UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA

Ph.D. Course in
Drugs, Biomolecules and Health Products
XXXI CYCLE

An integrated approach for physico-chemical profiling in drug discovery

Coordinator:
Chiar.mo Prof. Marco Mor

Supervisor:
Chiar.ma Prof.ssa Federica Vacondio

External Advisor:
Chiar.ma Dr.ssa Valentina Mileo

Ph.D. student: Francesca Ferlenghi

2015-2018

Summary

Abstract

In the last 20 years, small molecules developed as potential new drugs have shown an increasing chemical complexity in comparison to compounds optimized before the year 2000. Current research trend is well exemplified by ampholytes, for which molecular and biological characterisation is anything but trivial.

The aim of the present Ph.D. thesis was to outline an in-house integrated work-flow for physico-chemical profiling of new drug candidates at discovery and early development stage. Screening included ionization constants (pKas), lipophilicity (log *P* or log $D_{7.4}$) and solubility.

pKas were measured by combining potentiometry and spectrophotometry. For special needs, deductive approach or an *ad hoc* UV-metric assay was set up to determine site-specific constants and ion pairing propensity, respectively.

Log *P* was assessed by potentiometry or by an automatic shake-flask method. The latter had been optimized on a robotic system to save substance, time and manual liquid-handling; sample analysis was performed by liquid chromatography coupled with mass spectrometry.

Finally, solubility was measured by turbidimetric assay or by UV-spectrophotometry and characterisation of solution-mediated interconversion between different solid forms was attempted applying X-ray powder diffraction and Raman spectroscopy.

This in-house screening of physico-chemical properties proved valuable for a pre-clinical research department, since it provided reliable data for pre-development (e.g. guidelines for salt screening and solid form optimization). Moreover, experimental results from the present work will be included in the calibrating set of programs for in-house prediction of molecular properties (pKas, lipophilicity) and for *in-vitro-in-vivo* correlation (solubility). In fact, inclusion of discovery compounds is the key to develop expert systems and that is why experimental assays for measuring physico-chemical parameters are still warranted.

1. Relevance of physico-chemical profiling in pre-clinical research

1.1 Role of physico-chemical properties in current drug discovery scenario

Combinatorial chemistry, also named high-throughput organic synthesis or multiple parallel synthesis (MPS), has emerged in the 1990s and represents the current drug discovery paradigm. That strategy provides large numbers (well above 10^5) of new chemical entities (NCEs) that need to be tested. To meet this need, implementation of robotic systems, development of novel assay technologies and extensive miniaturization has been occurring over the last decade, especially as regards lipophilicity [1] and solubility profiling [2, 3]. This series of automatic screening techniques are collectively named high-throughput screening assays (HTS).

Drug discovery pathway is a long step-wise process: first, screening assays are performed on hits compounds and those showing activity during HTS are selected. Multiple hits belonging to the same chemical family set the basis of a lead series (*i.e.* structures that are suitable for further chemical optimization and exhibit favourable patent situation). Leads are further optimized and then access the drug candidate status that typically regards one over 10^4 screened compounds. Then approval from regulatory agencies is needed to enter clinical trials and if all the three phases proved successful (*i.e.* usually 1 over 10 candidates) the compound is approved and becomes a marketed drug [4].

Over the last two decades, small molecules developed as potential new drugs have shown an increasing chemical complexity in comparison to compounds optimized before the year 2000.

Recently developed drug candidates have higher molecular weight (MW) and a higher number of heavy atoms and ring structures than before. In addition, they are more flexible, more hydrophobic and, as a consequence, less soluble, resulting in what Hann called “obese molecules” [5].

Ampholytes offer a prime example of this pharmaceutical research trend: the presence of several ionisable sites leads to the co-existence of differently charged states of the molecule. That can affect not only the physico-chemical behaviour of the potential drug, but also its biological activity and biopharmaceutical performance [6], [7], [8].

