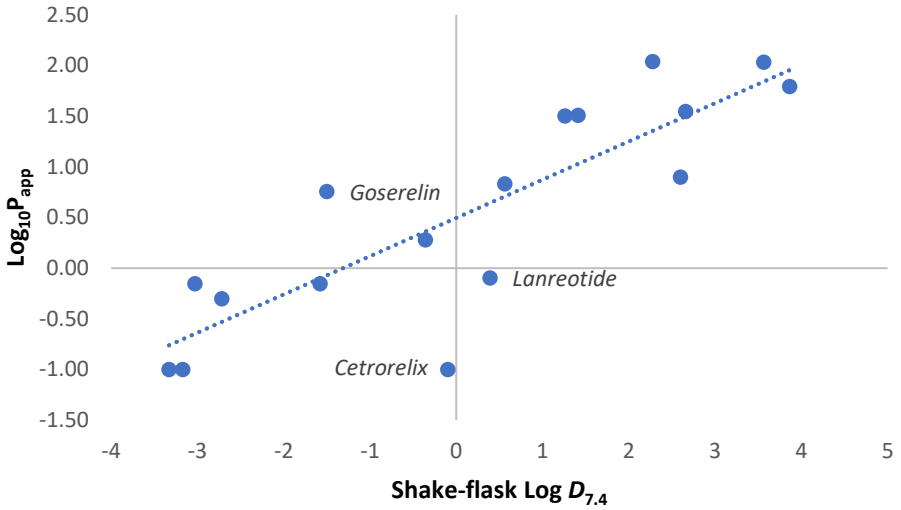


Figure 71. Correlation between shake-flask Log $D_{7.4}$ and $Log_{10}P_{app}$ for the 19 bRo5 pharmaceutical standards
*R^2 = 0.72; y = 0.38*shake-flask Log $D_{7.4}$ + 0.50 ±0.59*

A different scenario was observed for the correlation between permeability and Log $D_{7.4}$ measured using the shake-flask assay (Figure 71). In this case, the correlation was stronger, with an R^2 value of 0.72. Although the shake-flask method is more time-consuming, taking up to two days when including phase pre-saturation, stirring, and HPLC analysis, it applies gentler stirring compared to the pH-metric technique. Notably, no emulsion formation was observed with this method. Only two compounds, teicoplanin and caspofungin (Figure 51 and Figure 52), exhibited interfacial accumulation, preventing the determination of reliable Log $D_{7.4}$ values for these molecules.

Cetrorelix, although not showing clear interfacial problems or emulsions during the assay, still behaved as a significant outlier. This may be due to potential interactions with plastic surfaces (e.g., Eppendorf tubes) during the experiment.

Overall, a clearer and more consistent correlation was observed, in line with theoretical expectations. Unlike Log P_{oct}, which measures the lipophilicity of the neutral form only, Log $D_{7.4}$ accounts for the distribution of all ionized and unionized species at physiological pH. Given that most of the compounds in the dataset are ionizable and contain multiple ionizable groups, the conditions of the *shake-flask* assay at pH 7.4 more closely resemble those of the Caco-2 cell assay. This alignment likely contributes to the improved correlation between Log $D_{7.4}$ and permeability.

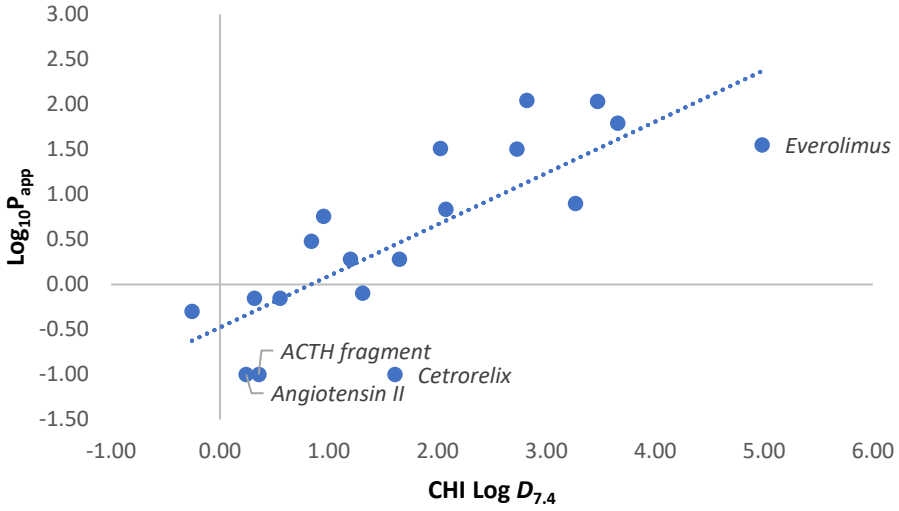


Figure 72. Correlation between CHI Log $D_{7.4}$ and $Log_{10}P_{app}$ for the 19 bRo5 pharmaceutical standards
*R^2 = 0.62; y = 0.57*CHI Log $D_{7.4}$ - 0.48 ±0.64*

As discussed in the previous chapter, chromatographic methods offer several advantages over traditional techniques for lipophilicity assessment. These include lower compound consumption, significant time savings, and the elimination of issues related to interfacial accumulation, which commonly affect pH-metric and *shake-flask* assays.

In the lipophilicity evaluation section, where all lipophilicity values, both *in-silico* and experimental, were compared using correlation plots, the highest correlation was observed between shake-flask Log $D_{7.4}$ and chromatographic CHI Log $D_{7.4}$ (R^2 = 0.86). *Shake-flask* Log $D_{7.4}$ also showed a good correlation with passive permeability measured in the Caco-2 assay (R^2 = 0.72), while CHI Log $D_{7.4}$ showed a slightly lower but still acceptable correlation (R^2 = 0.62), Figure 72. As in previous analyses, the two compounds with fixed permeability values (assigned due to BLOQ results in the permeability assay) and cetrorelix, acted as outliers and negatively affected the overall correlation.

Everolimus emerged as a notable outlier in the correlation between *shake-flask* Log $D_{7.4}$ and CHI Log $D_{7.4}$. It displayed an unusually high CHI value, suggesting strong interactions with the chromatographic column. Interestingly, everolimus also deviated slightly from the general trend in the correlation between CHI Log $D_{7.4}$ and permeability. This behavior may be attributed to its fully neutral character: it is the only compound in the dataset without ionizable groups at pH 7.4. This observation highlights the possibility that neutral compounds may engage in stronger non-ionic interactions with the stationary phase at pH 7.4, unlike ionizable compounds, which may remain partially or fully charged under the same conditions.

11.6.4 Categorical classification based on polarity and lipophilicity

Based on the correlation discussed in the previous chapter, it is evident that EPSA and TPSA (for polarity), as well as Log $D_{7.4}$ measured both by *shake-flask* and chromatographic CHI Log $D_{7.4}$, show a stronger correlation with permeability than other commonly used physico-chemical descriptors. This is particularly true when compared to calculated lipophilicity values such as XLog *P*, CLog *P*, CLog $D_{7.4}$, and pH-metric Log *P*. In the polarity context the separation between permeable and no-permeable compounds is crystal clear even if the linear correlation is below $R^2 = 0.50$.

These descriptors are typically interpreted using threshold ranges that aim to distinguish permeable from non-permeable compounds. For instance, as previously discussed, the TPSA range most often associated with passive permeability is between 60 and 90 Å², while values above this threshold generally indicate reduced permeability. For bRo5 compounds, however, some studies suggest that the acceptable TPSA threshold can be extended up to 140 Å² due to the presence of intramolecular interactions and conformational flexibility [51], [58], [59], [60]. Similarly, an EPSA value below 100 Å² has been proposed in several works as a useful cutoff to identify potentially permeable compounds [14], [16], [18].

A similar rationale is applied to lipophilicity: values of Log $D_{7.4}$ around 2-4, whether determined by the *shake-flask* method or CHI chromatographic method, are generally considered optimal for ensuring sufficient membrane permeability.

11.6.5 Stratification based on EPSA and TPSA

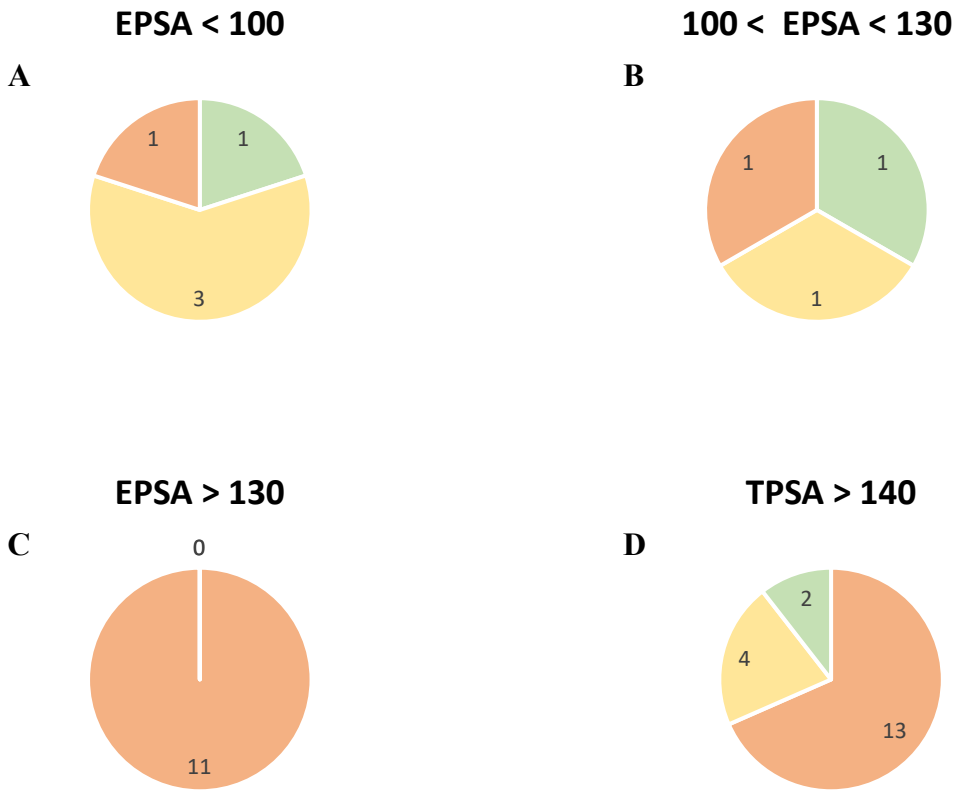


Figure 73. Permeability stratification based on EPSA < 100 Å (A), EPSA between 100 and 130 Å (B), EPSA > 130 Å (C) and TPSA > 140 Å (D). Red indicates compounds with low permeability (Papp < 1nm/s), orange indicates compounds with medium permeability (10 < Papp < 100 nm/s) and green indicates compounds with high permeability (Papp > 100 nm/s).

Based on the reference thresholds previously discussed, it is particularly noteworthy that all the compounds analyzed in this PhD project display TPSA values greater than 140 Å² (Figure 73-D). According to the established cut-off, these compounds should theoretically be non-permeable. However, six of them, everolimus, rifampicin, roxithromycin, saquinavir (moderate permeability), ritonavir and rifapentine (high permeability), demonstrate measurable permeability. This suggests that TPSA, while widely used as a descriptor for polarity and permeability prediction, is insufficient to reliably distinguish between permeable and non-permeable bRo5 compounds in this dataset.

In contrast, EPSA appears to provide a more discriminative metric. Among the compounds analyzed, five have EPSA values below 100 Å², three of which (everolimus, roxithromycin, and saquinavir) show moderate permeability, and one (ritonavir) shows high permeability. Only one compound in this group, erythromycin, has low permeability. Interestingly, erythromycin also shows an efflux ratio >3 even in the presence of efflux inhibitors, suggesting involvement of additional active efflux mechanisms. This implies that its intrinsic permeability might be higher than what was

experimentally measured, reinforcing the idea that EPSA < 100 could still be predictive of good permeability potential.

Among the 14 compounds with EPSA > 100 Å², 11 exhibit low permeability and EPSA values > 130 Å², supporting the threshold's predictive utility set up to 130 Å². The remaining three compounds, rifampicin, rifapentine, and rifaximin, show moderate, high, and low permeability, respectively. Notably, rifaximin also presents an efflux ratio > 2, which may again indicate active efflux mechanisms. As such, its true permeability could also be underestimated in the present study. Overall, these observations suggest that while TPSA fails to meaningfully distinguish permeable compounds within the bRo5 chemical space, EPSA offers a more reliable polarity descriptor, particularly when interpreting deviations through the lens of active transport processes.

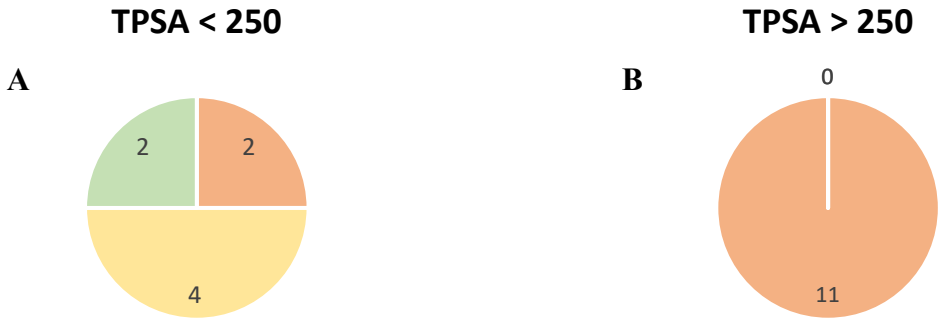


Figure 74. Permeability stratification based on TPSA < 250 Å (A) and TPSA > 250 Å (B).

Based on the analysis by Doak et al., bRo5 molecules are characterized by a higher molecular weight compared to conventional small molecules. This increase in size is typically accompanied by a higher number of hydrogen bond donors (HBD) and acceptors (HBA), resulting in increased overall polarity. Consequently, the polar surface area (PSA) of these molecules is often elevated. In their study, Doak et al. propose an extension of the traditional PSA cutoff from 140 Å² up to 250 Å² to better account for orally bioavailable bRo5 compounds [60].

Interestingly, when analyzing permeability data in relation to TPSA, it is evident that all compounds with a PSA above 250 Å² exhibit low permeability. In contrast, compounds with a TPSA below 250 Å² display a wider range of permeability profiles (Figure 74). For example, everolimus, rifampicin, roxithromycin, and saquinavir demonstrate moderate permeability, while ritonavir and rifapentine exhibit high permeability. Only two compounds with low permeability fall within the PSA < 250 Å² range: erythromycin and rifaximin. However, as previously discussed, the measured permeability of these two molecules may be underestimated due to limitations in the assay, specifically the inability of the P-glycoprotein inhibitor elacridar to fully block the active efflux transporters involved in their

disposition. Therefore, the true passive permeability of erythromycin and rifaximin could be higher than reported in the context of this PhD work.

As discussed in the polarity introduction chapter, TPSA is typically calculated by considering the contribution of nitrogen and oxygen atoms, as well as the hydrogen atoms directly bonded to them. In certain cases, sulfur atoms, particularly those involved in polar functionalities such as sulfoxides or sulfones, may also be included in the calculation.

However, the inclusion of sulfur has a minimal impact in the present dataset, as only six compounds among those analyzed contain sulfur atoms in their structure: ACTH fragment 4-10 (409 Å²), desmopressin (489 Å²), lanreotide (406 Å²), octreotide (383 Å²), oxytocin (450 Å²), and ritonavir (202 Å²). When sulfur is included in the TPSA calculation, the resulting values are slightly higher, but the difference does not significantly alter the overall trends or conclusions of this work.

11.6.6 Stratification based on shake-flask Log $D_{7.4}$ and CHI Log $D_{7.4}$

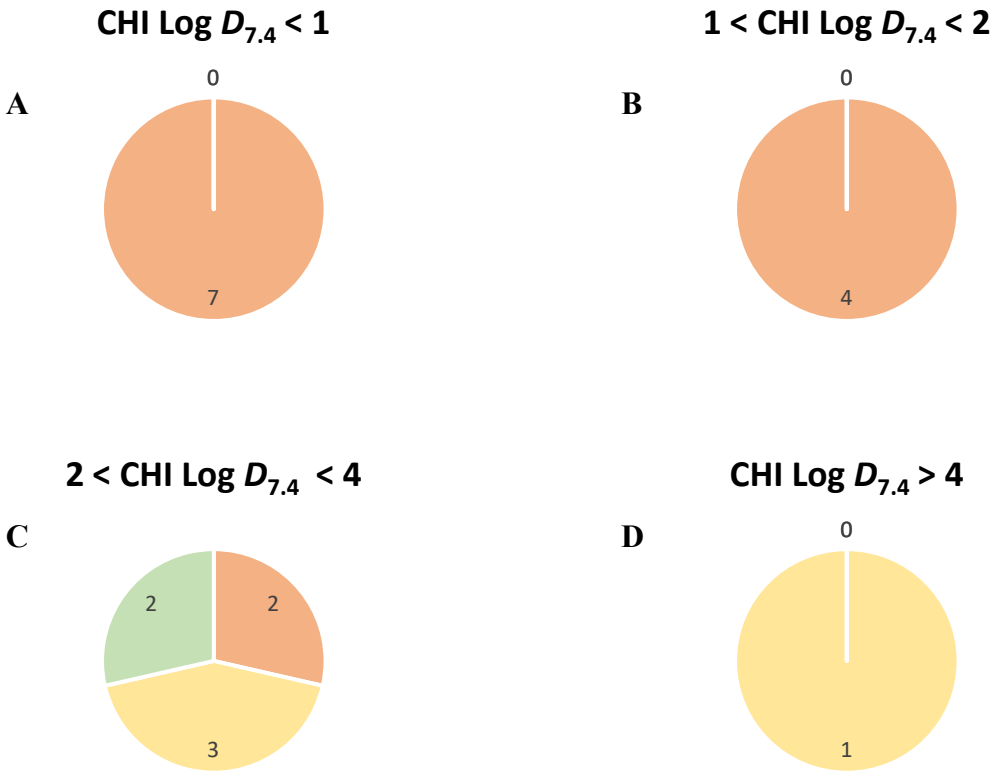


Figure 75. Permeability stratification based on CHI < 1 (A), CHI between 1 and 2 (B), CHI between 2 and 4 (C) and CHI > 4 (D)

While EPSA and TPSA values are not directly comparable due to differing measurement principles and units, EPSA typically yielding lower values, the situation is different for chromatographic and *shake-flask* lipophilicity. These two methods produce values on a comparable scale, allowing for the

application of similar interpretive ranges when assessing membrane permeability. Lipophilicity values between 2 and 4 are generally associated with favorable permeability, while values above 4 may reduce permeability due to excessive lipophilicity and membrane trapping. Conversely, values below 2 often indicate increased hydrophilicity and low passive diffusion potential. Based on these criteria, seven compounds in our dataset show low permeability with CHI Log $D_{7.4} < 1$. In the 1-2 range, four compounds remain poorly permeable.

In the favorable range between 2 and 4, permeability improves: rifapentine and ritonavir show high permeability, while rifampicin, roxithromycin and saquinavir show moderate permeability. Interestingly, erythromycin and rifaximin, although characterized by low-permeability compounds based on its permeability data, present a more complex profile. Despite their low passive permeability, experimental data indicate high efflux ratios, suggesting active transport phenomena. These results imply that there may be substrates of efflux transporters beyond P-glycoprotein, which was selectively inhibited using elacridar in the assay. Therefore, as discussed also for polarity categorization, their limited permeability may result not only from low passive diffusion, but also from active efflux via alternative pathways not accounted for in this experimental setup.

Above 4, everolimus stands out with very high CHI values. Despite its high lipophilicity, which could potentially reduce permeability due to membrane trapping or low solubility, it maintains moderate permeability, likely due to its neutral character and structural properties that favor both insertion into and release from membranes. Another possible explanation is its too strong interaction with stationary C18 phase during the CHI measurements (Figure 75).

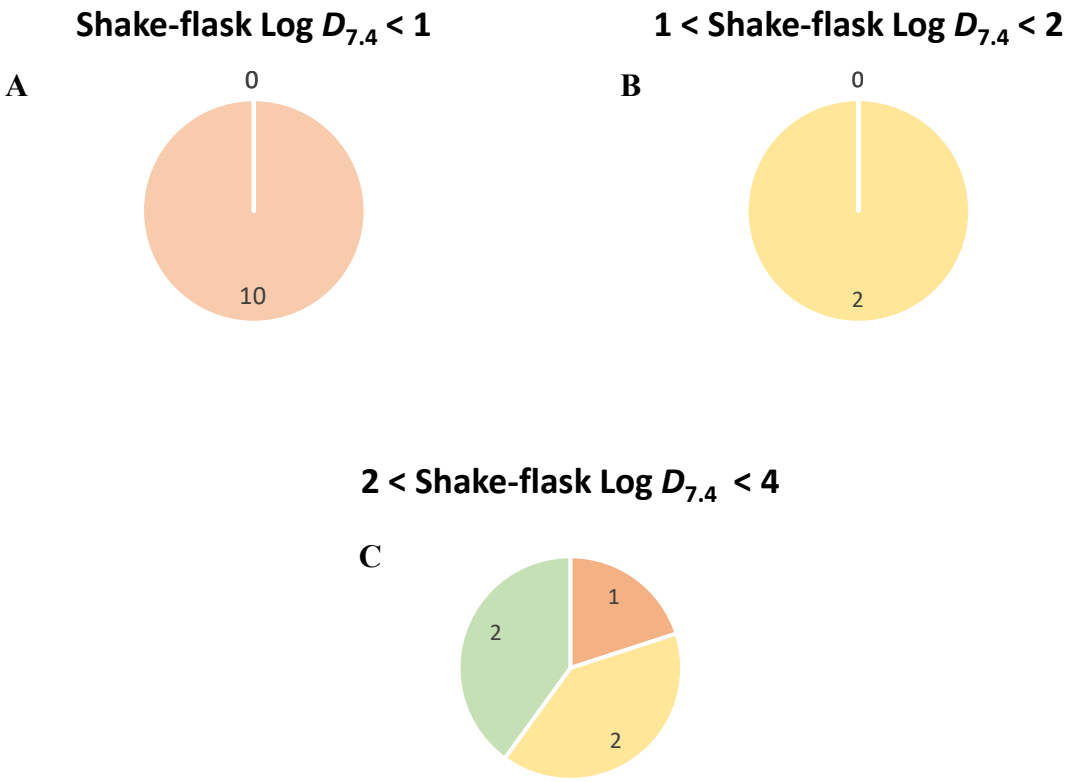


Figure 76. Permeability stratification based on shake-flask Log $D_{7.4}$ < 1 (A), between 1 and 2 (B), between 2 and 4 (C)

Using shake-flask Log $D_{7.4}$ data to categorize compounds by their permeability, the same classification range was applied: < 1 for hydrophilic compounds, 1-2 for moderately lipophilic compounds, and 2-4 for lipophilic compounds. No compounds in our dataset exhibit Log $D_{7.4}$ values above 4. In the < 1 range, ten compounds showed very low permeability, consistent with their high hydrophilicity. In the 1-2 range, two compounds displayed moderate permeability (roxithromycin and rifampicin), which aligns with expectations that increasing lipophilicity generally enhances passive membrane diffusion. In the 2-4 Log $D_{7.4}$ range, two compounds (rifapentine and ritonavir) exhibited high permeability, two (everolimus and saquinavir) showed moderate permeability, and one compound (rifaximin) showed unexpectedly low permeability. For rifaximin, this discrepancy may be explained by efflux transport: literature reports indicate that rifaximin is a substrate of multiple efflux transporters, some of which are not inhibited by elacridar, which was the only inhibitor used in this study. Therefore, its true passive permeability may be higher than the measured value. Erythromycin, on the other hand, falls into the <1 category with a shake-flask Log $D_{7.4}$ of 0.56, consistent with low permeability. However, this is in strong contrast to its chromatographic lipophilicity, which appears much higher. This discrepancy may be due to strong interactions with

the stationary phase in the chromatographic system, which are not mirrored in the *n*-octanol-water partitioning at pH 7.4 (Figure 76).

11.7 Conclusions

This PhD project aimed to explore the relationships between various physico-chemical descriptors and passive permeability for 19 bRo5 pharmaceutical standards, using the Caco-2 cell model with P-gp/BCRP inhibition. The study integrated both *in-silico* and experimental polarity and lipophilicity parameters to assess their individual and combined predictive performance.

The polarity descriptors evaluated included TPSA and EPSA. TPSA, although widely used in permeability prediction for small molecules, demonstrated limited predictive power for bRo5 compounds in this dataset. All compounds studied exhibited TPSA values above 140 Å², a level at which conventional models would predict poor permeability. However, six compounds (everolimus, rifampicin, roxithromycin, saquinavir, ritonavir, and rifapentine) displayed moderate to high permeability despite exceeding this threshold. In contrast, EPSA provided a clearer separation between permeable and non-permeable compounds. Compounds with EPSA below 130 Å² (e.g., everolimus, roxithromycin, ritonavir, saquinavir) generally exhibited measurable permeability. Most compounds with EPSA above 130 Å² were poorly permeable, supporting the utility of this descriptor in the bRo5 space.

A lipophilic analysis revealed that experimental methods outperformed computational estimates in predicting permeability. The shake-flask Log $D_{7.4}$ showed the strongest correlation with passive permeability ($R^2 = 0.72$), followed by CHI Log $D_{7.4}$ ($R^2 = 0.62$). Only the Log P_{oct} measured using potentiometry showed low correlation ($R^2 = 0.23$) highlighting the limits of this technique in lipophilicity measurement for bRo5 molecules. Among computational descriptors, XLog *P* showed the best correlation ($R^2 = 0.61$), outperforming CLog *P* ($R^2 = 0.20$) and CLog $D_{7.4}$ ($R^2 = 0.27$). Notably, the exclusion of the two low-permeability outliers (ACTH fragment and angiotensin II) improved the correlations for CLog *P* and CLog $D_{7.4}$ ($R^2 = 0.30$ and 0.35 respectively), although XLog *P* remained the most predictive. This outcome likely reflects the limitations of pK_a predictions in CLog $D_{7.4}$ and the simplicity of XLog *P*, which, although less refined, proved more reliable within this dataset.

The study also highlighted the role of active transport processes in modulating permeability. Compounds such as erythromycin and rifaximin demonstrated unexpectedly low permeability values

on the basis of polarity and lipophilicity data, but with high efflux ratios (> 3 and > 11 respectively, even in the presence of inhibitors), suggesting that active transport significantly underestimated their true passive permeability.

Overall, the findings emphasize the importance of combining polarity and lipophilicity descriptors, particularly EPSA and experimental Log $D_{7.4}$, for a more reliable prediction of passive permeability in bRo5 compounds. Traditional cutoffs for small molecules, such as TPSA < 140 Å², fail to accurately capture the permeability potential of larger, flexible structures capable of forming intramolecular hydrogen bonds or adopting conformations that mask polar surface area. The data align with the observations of Doak et al., who proposed an extension of PSA thresholds up to 250 Å² for orally bioavailable bRo5 molecules [60].

It remains unclear whether the observed permeability of certain bRo5 compounds with TPSA values above traditional thresholds (i.e., >140 Å²) can be generalized to the broader chemical space, or if this finding is specifically related to their ability to form intramolecular hydrogen bonds (IMHBs), which reduce the exposure of polar surface area and facilitate passive diffusion. Unfortunately, the current dataset lacks compounds with TPSA values in the 140-250 Å² range that also exhibit low permeability and are not substrates of active efflux transporters. This limitation prevents a definitive assessment of whether higher TPSA values are broadly compatible with permeability in the absence of structural features such as IMHB formation or active transport mechanisms.

In conclusion, this work demonstrates that no single descriptor can fully explain or predict passive permeability in complex chemical spaces. Instead, integrating experimental lipophilicity data (*shake-flask* or CHI), polarity measures (EPSA), and awareness of efflux transporter involvement provides a completer and more accurate picture. These insights can help refine screening strategies in drug discovery for bRo5 compounds, supporting the development of predictive models that better reflect the interplay of structure, physico-chemical properties, and biological transport mechanisms.

12 Caco-2 permeability prediction for Chiesi's compounds

As part of this PhD project, conducted in collaboration with Chiesi Farmaceutici S.p.A., chromatographic parameters have been increasingly integrated into the company's early drug discovery screening cascade, due to their numerous advantages. Over the past few years, approximately five thousand compounds from different research projects have been evaluated in terms of their chromatographic polarity and lipophilicity.

These data are crucial for expanding the internal chromatographic dataset and assessing the ability of these parameters to help identify the most promising candidates, particularly in relation to passive

permeability. However, the availability of permeability data is significantly more limited compared to the chromatographic measurements. In total, 544 compounds were available for the evaluation of permeability prediction.

12.1.1 Creation of an iterative rule for permeability prediction

To correctly classify compounds with good or acceptable permeability, and to ensure a more comprehensive dataset, a threshold of $Log_{10}P_{app} > 1.70$ (Caco2 AB = 50 nm/s) has been established as the criterion for moderate-high permeability. It should be observed that while Lipinski's rule demonstrates sensitivity, it lacks specificity: compounds exhibiting favorable Caco-2 permeability typically meet Lipinski's criteria (Lipinski = 1), yet numerous compounds classified as Lipinski = 1 display a wide range of Caco-2 permeability values. In order to assess the ability of a model (e.g. the Lipinski's rule or a combination of variables such as EPSA and CHI Log $D_{7.4}$) to correctly describe an observable property (in this case that $Log_{10}P_{app} > 1.70$), we employ the following statistical parameters, based on the cases in which the attribution of the observable property is correct, encompassing both true positives (TP: number of cases where the observed property is positive and correctly identified) and true negatives (TN: number of cases where the property is negative and correctly recognized).

Sensitivity (o TP ratio): *TP/(P = total number of positives)*
Specificity (o TN ratio): *TN/(N = total number of negatives)*
Accuracy: *(TP+TN)/(P+N)*

For Lipinski's rule, considering the set of 544 compounds for which $Log_{10}P_{app} > 1.70$, EPSA and CHI Log $D_{7.4}$ values are available, the following parameters are found:

Sensitivity: 86.2% (112 of the 130 compounds with $Log_{10}P_{app} > 1.70$ pass the Lipinski's rule
Specificity: 30% (124 of the 414 compounds with $Log_{10}P_{app} < 1.70$ don't pass the Lipinski's rule
Accuracy: 43.4% (only 236 of the 544 compounds are correctly predicted by Lipinski's rule)

This result indicates that while if a compound has a moderate or high (> 50nm/s) Caco-2 A>B value it is likely to pass Lipinski's rule, the fact that it meets the rule does not guarantee that the compound will have a value > 50nm/s, as only one-quarter of compounds with Caco2 below the threshold (negatives) fail to meet the rule, while three-quarters represent false positives (compounds classified as negatives that nonetheless satisfy the rule).

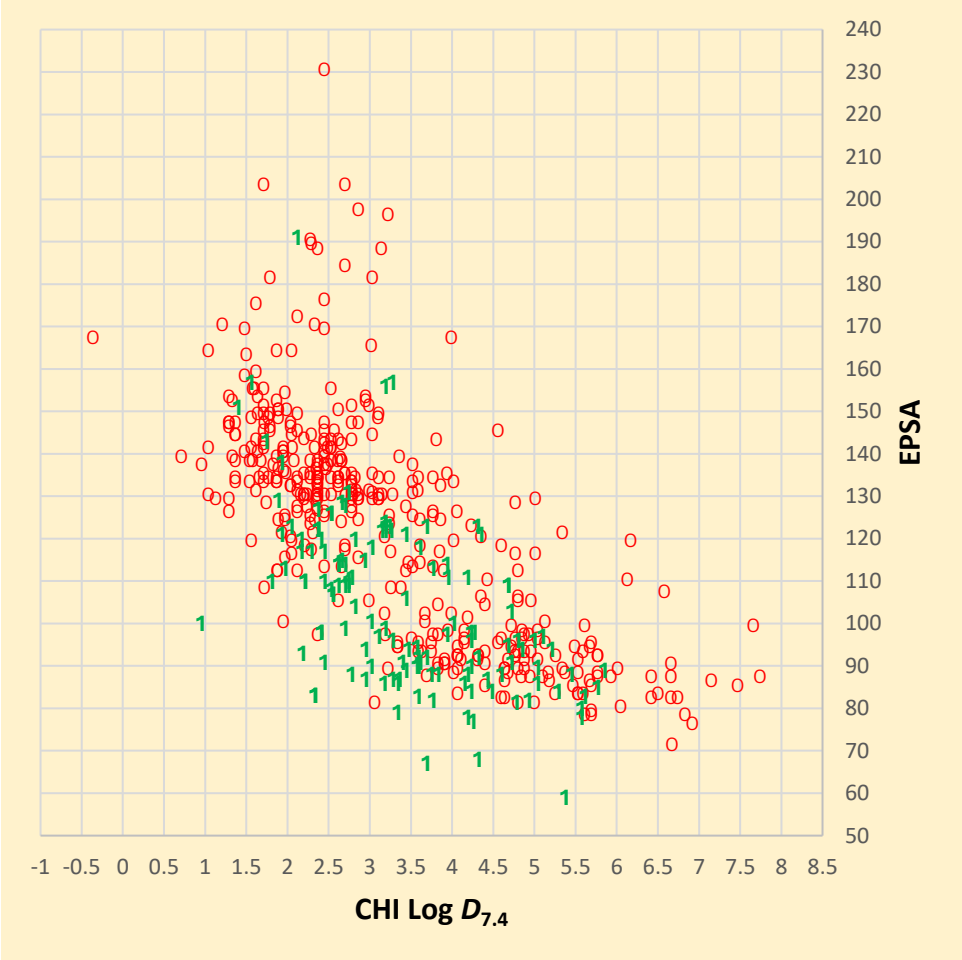


Figure 77. Arrangement in the EPSA and CHI Log D7.4 space of compounds with moderate or high permeability (1 = Caco2 AB > 50 nm/s) and with low permeability (0 = Caco AB < 50 nm/s)

To find a better model, which implements the use of EPSA and CHI Log $D_{7.4}$ chromatographic descriptors, possibly exploiting the calculated variables, the following steps were taken.

As can be seen from the two graphs (Figure 77 and the zoomed one Figure 78), CHI Log $D_{7.4}$ (x-axis) vs EPSA (y-axis), in which permeability classification is indicated as a binary value (1 = $Log_{10}P_{app}$ > 1.70; 0 = $Log_{10}P_{app}$ < 1.70), outside the rectangle defined by the limits (EPSA < 130; 2 < CHI Log $D_{7.4}$ < 6) there are very few positive compounds, with only 15 compounds with EPSA slightly above 130. These observations indicate that specific thresholds of CHI Log $D_{7.4}$ and EPSA are associated with higher permeability.

Within the space defined by EPSA < 130 and 2 < CHI Log $D_{7.4}$ < 6, both highly permeable (1) and less permeable compounds (0) are present, with no clear distinction between the two classes in terms of the spatial distribution of the chromatographic variables.

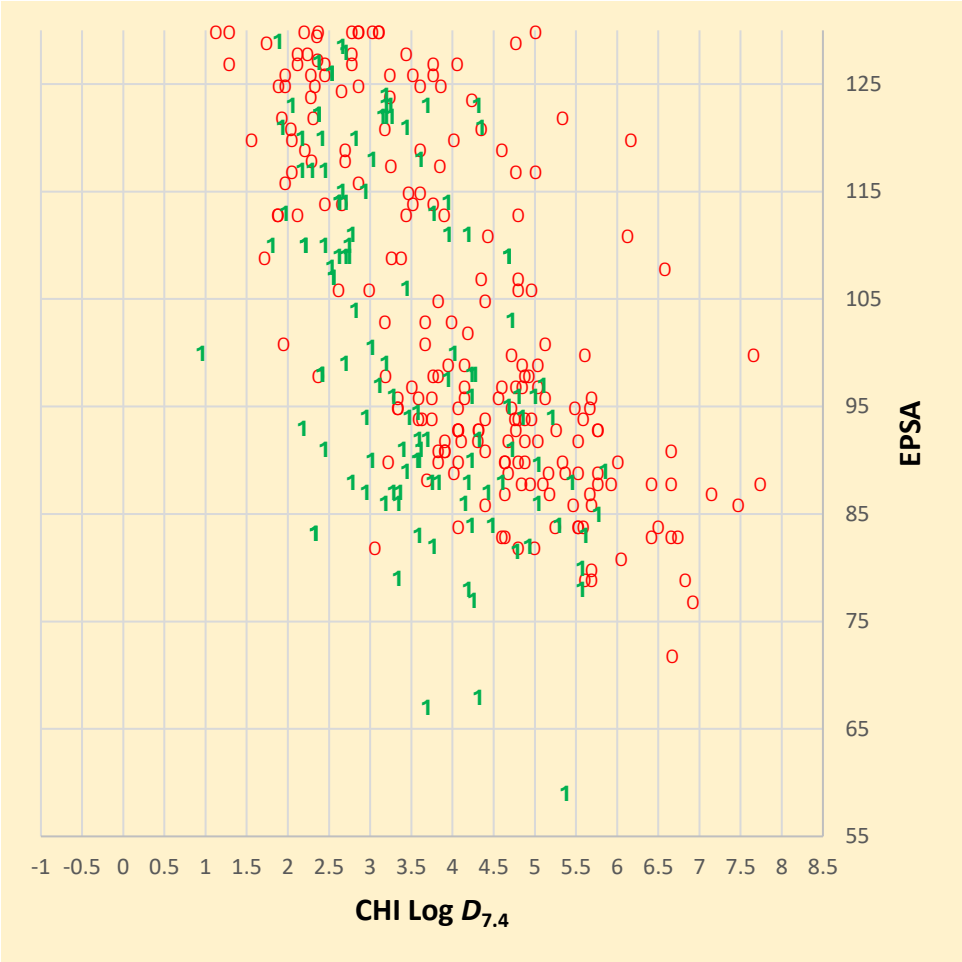


Figure 78. More spatially magnified distribution in the chromatographic parameters space

At this point, to get an estimate of the probability that exploits all available information, each compound within the "positive" area is assigned a *score* obtained from the MRA using y = $Log_{10}P_{app}$ and X = (CHI Log $D_{7.4}$; EPSA; Lipinski) where Lipinski's rule is an indicator variable set to 1 or 0 depending on whether the compound passes the rule (Figure 79).

At this point, by setting a threshold value of 1 for the score to selectively identify the majority of compounds with $Log_{10}P_{app}$ > 1.70, a new iterative rule can be defined to classify compounds likely to have Caco2 permeability value above 50 nm/s.

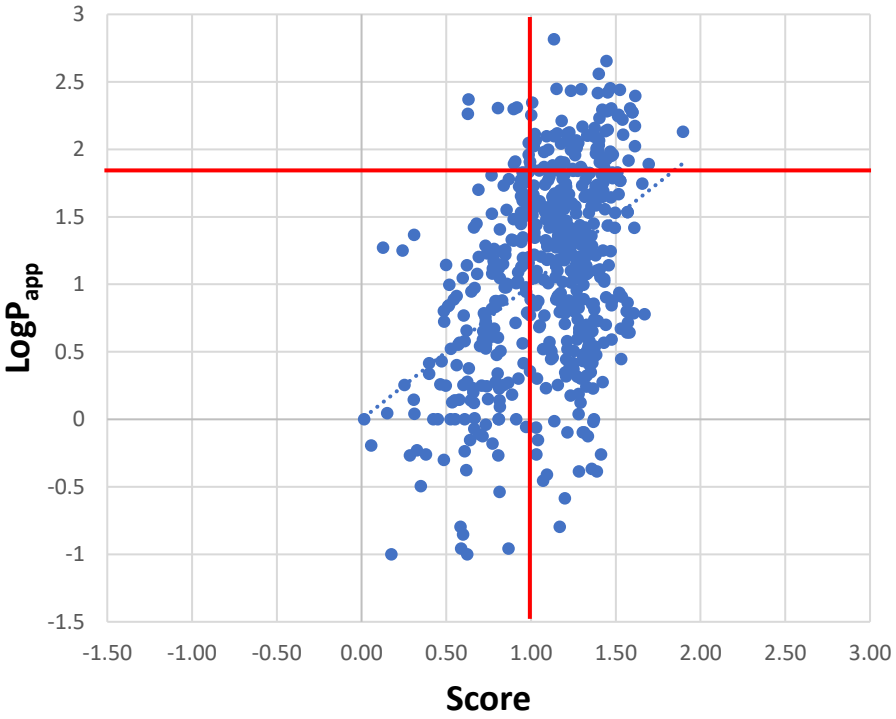


Figure 79. Relationship between the predictive Score (x-axis) and the apparent permeability coefficient (LogPapp) measured in Caco-2 cells (y-axis). The Score represents a descriptor derived from multivariate models integrating molecular properties related to passive diffusion. The red lines indicate the classification thresholds (Score > 1 and LogPapp > 1.70), above which compounds are identified as highly permeable.

This rule applies in 2 steps:

- If EPSA > 130 or CHI Log $D_{7.4}$ < 2 or CHI Log $D_{7.4}$ > 6, the class = 0
- For (EPSA < 130; 2 < CHI Log $D_{7.4}$ < 6) and Score > 1, the class = 1, otherwise class = 0

The parameters of such a level are:

Sensitivity: 79.2% (103 of the 130 compounds with $Log_{10}P_{app}$ > 1.70 pass the iterative rule)

Specificity: 74.6% (309 of the 414 compounds with $Log_{10}P_{app}$ < 1.70 do not pass the iterative rule)

Accuracy: 75.7% (412 of the 544 compounds are correctly predicted by the iterative rule)

In conclusion: compared to Lipinski's rule, considering the EPSA and CHI Log $D_{7.4}$ of the compounds it is possible to significantly decrease the number of false positives, thus increasing specificity at the expense of a limited decrease in sensitivity, with a substantial improvement in prediction accuracy.

12.1.2 Bidimensional heatmap and logistic regression using EPSA and CHI Log $D_{7.4}$ for permeability prediction

To better understand how physico-chemical properties influence the likelihood of class 1 membership, a preliminary analysis was performed by discretizing the chemical space into two-

dimensional bins defined by selected descriptors (e.g., CHI Log $D_{7.4}$ and EPSA). Four heatmaps were generated to visualize compound distribution and class probabilities.

The first two heatmaps (Figure 80) show the absolute number of compounds falling into each bin, when the first plot show the compounds with permeability > 50 nm/s while the second plot shows the distribution of all compounds in the space. These plots reveal the density and clustering of data points within specific regions of the descriptor space. It is evident that the dataset is not uniformly distributed, with a concentration of compounds in well-defined areas and sparse populations elsewhere.

The third and fourth heatmap plots (Figure 81) represents the empirical probability of class 1 within each bin, calculated as the ratio of class 1 compounds to the total number of compounds in that bin. This visualization highlights regions of high and low class 1 enrichment. Notably, in Figure 80 and Figure 81 (left heatmap) a gradual transition can be observed, with certain combinations of descriptors strongly associated with class 1, while others are predominantly populated by class 0 compounds. On the contrary in Figure 81 (right heatmap) is possible to see how Lipinski's rule is not able to discriminate compounds in the dataset, and in all descriptors combination there are compounds following the rule.

These empirical trends provide a rationale for the application of interpolation methods, such as logistic regression, to model the probability surface across the descriptor space. The logistic models presented in the following sections aim to capture and generalize these observed patterns.

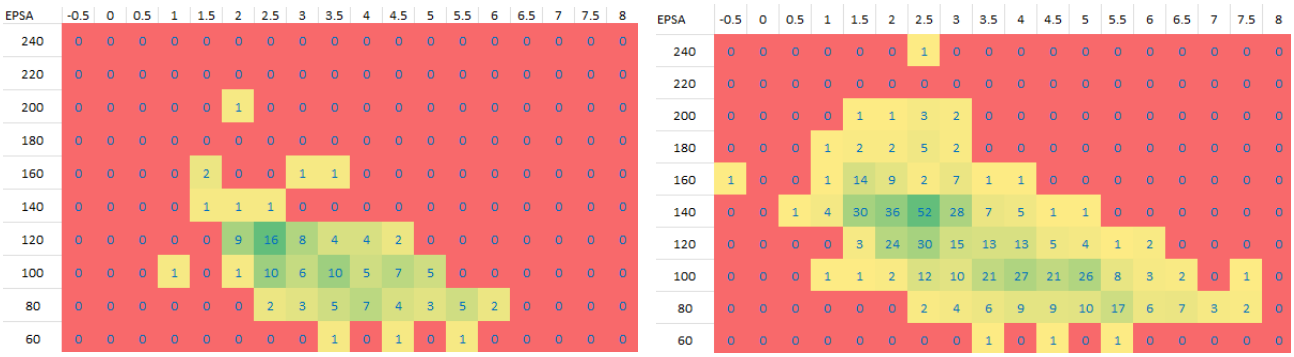


Figure 80. Heatmap plot showing the distribution in the EPSA and CHI Log $D_{7.4}$ space of compounds (and the number of those compounds) with Caco-2 > 50 nm/s (left plot), and the distribution (and the total number) in the EPSA and CHI Log $D_{7.4}$ space

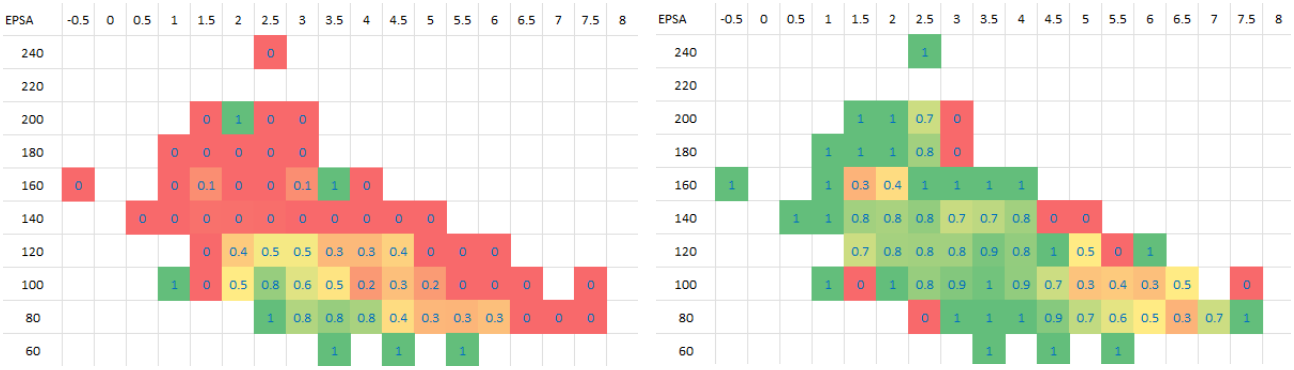


Figure 81. Heatmap plot showing the probability distribution in the EPSA and CHI Log $D_{7.4}$ space of with Caco-2 > 50 nm/s (left plot), and the probability distribution in the EPSA and CHI Log $D_{7.4}$ space of compounds that follow the Lipinski's rule

To investigate the influence of selected physico-chemical descriptors on the probability of a compound belonging to class 1 (moderate or high permeability, P_{app} > 50 nm/s) or class 0 (low permeability P_{app} < 50 nm/s), logistic regression was applied using EPSA and CHI Log $D_{7.4}$ as independent variables. The resulting probability curves (pcalc) are shown in the figures below (Figure 82 and Figure 83).

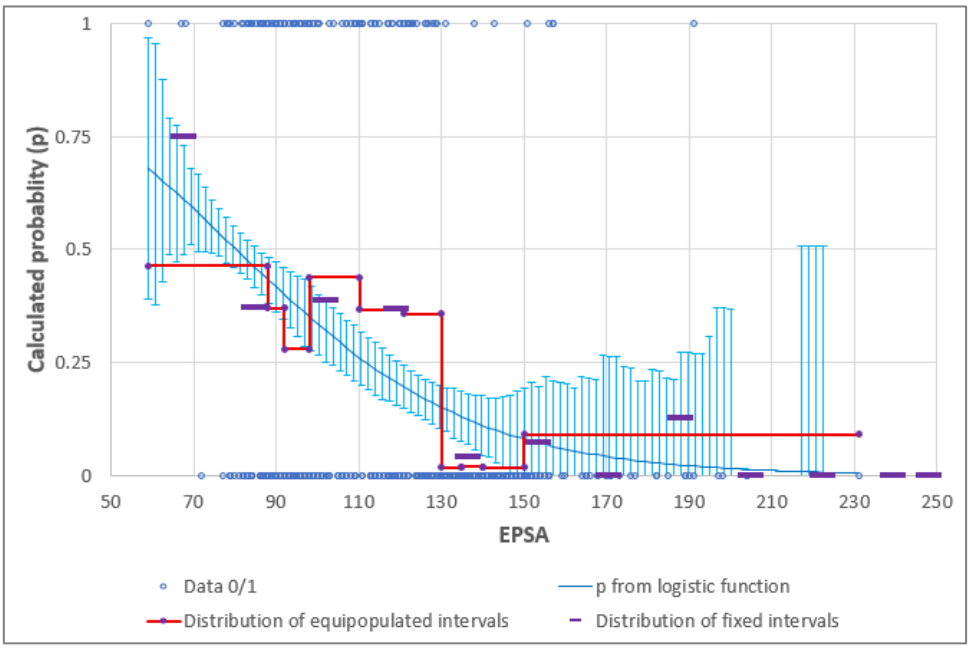


Figure 82. Probability curve as a result of logistic regression using EPSA data to predict moderate or high permeability class (P_{app} > 50) of the Chiesi's compounds

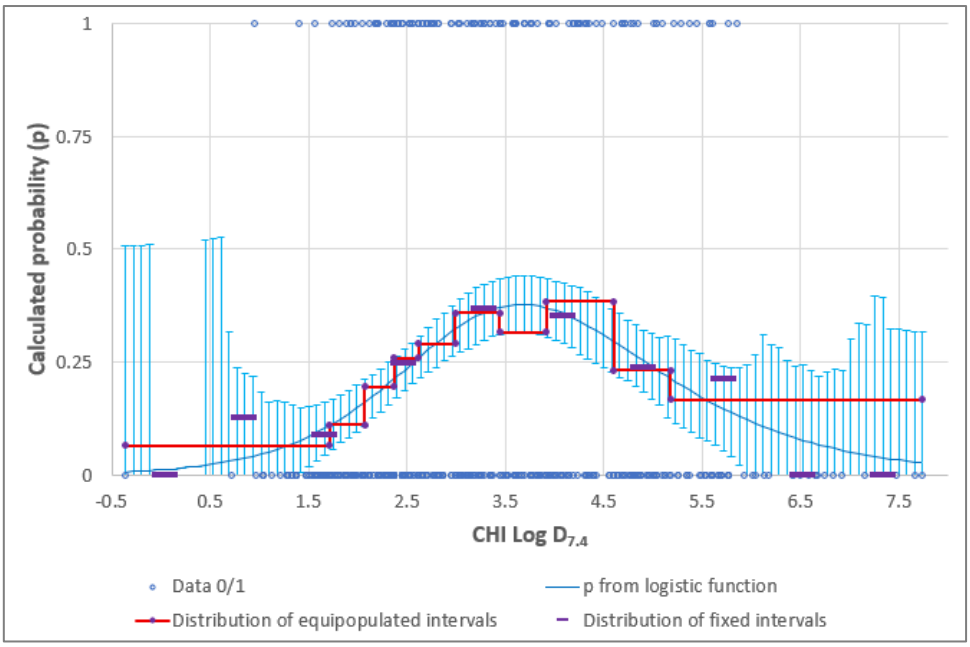


Figure 83. Probability curve as a result of logistic regression using CHI Log $D_{7.4}$ data to predict moderate or high permeability class (P_{app} > 50) of the Chiesi's compounds

In each plot:

- The *blue circles* represent the original binary class values (0 or 1) for each compound
- The *light blue line* shows the predicted probability obtained from logistic regression, with vertical error bars indicating the standard error based on the numerosity of compounds for each CHI Log $D_{7.4}$ values
- The *red line* corresponds to empirical probabilities calculated using *equal-frequency binning* (i.e., each bin contains the same number of compounds).
- The *purple line* shows empirical probabilities based on *fixed-width binning* (i.e., constant intervals on the x-axis)

The EPSA plot (Figure 82) shows an inverse relationship between the descriptor and the class 1 probability. High probabilities are associated with low EPSA values (below 100 Å²), while probabilities sharply drop for EPSA values above 130-140 Å². This behavior aligns with the hypothesis that excessive exposed polarity negatively impacts the property under study, such as passive permeability.

Both empirical estimates closely follow the logistic curve, providing additional validation for the model.

In the CHI Log $D_{7.4}$ plot (Figure 83), the predicted probability increases between 2.5 and 3.5, reaching a local maximum near 3.5 (bell-shape trend). This suggests that compounds with moderate

lipophilicity (as estimated by CHI Log $D_{7.4}$) are more likely to belong to class 1. Outside this optimal range, the probability decreases, indicating that both very low and very high lipophilicity values may be less favorable.
The empirical curves (red and purple) support the general shape of the logistic regression, showing good alignment in the central region of the distribution.

These results, based on the analyzed dataset, demonstrate that chromatographic parameters can effectively distinguish between permeable and non-permeable compounds within a certain range. In contrast, Lipinski's rule does not appear to provide the same level of discrimination based on permeability in this context. The chromatographic values evaluated are already being implemented in the company's screening cascade and could serve as the foundation for the development of predictive, iterative models. However, due to the chemical diversity of the compounds across different projects, it will be essential to tailor models and define specific threshold ranges for each chemical series or molecular class. This approach will improve the predictive power and increase the ability to distinguish permeable from non-permeable compounds for every project.
These promising findings will be further explored and validated in future work.

13 Experimental section - Chromatographic retention time prediction using physico-chemical properties

13.1 Aim

During my PhD, I spent six months at Uppsala University (Sweden), within the Department of Analytical Pharmaceutical Chemistry, headed by Professor Mikael Hedeland. During this research stay, I was involved in a project conducted in collaboration with separation scientists from AstraZeneca (Gothenburg R&D site), focusing on multivariate analysis and the development of a predictive model for chromatographic retention times. The study was based on a dataset of 39 compounds, acids, bases, two neutral and four zwitterions, aimed to correlate their calculated physico-chemical properties with their retention behavior. The ultimate goal of the project was to better understand how compounds behave under different chromatographic conditions and to enable the prediction of retention times without the need for extensive method development. By relying on molecular descriptors alone, the model could help anticipate chromatographic profiles of APIs and their related impurities. This predictive capability can significantly streamline method development, reduce experimental workload, and support robust separation strategies, which are particularly valuable in quality control and early-stage pharmaceutical development.

13.2 Materials

To measure retention time of the 39 pharmaceutical standard, the following chemical reagents were used: acetonitrile (Sigma-Merck), MilliQ water (Sigma-Merck), ammonium acetate (Sigma-Aldrich), formic acid (Sigma-Aldrich) and trifluoroacetic acid (Sigma-Aldrich).
The following instruments were used: Aquity UPLC (Waters), analytical balance (Mettler-toledo S.r.l. characterized by a sensibility of 0.1 mg and with a capacity of 100 mg), 1.5 mL Eppendorf vials (Eppendorf), weighting vessels (Sigma-Merck), pipettes P100, P1000 (Gilson), smart spatula (Sigma-Merck), ultrasound sonicator and vortex.

The following compounds were characterized for their chromatographic retention time using twelve different chromatographic methods: acetylsalycilic acid (Sigma-Aldrich), alprenolol (Sigma-Aldrich), amlodipine (provided by AstraZeneca), amoxicilline (Sigma-Aldrich), atenolol (Sigma-Aldrich), benzothiadiazine related compound A (provided by AstraZeneca), bumetanide (Sigma-Aldrich), candersartan (provided by AstraZeneca), candesartan related compound A (I) (provided by AstraZeneca), candesartan related compound B (II) (provided by AstraZeneca), candesartan related compound D (IV) (provided by AstraZeneca), candesartan related compound F (VI) (provided by

AstraZeneca), carbamazepine (Sigma-Aldrich), captopril (Sigma-Aldrich), cetirizine (Sigma-Aldrich), clopidogrel (provided by AstraZeneca), dabigatran (provided by AstraZeneca), dexamethasone (Sigma-Aldrich), diclofenac (Sigma-Aldrich), enalapril (provided by AstraZeneca), felodipine (provided by AstraZeneca), felodipine related compound A (H152/37) (provided by AstraZeneca), furosemide (provided by AstraZeneca), hydrochlorotiazide (provided by AstraZeneca), hydrochlorotiazide related compound A (provided by AstraZeneca), ibuprofen (Sigma-Aldrich), indomethacin (Sigma-Aldrich), losartan (provided by AstraZeneca), metoprolol (provided by AstraZeneca), omeoprazole (provided by AstraZeneca), omeprazole related compound A (H168/66) (provided by AstraZeneca), paracetamol (Sigma-Aldrich), phenylbutazone (Sigma-Aldrich), phenytoin (Sigma-Aldrich), pindolol (Sigma-Aldrich), propranolol (Sigma-Aldrich), salycilic acid (Sigma-Aldrich), timolol (Sigma-Aldrich).

13.3 Methods

Twelve chromatographic methods, using three different columns and four different aqueous mobile phases have been used. The column used were Acquity UPLC BEH C18 1.7um (2.1 x 100mm), Acquity UPLC HSS T3 1.8um (2.1 x 100mm) and InfinityLab poroshell 120 EC-C18 2.7um (3 x 100mm), while the mobile phase used were ammonium acetate solution 20mM, formic acid solution 0.1% and two trifluoroacetic acid solutions 0.1% and 0.03%. A gradient has been used as follows:

Table 14. Chromatographic gradient used to measure the retention time in reverse phase of compounds in the dataset

Time (min)	A - Aqueous solution (%)	B - Acetonitrile (%)
0.00	90	10
1	90	10
9	10	90
9.10	90	10
10	90	10

Injection volume was 3uL, column temperature maintained at 40°C and 0.5 mL/min as flow rate. The sample has been dissolved in a 50:50 solution of water/acetonitrile but on the basis of the solubility of the compounds the water percentage has been increased in case of high solubility to reach as much as possible the same condition as time zero, while acetonitrile has been increased in case of

compounds with low aqueous solubility. After solvation, a vortex mixer was used for mixing. Sonication and gentle heating were used to promote the dissolution of compounds with solubility issues. SIMCA®, a multivariate data analysis software from Sartorius, was used for data analysis.

13.4 Results

At the end of the HPLC-UV analysis, all the retention times (Rt) obtained for each compound under the 12 different chromatographic conditions (3 columns × 4 mobile phases), together with the corresponding physico-chemical properties, were collected and organized into a single dataset. This dataset was then processed using the SIMCA® software to perform multivariate statistical analyses. The objective was to explore and model the relationship between physico-chemical properties and chromatographic retention behavior, with the ultimate goal of identifying key molecular descriptors influencing retention time and building predictive models.

13.4.1 Principal component analysis (PCA)

A Principal Component Analysis (PCA) was first applied as an exploratory tool to visualize patterns within the dataset, detect potential outliers, and identify the variables contributing most to the variability in retention time across the different compounds (Figure 84).

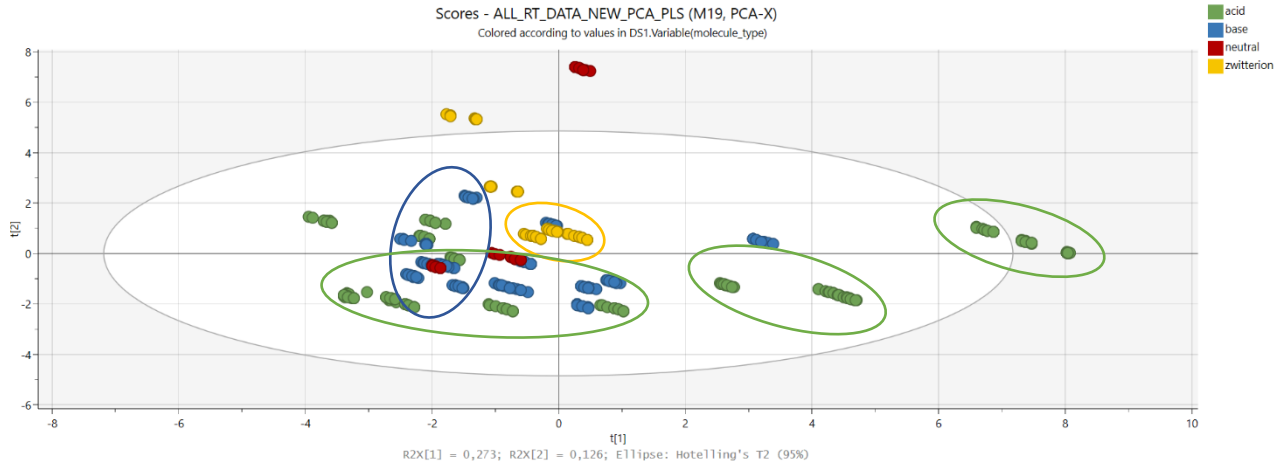


Figure 84. Score plot of PCA analysis of the dataset used, to highlight distribution of the data in the PCs axis

PCA transforms the original variables into new, uncorrelated variables called principal components (PCs), which are linear combinations of the original variables and are ranked according to the amount of variance they explain. In this case, the first principal component (PC1) explained 27% of the total variance, while the second component (PC2) accounted for 13%. Although these values are relatively low, they still provide meaningful insight into the structure of the data.

In the score plot, different clusters of compound classes were grouped using color-coded circles, yellow for loop diuretics, blue for β-antagonists and green for NSAIDs and sartan classes.
Two compounds, dexamethasone (red dots) and amoxicillin (yellow dots), did not cluster with any of the main chemical classes and appeared as outliers, indicating that their retention behavior is not well captured by the first two principal components.

To interpret the contribution of each variable (both chromatographic and physico-chemical) to the PCs, the loadings plot was used (Figure 85). Variables located far from the origin of the plot exert a stronger influence on the model and are therefore more important in explaining the variance.

Figure 85. Loadings plot used to verify the importance of each variable in the PCs plot

The dModX plot (Figure 86) further supported the identification of outliers, showing which compounds were poorly described by the model, above the red line. Although the overall model captured general trends in the dataset, certain compounds such as captopril (green), carbamazepine (red), and bumetanide (yellow) showed significant deviation from the model's prediction, suggesting that additional factors may influence their retention times.

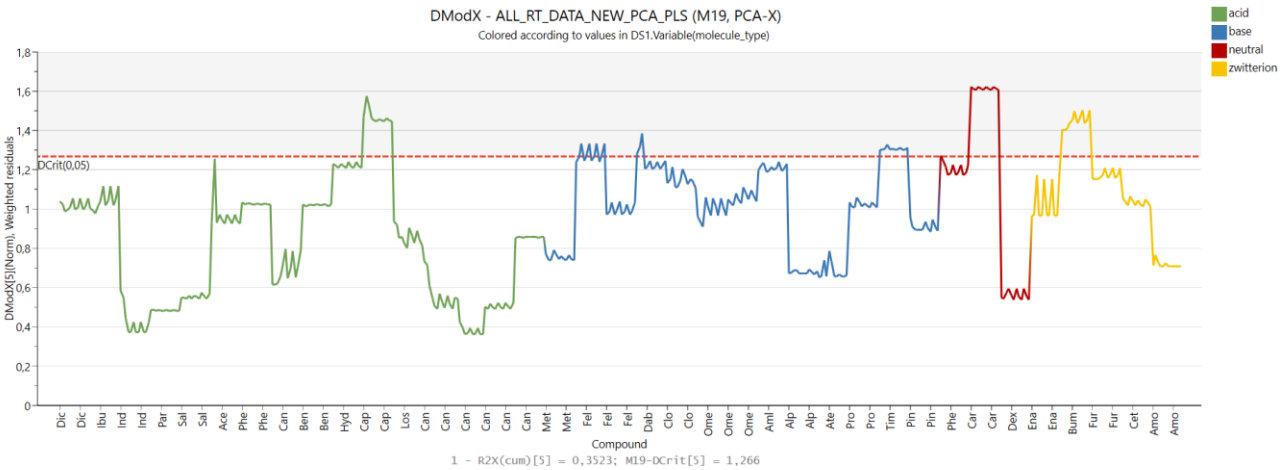


Figure 86. DModX plot used to verify the ability of each PC in the model to describe the data

After the PCA analysis was used to simplify the data creating new PCs, and to visualize pattern, distribution and the presence of outlier, based on these results, a Partial Least Square (PLS) analysis aimed to create a model for the retention time prediction has been used.

13.4.2 Partial Least Square analysis (PLS)

After this initial exploratory step, a Partial Least Squares (PLS) regression was applied to build predictive models for the retention time of the compounds, using their physico-chemical properties as input variables. Unlike PCA, which focuses on maximizing variance within X (independent variables), PLS aims to maximize the covariance between X and Y (in this case, between molecular descriptors and retention time), thereby identifying the most relevant descriptors for prediction.
This method allowed for the construction of regression models tailored to each chromatographic method, assessing how well the retention behavior of the compounds can be explained and predicted by their chemical properties. The PLS analysis also highlighted the relative importance of each descriptor in determining retention time, offering mechanistic insight into how molecular features such as polarity, hydrogen bonding, and lipophilicity interact with the chromatographic system.

During the construction of the PLS model, some physico-chemical descriptors with negligible contribution were excluded from the analysis in order to improve the model's performance and reduce noise. Also, the compounds without all the Rt using all the conditions have been excluded. These variables were identified based on their low Variable Importance in Projection (VIP) scores and minimal loading values, indicating limited influence on the prediction of retention time.

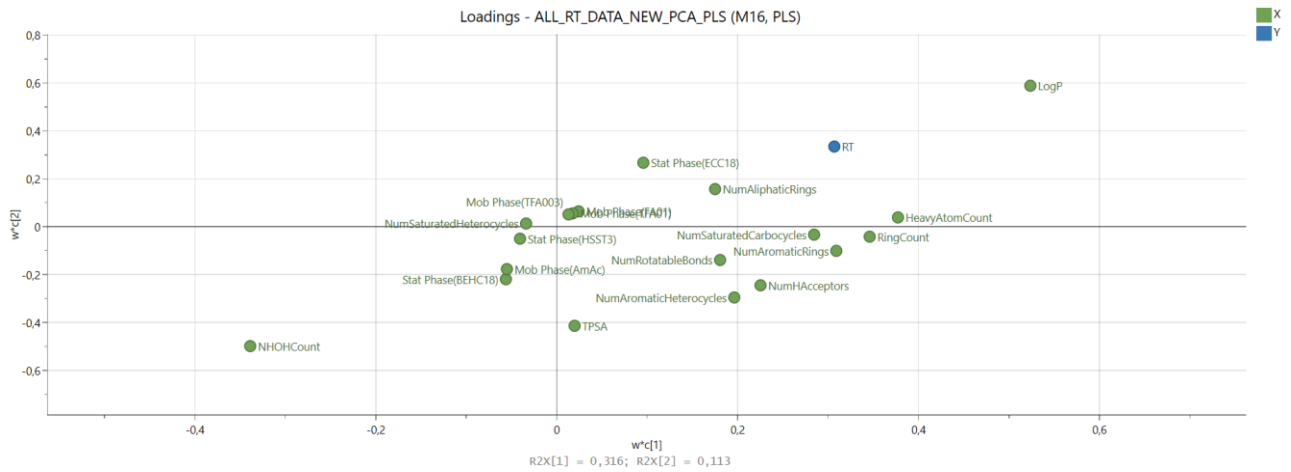


Figure 87. Loadings plot used to verify the importance of each variable to the model and the relative correlation to the Y variable (retention time)

The loadings plot (Figure 87) revealed that two physico-chemical properties played a major role in explaining the variation in retention time: Log *P*, which showed a direct correlation with retention time, compounds with higher lipophilicity tend to have longer retention times, especially in reversed-phase chromatographic systems. The number of hydrogen bond donors (i.e., NH and OH groups), which was inversely correlated with retention time, likely due to increased polarity and interaction with the mobile phase, leading to faster elution. These findings align with expected chromatographic behavior, where hydrophobic interactions drive longer retention on non-polar stationary phases, while polar functional groups enhance solubility in the mobile phase, reducing retention.

In the dModX plot (Figure 88), the majority of the compounds fall below the red confidence limit, indicating that the model describes them well and that the prediction is reliable. However, two compounds, carbamazepine and dexamethasone, lie above the threshold, suggesting that they are not well explained by the model. This may be due to the absence of specific physico-chemical variables that adequately capture their chromatographic behavior.

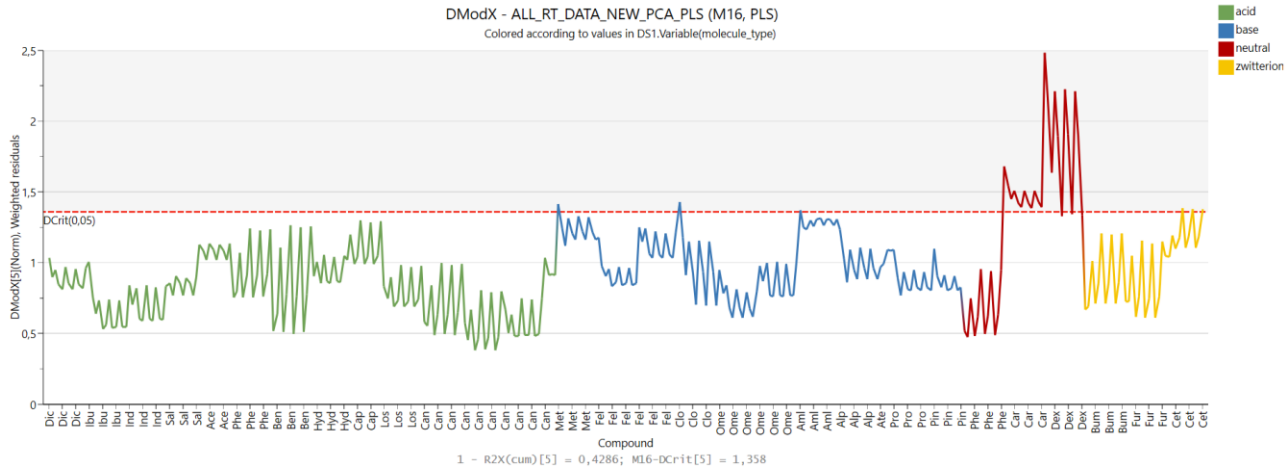


Figure 88. DModX plot used to verify the ability of the model to describe the data and the presence of outliers

To further investigate the contribution of each physico-chemical descriptor to the prediction of retention time, two key plots were considered: the Variable Importance in the Projection (VIP) plot and the regression coefficient plot. Both plots provide complementary information on the role of each variable in the model, but from different perspectives.

The VIP plot (Figure 89) reflects the overall importance of each variable across all components in the PLS model. It integrates both the explained variance of the response variable and the weight of each component, offering a global view of variable relevance. Typically, a VIP score greater than 1 indicates that the variable is significantly contributing to the model's predictive performance.

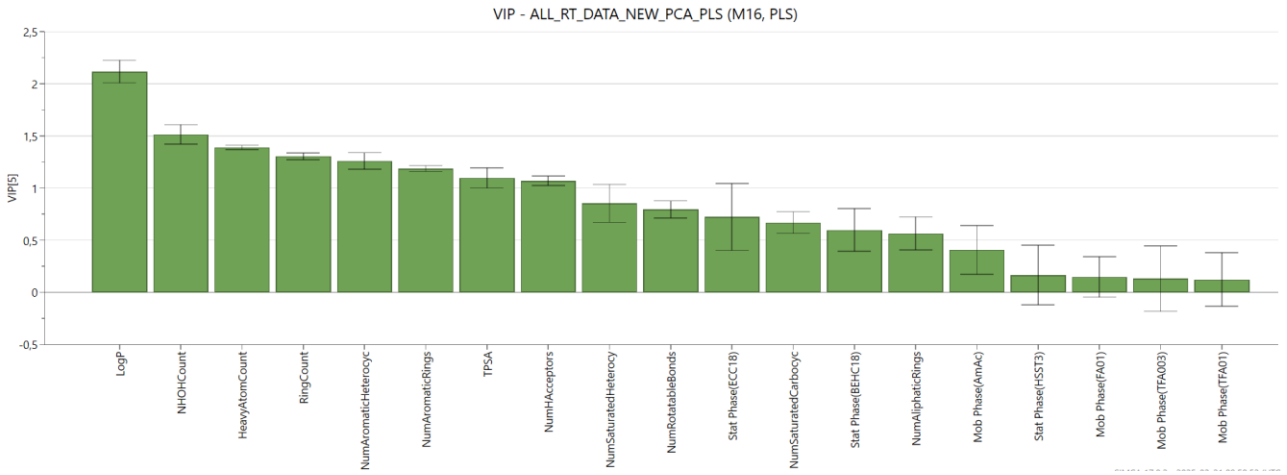


Figure 89. VIP plot for the evaluation of the important of each variable in the model

On the other hand, the regression coefficient plot (Figure 90) illustrates the direct influence of each variable on the response (i.e., retention time). It shows how each descriptor contributes to the prediction equation, both in magnitude and in direction. A positive coefficient indicates a direct correlation (an increase in the variable leads to an increase in Rt), while a negative coefficient indicates an inverse relationship. Taken together, the VIP and regression coefficient plots help identify the most influential variables in the model, both in terms of global importance and specific predictive effect on chromatographic retention.

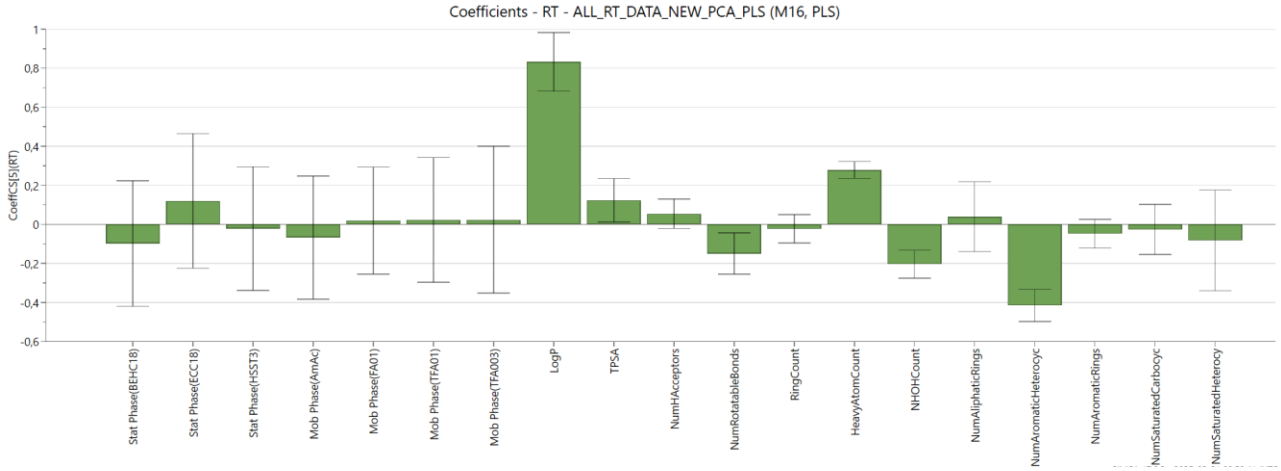


Figure 90. Regression coefficient plot used together with the VIP plot to verify the importance of each variable in the model

13.5 Conclusion

At the end of the PCA analysis and PLS model development, two key performance metrics were evaluated: the coefficient of determination (R^2) and the predictive ability (Q^2). The first one represents the proportion of variance in the observed data that is explained by the model, providing a measure of the goodness-of-fit. The second one, calculated through cross-validation, estimates the model's predictive capability, indicating how well the model can generalize to new, unseen data.

In our case, the PLS model built using the selected dataset yielded an R^2 value of 0.86, indicating a strong correlation between observed and predicted retention times (Figure 91).

The Q^2 value was 0.80, which meets the predefined goal of the project and suggests a good starting point for further model development. These results were obtained using twelve physico-chemical descriptors out of the initial nineteen, after excluding seven variables with negligible contribution to the model. Additionally, 345 out of 444 compounds were retained as observations. Compounds lacking retention time values, either due to very early elution or poor peak shape, were excluded.

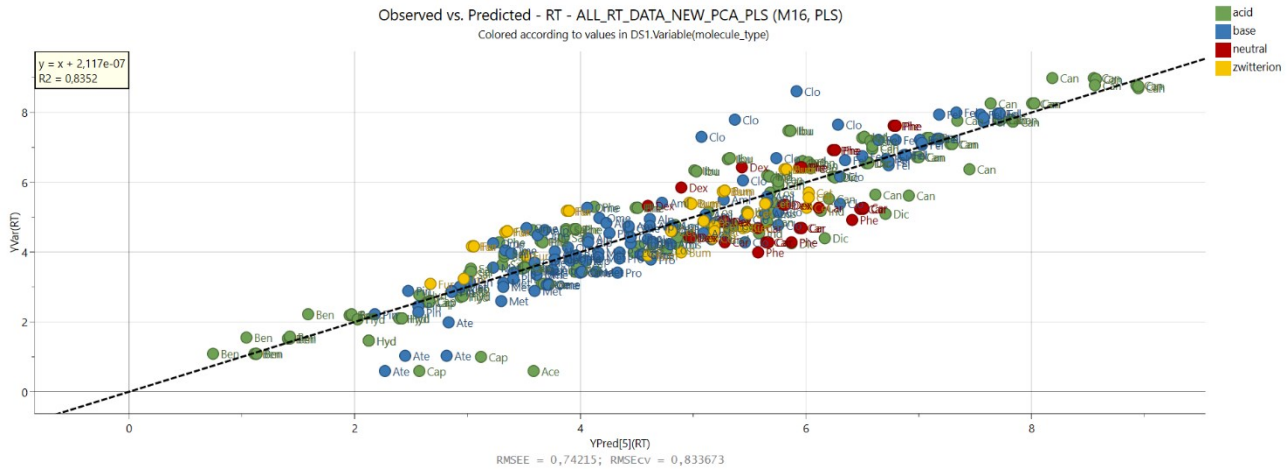


Figure 91. Correlation between observed and predicted Rt time of the compounds in the dataset

Table 15 compares the model performance using all twelve selected descriptors with the performance of a reduced model based on the seven most relevant variables, identified through VIP scores and regression coefficients. The difference in model performance was minimal. Interestingly, when constructing a model using only the calculated Log *P* value, the R^2 and Q^2 values were only slightly lower, suggesting that lipophilicity is the most influential parameter in determining chromatographic retention time under the studied conditions. The remaining physico-chemical properties appear to contribute marginally, or in some cases may act as noise, reducing the overall predictive accuracy of the model.

Another relevant finding concerns the comparison and correlation of retention times obtained from four different chromatographic conditions, three employing acidic additives in the mobile phase and one using ammonium acetate, as well as three different stationary phases. The results show a very strong correlation ($R^2 = 0.99$) among the retention times from the acidic conditions across all columns, and between those obtained under ammonium acetate conditions. This suggests that the chromatographic behavior of the compounds is consistent across these settings, and that the model could be significantly simplified by retaining only two representative conditions: one stationary phase combined with two mobile phases (ammonium acetate and an acid).

Table 15. R^2 and Q^2 values using twelve different variables, seven different variables and using Log P only

	12 variables	7 variables	1 variable (Log *P*)
R^2	0.84	0.83	0.75
Q^2	0.80	0.80	0.73

Figure 92 shows a clear separation between acidic and basic compounds when comparing the average retention times measured under ammonium acetate (20 mM) across three different columns to those measured using 0.1% formic acid under the same stationary phases. The separation between acidic and basic compounds is striking. Acidic compounds elute more rapidly in acidic condition, while basic compounds elute faster under ammonium acetate conditions. This behavior is likely influenced by the compounds' pK_a values. Acidic compounds tend to be deprotonated under the buffered (ammonium acetate) conditions, which reduces their interaction with the hydrophobic C18 stationary phase. Under acidic conditions, these compounds are predominantly in their neutral form, enhancing their retention due to stronger hydrophobic interactions. Conversely, basic compounds are mostly protonated under acidic conditions, leading to decreased retention.

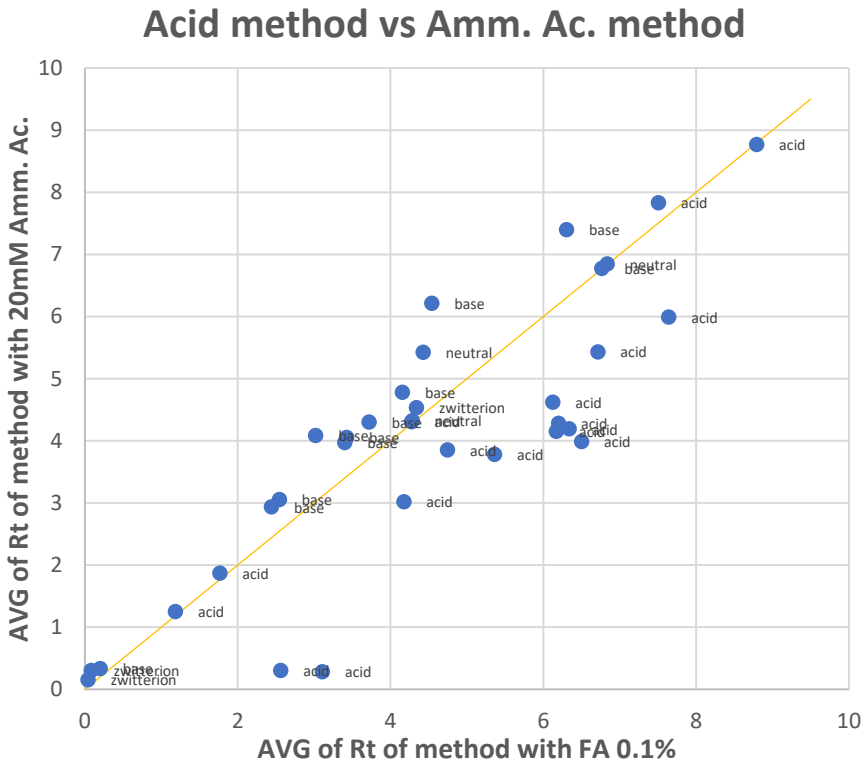


Figure 92. Correlation between average retention time using the method with formic acid 0.1% and the average retention time using the method with ammonium acetate 20mM

In future model developments, these insights should be taken into account. Moreover, the inclusion of experimentally measured physico-chemical properties, preferably determined using chromatographic techniques, could further enhance the model's predictive performance. Expanding the dataset size would also contribute to improving the robustness and generalizability of the models; however, this may come at the cost of reduced predictivity. Therefore, the development of specific

models tailored to different classes of compounds could represent a more effective strategy to accurately predict retention behavior based on compound-specific features.

14 General conclusions

This PhD project explored both experimental and computational determination of key physico-chemical properties -pK_a, lipophilicity, and polarity- and their impact on the permeability, screening strategies, and predictive modeling in early-stage drug discovery. Although these properties are conceptually distinct, they are closely interconnected and collectively shape a molecule's ADME profile.

The accurate experimental determination of pK_a using the SiriusT3™ instrument proved to be a cornerstone in the physico-chemical characterization workflow. It enabled the reliable assessment of compounds with diverse ionization patterns, including peptides and macrocycles, and provided essential input for subsequent Log *P* and Log *D* determinations. The value of integrating experimental pK_a measurements lies in the increased internal consistency of lipophilicity assessments and in enhanced confidence in permeability and solubility predictions, two key parameters for formulation and candidate selection.

In the analysis of lipophilicity, the study demonstrated the superior robustness and biological relevance of chromatographic CHI Log $D_{7.4}$ compared to traditional pH-metric and shake-flask methods, particularly when dealing with bRo5 compounds. CHI Log $D_{7.4}$ provided reliable results across structurally diverse molecules and exhibited strong correlation with shake-flask data. However, significant discrepancies emerged when comparing experimental data with computational estimates (CLog *P*, XLog *P*, CLog *D*), highlighting the limitations of *in-silico* methods for complex, flexible, amphiphilic or multi-ionizable molecules. These findings reinforce the need for experimental validation, especially when modeling the behavior of advanced chemical scaffolds.

The evaluation of polarity through EPSA measurements offered further insight into membrane permeability potential, particularly when integrated with calculated TPSA values. Although TPSA can be useful as a quick screening tool, EPSA provides a more sensitive and experimentally grounded descriptor, capable of capturing subtle conformational effects and surface interactions. The high reproducibility and throughput of the EPSA method (Wang et al. protocol) enabled its implementation in industrial screening workflows and demonstrated its added value in the identification of intramolecular hydrogen bonds (IMHBs), a feature often invisible to computational models. Together with the $\Delta\text{Log}\,P_{\text{oct-tol}}$ descriptor, EPSA showed good capacity to detect IMHBs and to predict permeability trends, especially in the intermediate polarity space.

From a modeling perspective, the multivariate analysis (PCA and PLS) highlighted the central role of lipophilicity as the primary driver of chromatographic retention. The development of predictive PLS models using selected physico-chemical parameters yielded strong performance ($R^2 = 0.86$, $Q^2 = 0.80$), confirming the relevance of lipophilicity in retention time classification. Interestingly, retention behavior across different chromatographic conditions showed consistent trends, allowing the identification of simplified yet highly informative settings. In particular, the retention shifts observed between acidic and buffered mobile phases could be rationalized by considering the compounds' ionization states, reinforcing the importance of pK_a as a mechanistic parameter in chromatographic and biological behavior.

In conclusion, this research demonstrates the value of integrating chromatographic physico-chemical profiling into the early stages of drug development, also for chemically complex compounds that lie beyond the traditional Ro5 space. By combining high-quality experimental data with modeling approaches, the project laid the groundwork for the development of more accurate, context-specific predictive models for permeability prediction and compound prioritization. The insights gained from this work not only contributed to the optimization of internal screening strategies but also highlighted the limitations of relying exclusively on computational tools, advocating for a balanced and data-informed approach to early drug discovery.

15 Bibliography

[1] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, “Innovation in the pharmaceutical industry: New estimates of R&D costs,” *J Health Econ*, vol. 47, pp. 20–33, May 2016, doi: 10.1016/j.jhealeco.2016.01.012.

[2] O. J. Wouters, M. McKee, and J. Luyten, “Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018,” *J Am Med Assoc*, vol. 323, no. 9, pp. 844–853, Mar. 2020, doi: 10.1001/jama.2020.1166.

[3] C. H. Wong, K. W. Siah, and A. W. Lo, “Estimation of clinical trial success rates and related parameters,” *Biostatistics*, vol. 20, no. 2, pp. 273–286, Apr. 2019, doi: 10.1093/biostatistics/kxx069.

[4] M. Hay, D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal, “Clinical development success rates for investigational drugs,” *Nat Biotechnol*, vol. 32, no. 1, pp. 40–51, Jan. 2014, doi: 10.1038/nbt.2786.

[5] M. M. Hann, “Molecular obesity, potency and other addictions in drug discovery,” *Medchemcomm*, vol. 2, no. 5, pp. 349–355, May 2011, doi: 10.1039/c1md00017a.

[6] A. Pagliara, P.-A. Carrupt, G. Caron, P. Gaillard, and B. Testa, “Lipophilicity Profiles of Ampholytes,” *Chem Rev*, vol. 97, no. 8, Dec. 1997, doi: 10.1021/cr9601019.

[7] A. Pagliara *et al.*, “Molecular properties and pharmacokinetic behavior of cetirizine, a zwitterionic H1-receptor antagonist,” *J Med Chem*, vol. 41, no. 6, pp. 853–863, 1998, doi: 10.1021/jm9704311.

[8] P. D. Leeson and B. Springthorpe, “The influence of drug-like concepts on decision-making in medicinal chemistry,” *Nat Rev Drug Discov*, vol. 6, no. 11, pp. 881–890, Nov. 2007, doi: 10.1038/nrd2445.

[9] C. A. Lipinski, B. W. Dominy, and P. J. Feeney, “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings,” *Adv Drug Deliv Rev*, vol. 23, no. 1–3, pp. 3–25, Jan. 1997, doi: 10.1016/s0169-409x(00)00129-0.

[10] M. Rossi Sebastiano *et al.*, “Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs beyond the Rule of 5,” *J Med Chem*, vol. 61, no. 9, pp. 4189–4202, May 2018, doi: 10.1021/acs.jmedchem.8b00347.

[11] M. Rossi Sebastiano, D. Garcia Jimenez, M. Vallaro, G. Caron, and G. Ermondi, “Refinement of Computational Access to Molecular Physicochemical Properties: From Ro5 to bRo5,” *J Med Chem*, vol. 65, no. 18, pp. 12068–12083, Sep. 2022, doi: 10.1021/acs.jmedchem.2c00774.

[12] T. W. Johnson, K. R. Dress, and M. Edwards, “Using the Golden Triangle to optimize clearance and oral absorption,” *Bioorg Med Chem Lett*, vol. 19, no. 19, pp. 5560–5564, Oct. 2009, doi: 10.1016/j.bmcl.2009.08.045.

[13] S. Mignani, S. Huber, H. Tomás, J. Rodrigues, and J. P. Majoral, “Compound high-quality criteria: a new vision to guide the development of drugs, current situation,” Apr. 01, 2016, *Elsevier Ltd*. doi: 10.1016/j.drudis.2016.01.005.

[14] Y. T. Wang *et al.*, “High-Throughput SFC-MS/MS Method to Measure EPSA and Predict Human Permeability,” *J Med Chem*, vol. 67, no. 16, pp. 13765–13777, Aug. 2024, doi: 10.1021/acs.jmedchem.4c00571.

[15] S. Neervannan, “Preclinical formulations for discovery and toxicology: physicochemical challenges,” *Expert Opin Drug Metab Toxicol*, vol. 2, no. 5, pp. 715–731, 2006, doi: 10.1517/17425255.2.5.715.

[16] G. H. Goetz and M. Shalaeva, "Leveraging chromatography based physicochemical properties for efficient drug design," *ADMET DMPK*, vol. 6, no. 2, pp. 85–104, Jun. 2018, doi: 10.5599/admet.529.

[17] G. H. Goetz, M. Shalaeva, G. Caron, G. Ermondi, and L. Philippe, "Relationship between passive permeability and molecular polarity using block relevance analysis," *Mol Pharm*, vol. 14, no. 2, pp. 386–393, Feb. 2017, doi: 10.1021/acs.molpharmaceut.6b00724.

[18] G. H. Goetz, L. Philippe, and M. J. Shapiro, "EPSA: A novel supercritical fluid chromatography technique enabling the design of permeable cyclic peptides," *ACS Med Chem Lett*, vol. 5, no. 10, pp. 1167–1172, Oct. 2014, doi: 10.1021/ml500239m.

[19] G. H. Goetz, L. Philippe, and M. J. Shapiro, "EPSA: A novel supercritical fluid chromatography technique enabling the design of permeable cyclic peptides," *ACS Med Chem Lett*, vol. 5, no. 10, pp. 1167–1172, Oct. 2014, doi: 10.1021/ml500239m.

[20] G. Caron, M. Vallaro, and G. Ermondi, "High throughput methods to measure the propensity of compounds to form intramolecular hydrogen bonding," *Medchemcomm*, vol. 8, no. 6, pp. 1143–1151, 2017, doi: 10.1039/c7md00101k.

[21] C. Dardonville, "Automated techniques in pKa determination: Low, medium and high-throughput screening methods," *Drug Discov Today Technol*, vol. 27, pp. 49–58, Jul. 2018, doi: 10.1016/j.ddtec.2018.04.001.

[22] D. T. Manallack, "The pKa Distribution of Drugs: Application to Drug Discovery," *Perspect Medicin Chem*, vol. 1, pp. 25–38, Sep. 2007.

[23] E. Sjögren *et al.*, "In vivo methods for drug absorption - Comparative physiologies, model selection, correlations with in vitro methods (IVIVC), and applications for formulation/API/excipient characterization including food effects," *European Journal of Pharmaceutical Sciences*, vol. 57, no. 1, pp. 99–151, Jun. 2014, doi: 10.1016/j.ejps.2014.02.010.

[24] G. A. Mtewa *et al.*, "Fundamental Methods in Drug Permeability, pKa, LogP and LogDx Determination," *Journal of Drug Research and Development*, vol. 5, no. 1, pp. 1–6, Nov. 2018, doi: dx.doi.org/10.16966/2470-1009.146.

[25] G. Caron, G. Ermondi, and R. Scherrer, "Lipophilicity, Polarity, and Hydrophobicity," *Comprehensive Medicinal Chemistry II*, vol. 5, pp. 425–452, Nov. 2006, doi: 10.1016/B0-08-045044-X/00135-8.

[26] G. E. Han and R. Priefer, "A systematic review of various pKa determination techniques," *Int J Pharm*, vol. 635, Mar. 2023, doi: 10.1016/j.ijpharm.2023.122783.

[27] J. Reijenga, A. van Hoof, A. van Loon, and B. Teunissen, "Development of methods for the determination of pKa values," *Anal Chem Insights*, vol. 8, no. 1, pp. 53–71, Aug. 2013, doi: 10.4137/ACI.S12304.

[28] M. Yasuda, "Dissociation Constants of Some Carboxylic Acids in Mixed Aqueous Solvents," *Bull Chem Soc Jpn*, vol. 32, no. 5, pp. 429–432, May 1959, doi: doi.org/10.1246/bcsj.32.429.

[29] A. Avdeef *et al.*, "PH-metric log P 11. pK a determination of water-insoluble drugs in organic solvent-water mixtures," *J Pharm Biomed Anal*, vol. 20, pp. 631–641, 1999, doi: doi.org/10.1016/S0731-7085(98)00235-0.

[30] M. AG *et al.*, "Fundamental Methods in Drug Permeability, pKa, LogP and LogDx Determination," *Journal of Drug Research and Development*, vol. 5, no. 1, 2019, doi: 10.16966/2470-1009.146.

[31] J. C. Dearden and G. M. Bresnen, "The Measurement of Partition Coefficients," *Quant. Struct.-Act. Relat.*, vol. 7, pp. 133–144, 1988, doi: doi.org/10.1002/qsar.19880070304.

[32] X. Liu, B. Testa, and A. Fahr, "Lipophilicity and its relationship with passive drug permeation," *Pharm Res*, vol. 28, no. 5, pp. 962–977, May 2011, doi: 10.1007/s11095-010-0303-7.

[33] K. Valko, C. M. Du, C. Bevan, D. P. Reynolds, and M. H. Abraham, "Rapid Method for the Estimation of Octanol/Water Partition Coefficient (Log P oct) from Gradient RP-HPLC Retention and a Hydrogen Bond Acidity Term (Σα 2 H)," 2001.

[34] K. L. Valko, "Application of biomimetic HPLC to estimate in vivo behavior of early drug discovery compounds," *Future Drug Discov*, vol. 1, no. 1, pp. 1–14, Jul. 2019, doi: 10.4155/fdd-2019-0004.

[35] G. H. Goetz and M. Shalaeva, "Leveraging chromatography based physicochemical properties for efficient drug design," 2018, *International Association of Physical Chemists*. doi: 10.5599/admet.529.

[36] J. Comer and K. Tam, "Lipophilicity Profiles: Theory and Measurement," in *Pharmacokinetic Optimization*, B. Testa, H. van de Waterbeemd, G. Folkers, and R. Guy, Eds., 2001, pp. 275–304. doi: 10.1002/9783906390437.ch17.

[37] R. J. Young, D. V. S. Gree, C. N. Luscombe, and A. P. Hill, "Getting physical in drug discovery II: The impact of chromatographic hydrophobicity measurements and aromaticity," *Drug Discov Today*, vol. 16, no. 17–18, pp. 822–830, Sep. 2011, doi: 10.1016/j.drudis.2011.06.001.

[38] K. Valkó, C. Bevan, and D. Reynolds, "Chromatographic Hydrophobicity Index by Fast-Gradient RP-HPLC: A High-Throughput Alternative to log P/log D," *Anal Chem*, vol. 69, no. 11, pp. 2022–2029, Jun. 1997, doi: doi:10.1021/ac961242d.

[39] K. L. Valkó, "Lipophilicity and biomimetic properties measured by HPLC to support drug discovery," *J Pharm Biomed Anal*, vol. 130, pp. 35–54, Oct. 2016, doi: 10.1016/j.jpba.2016.04.009.

[40] M. H. Abraham, W. E. Acree, A. J. Leo, D. Hoekman, and J. E. Cavanaugh, "Water-solvent partition coefficients and Delta Log P values as predictors for blood-brain distribution; application of the Akaike information criterion," *J Pharm Sci*, vol. 99, no. 5, pp. 2492–2501, May 2010, doi: 10.1002/jps.22010.

[41] C. P. Tinworth and R. J. Young, "Facts, Patterns, and Principles in Drug Discovery: Appraising the Rule of 5 with Measured Physicochemical Data," Sep. 24, 2020, *American Chemical Society*. doi: 10.1021/acs.jmedchem.9b01596.

[42] S. Bunally and R. J. Young, "The role and impact of high throughput biomimetic measurements in drug discovery," 2018, *International Association of Physical Chemists*. doi: 10.5599/admet.530.

[43] S. B. Bunally, C. N. Luscombe, and R. J. Young, "Using Physicochemical Measurements to Influence Better Compound Design," Sep. 01, 2019, *SAGE Publications Inc.* doi: 10.1177/2472555219859845.

[44] D. A. Degoey, H. J. Chen, P. B. Cox, and M. D. Wendt, "Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection," Apr. 12, 2018, *American Chemical Society*. doi: 10.1021/acs.jmedchem.7b00717.

[45] A. Guillot *et al.*, "Lipophilicity Determination of Highly Lipophilic Compounds by Liquid Chromatography," *Chem Biodivers*, vol. 6, no. 11, pp. 1828–1836, 2009, doi: 10.1002/cbdv.200900115.

[46] R. P. Austin, A. M. Davis, and C. N. Manners, "Partitioning of ionizing molecules between aqueous buffers and phospholipid vesicles," *J Pharm Sci*, vol. 84, no. 10, pp. 1180–1183, 1995, doi: 10.1002/jps.2600841008.

[47] K. Palm, K. Luthman, A.-L. Ungell, and G. Strandlund, “Correlation of drug absorption with molecular surface properties,” *J Pharm Sci*, vol. 85, no. 1, pp. 32–39, Jan. 1996, doi: 10.1021/js950285r.

[48] D. E. Clark, “What has polar surface area ever done for drug discovery?,” *Future Med Chem*, vol. 3, no. 4, pp. 469–484, Mar. 2011, doi: 10.4155/fmc.11.1.

[49] G. H. Goetz *et al.*, “High throughput method for the indirect detection of intramolecular hydrogen bonding,” *J Med Chem*, vol. 57, no. 7, pp. 2920–2929, Apr. 2014, doi: 10.1021/jm401859b.

[50] H. van de Waterbeernd er al, H. van de Waterbeemd, G. F. Camenisch Hoffmann-, G. Folkers, and O. A. Raevsky, “Estimation of Caco-2 Cell Permeability using Calculated Molecular Descriptors,” *Quant. Struct.-Act. Relat*, vol. 15, pp. 480–490, 1996, doi: doi.org/10.1002/qsar.19960150604.

[51] K. Palm, P. Stenberg, K. Luthman, and P. Artursson, “Polar molecular surface properties predict the intestinal absorption of drugs in humans,” *Pharm Res*, vol. 14, no. 5, pp. 568–571, May 1997, doi: 10.1023/a:1012188625088.

[52] H. van de Waterbeemd and M. Kansy, “Hydrogen-Bonding Capacity and Brain Penetration,” *CHIMIA International Journal for Chemistry*, vol. 46, pp. 299–303, Jul. 1992, doi: 10.2533/chimia.1992.299.

[53] D. E. Clark, “Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption,” *J Pharm Sci*, vol. 88, no. 8, pp. 807–814, 1999, doi: 10.1021/js9804011.

[54] P. Ertl, B. Rohde, and P. Selzer, “Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties,” *J Med Chem*, vol. 43, no. 20, pp. 3714–3717, Oct. 2000, doi: 10.1021/jm000942e.

[55] S. Winiwarter, N. M. Bonham, F. Ax, A. Hallberg, H. Lennernäs, and A. Karlén, “Correlation of human jejunal permeability (in vivo) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach,” *J Med Chem*, vol. 41, no. 25, pp. 4939–4949, Dec. 1998, doi: 10.1021/jm9810102.

[56] V. Fedi *et al.*, “Discovery of a new series of potent and selective linear tachykinin NK 2 receptor antagonists,” *J Med Chem*, vol. 50, no. 20, pp. 4793–4807, Oct. 2007, doi: 10.1021/jm070289w.

[57] C. R. W. Guimarães, A. M. Mathiowetz, M. Shalaeva, G. Goetz, and S. Liras, “Use of 3D properties to characterize beyond rule-of-5 property space for passive permeation,” *J Chem Inf Model*, vol. 52, no. 4, pp. 882–890, Apr. 2012, doi: 10.1021/ci300010y.

[58] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, “Molecular properties that influence the oral bioavailability of drug candidates,” *J Med Chem*, vol. 45, no. 12, pp. 2615–2623, Jun. 2002, doi: 10.1021/jm020017n.

[59] J. Kelder, P. D. Grootenhuis, D. M. Bayada, L. P. Delbressine, and J. P. Ploemen, “Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs,” *Pharm Res*, vol. 16, no. 10, pp. 1514–1519, Oct. 1999, doi: 10.1023/a:1015040217741.

[60] B. C. Doak, B. Over, F. Giordanetto, and J. Kihlberg, “Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates,” *Chem Biol*, vol. 21, no. 9, pp. 1115–1142, Sep. 2014, doi: 10.1016/j.chembiol.2014.08.013.

[61] T. Steiner, “The Hydrogen Bond in the Solid State,” *Angewandte Chemie*, vol. 41, no. 1, pp. 49–76, Jan. 2002, doi: 10.1002/1521-3773(20020104)41:1<48::aid-anie48>3.0.co;2-u.

[62] C. Deliang *et al.*, "Effects of intramolecular hydrogen bonds on lipophilicity," *European Journal of Pharmaceutical Sciences*, vol. 130, pp. 100–106, Mar. 2019, doi: 10.1016/j.ejps.2019.01.020.

[63] M. Shalaeva *et al.*, "Integrating intramolecular hydrogen bonding (IMHB) considerations in drug discovery using ΔlogP as a tool," *J Med Chem*, vol. 56, no. 12, pp. 4870–4879, Jun. 2013, doi: 10.1021/jm301850m.

[64] D. Garcia Jimenez *et al.*, "Chamelogk: A Chromatographic Chameleonicity Quantifier to Design Orally Bioavailable Beyond-Rule-of-5 Drugs," *J Med Chem*, vol. 66, no. 15, pp. 10681–10693, Jul. 2023, doi: 10.1021/acs.jmedchem.3c00823.

[65] T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey, and M. P. Jacobson, "Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: Successful in silico prediction of the relative permeabilities of cyclic peptides," *J Am Chem Soc*, vol. 128, no. 43, pp. 14073–14080, Nov. 2006, doi: 10.1021/ja063076p.

[66] P. V. Desai, T. J. Raub, and M. J. Blanco, "How hydrogen bonds impact P-glycoprotein transport and permeability," Nov. 01, 2012. doi: 10.1016/j.bmcl.2012.08.059.

[67] M. H. Abraham and R. J. Abraham, "A simple and facile NMR method for the determination of hydrogen bonding by amide N-H protons in protein models and other compounds," *New Journal of Chemistry*, vol. 41, no. 14, pp. 6064–6066, 2017, doi: 10.1039/c7nj01044c.

[68] R. C. Young *et al.*, "Development of a New Physicochemical Model for Brain Penetration and Its Application to the Design of Centrally Acting H2 Receptor Histamine Antagonists," *J Med Chem*, vol. 31, no. 3, pp. 656–671, Mar. 1988, doi: 10.1021/jm00398a028.

[69] K. Valkó, "Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution," *J Chromatogr A*, vol. 1037, no. 1–2, pp. 299–310, May 2004, doi: 10.1016/j.chroma.2003.10.084.

[70] J. W. Mcfarland *et al.*, "Quantitative Structure-Activity Relationships among Macrolide Antibacterial Agents: In Vitro and in Vivo Potency against Pasteurella multocida," *J Med Chem*, vol. 40, no. 9, p. 28, 1997, doi: 10.1021/jm960436i.

[71] R. D. Vaishya, A. Mandal, M. Gokulgandhi, S. Patel, and A. K. Mitra, "Reversible hydrophobic ion-paring complex strategy to minimize acylation of octreotide during long-term delivery from PLGA microparticles," *Int J Pharm*, vol. 489, no. 1–2, pp. 237–245, Jul. 2015, doi: 10.1016/j.ijpharm.2015.04.075.

[72] G. Ermondi *et al.*, "Rifampicin as an example of beyond-rule-of-5 compound: Ionization beyond water and lipophilicity beyond octanol/water," *European Journal of Pharmaceutical Sciences*, vol. 161, Jun. 2021, doi: 10.1016/j.ejps.2021.105802.

[73] A. Denninger, U. Westedt, and K. G. Wagner, "Shared IVIVR for five commercial enabling formulations using the bipha+ biphasic dissolution assay," *Pharmaceutics*, vol. 13, no. 2, pp. 1–22, Feb. 2021, doi: 10.3390/pharmaceutics13020285.

[74] J. Gharbi-Benarous, M. Delaforge, C. K. Jankowski, and J.-P. Girault, "A Comparative NMR Study between the Macrolide Antibiotic Roxithromycin and Erythromycin A with Different Biological Properties," *J Med Chem*, vol. 34, no. 3, pp. 1117–1125, 1991, doi: 10.1021/jm00107a036.

[75] K. M. Hosny, "Development of Saquinavir Mesylate nanoemulsion-loaded transdermal films: Two-step optimization of permeation parameters, characterization, and ex vivo and in vivo evaluation," *Int J Nanomedicine*, vol. 14, pp. 8589–8601, 2019, doi: 10.2147/IJN.S230747.

[76] M. C. Dodd, M. O. Buffle, and U. Von Gunten, "Oxidation of antibacterial molecules by aqueous ozone: Moiety-specific reaction kinetics and application to ozone-based wastewater treatment," *Environ Sci Technol*, vol. 40, no. 6, pp. 1969–1977, Mar. 2006, doi: 10.1021/es051369x.

[77] K. L. Valko, "Application of biomimetic HPLC to estimate in vivo behavior of early drug discovery compounds," *Future Drug Discov*, vol. 1, no. 1, Jul. 2019, doi: 10.4155/fdd-2019-0004.

[78] R. J. Young, D. V. S. Green, C. N. Luscombe, and A. P. Hill, "Getting physical in drug discovery II: The impact of chromatographic hydrophobicity measurements and aromaticity," Sep. 2011. doi: 10.1016/j.drudis.2011.06.001.

[79] A. H. Kahns, A. Buur, and H. Bundgaard, "Prodrugs of peptides. 18. Synthesis and evaluation of various esters of desmopressin (dDAVP)," *Pharm Res*, vol. 10, no. 1, pp. 68–74, Jan. 1993, doi: 10.1023/a:1018973029651.

[80] K. Valko, C. M. Du, C. Bevan, D. P. Reynolds, and M. H. Abraham, "Rapid Method for the Estimation of Octanol/Water Partition Coefficient (Log P oct) from Gradient RP-HPLC Retention and a Hydrogen Bond Acidity Term (Σα 2 H)," *Curr Med Chem*, vol. 8, no. 9, pp. 1137–1146, 2001.

[81] K. L. Valkó, "Lipophilicity and biomimetic properties measured by HPLC to support drug discovery," *J Pharm Biomed Anal*, vol. 130, pp. 35–54, Oct. 2016, doi: 10.1016/j.jpba.2016.04.009.

[82] K. Taguchi, E. Fukusaki, and T. Bamba, "Supercritical fluid chromatography/mass spectrometry in metabolite analysis," *Bioanalysis*, vol. 6, no. 12, pp. 1679–1689, 2014, doi: 10.4155/bio.14.120.

[83] K. M. Shiva, M. Suraj, S. Meenakshi, S. Sushant Kumar, and D. Siddharth, "Solid dispersion: A new approach for bettering poor drug solubility," *NeuroQuantology*, vol. 20, no. 15, pp. 5567–5573, Nov. 2022, doi: 10.14704/NQ.2022.20.15.NQ88560.

[84] G. L. Amidon, H. Lennernäs, V. P. Shah, and J. R. Crison, "A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability," *Pharm Res*, vol. 12, no. 3, pp. 413–420, Oct. 1994, doi: 10.1023/a:1016212804288.

[85] E. H. Kerns, L. Di, S. Petusky, M. Farris, R. Ley, and P. Jupp, "Combined application of parallel artificial membrane permeability assay and Caco-2 permeability assays in drug discovery," *J Pharm Sci*, vol. 93, no. 6, pp. 1440–1453, 2004, doi: 10.1002/jps.20075.

[86] B. Press and D. Di Grandi, "Permeability for intestinal absorption: Caco-2 assay and related issues," *Curr Drug Metab*, vol. 9, no. 9, pp. 893–900, Nov. 2008, doi: 10.2174/138920008786485119.

[87] S. Hariharan, S. Gunda, G. P. Mishra, D. Pal, and A. K. Mitra, "Enhanced corneal absorption of erythromycin by modulating P-glycoprotein and MRP mediated efflux with corticosteroids," *Pharm Res*, vol. 26, no. 5, pp. 1270–1282, May 2009, doi: 10.1007/s11095-008-9741-x.

[88] K. Joyce, "Center for Drug Evaluation and Research; Application Number: 22-554; Clinical Pharmacology and Biopharmaceutics Review(s)," Jun. 2009. [Online]. Available: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2010/022554orig1s000sumr.pdf