The increasing molecular complexity of drug candidates has led medicinal chemists to adopt a holistic approach to discovery, in which absorption, distribution, metabolism, elimination and toxicity (ADMET) are optimized in parallel with target affinity [9]. Combining structure-based design, which controls biological activity and potency, and property-based design, which manipulates structural features trying to optimize absorption and pharmacokinetics, has currently become the preferred approach in lead optimization. In this way, the main causes for attrition in the late 1990s (*i.e.* pharmacokinetics and bioavailability) have been largely controlled, but, owing to molecular obesity, toxicity and promiscuity are now regarded as the main issues [10].

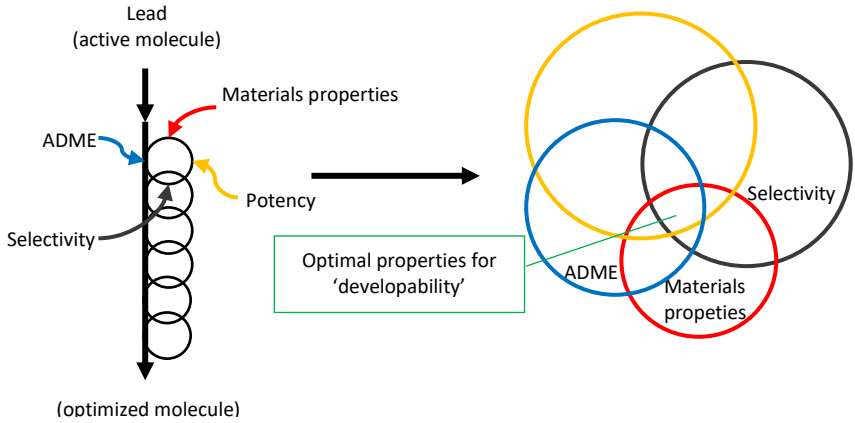


Figure 1: Graphical scheme of holistic approach to drug discovery. Picture from Kerns E. H and Di L., Drug-like properties: concepts, structure, design and methods from ADME to toxicity optimization, 2008, Elsevier Inc.

Estimation of how likable it is for a compound to exhibit low permeability or absorption on the basis of its physico-chemical properties is a concept regarded as "drug-likeness", introduced by Lipinski in 1997 referring to

orally administered drugs [11]. He and co-workers analysed a wide database of drugs and late-stage drug candidates, seeking for which properties were responsible for attrition. According to Lipinski and co-authors, a compound is expected to undergo high passive diffusion and oral absorption if it fits a concise set of calculated parameters (*i.e.* molecular weight MW ≤ 500; calculated partition coefficient clog *P* ≤ 5, number of hydrogen donors ≤ 5, and number of hydrogen acceptors ≤ 10). Those are known as "Rule-of-five" (Ro5) guidelines for drug-likeness and had a dramatic impact on medicinal chemistry and, in particular, on combinatorial chemistry and High Throughput Screening (HTS) [12].

It has been pointed out, however, that Ro5 could fail in predicting the fate of drug candidates in human environment: for instance, orally administered drugs such as tetracycline and rifampicin are absorbed through intestinal barrier despite violating established guidelines [13]. Moreover, Lipinski himself noticed that natural products metabolites (e.g. cyclosporine A) act as "chemical-chameleon". That means they are able to remarkably change shape, polarity and intra-molecular hydrogen bond pattern according to the surroundings, thus being absorbed despite violating Ro5 chemical space [14].

Ro5-type models have provided useful tools for estimation of "drug-likeness" for compound and "druggability" for target predictions and recently similar criteria for non-oral routes of delivery (e.g. ophthalmic, nasal and transdermal routes) have been highlighted [15].

Identification and expansion of chemical areas that contain specific drug-like molecules with adequate physicochemical properties is truly challenging although useful in drug-discovery. That urgency has led to different declination of Ro5 to identify discrete molecularly-diverse areas to serve medicinal chemistry: added to drug-like chemical space defined by Lipinski's Ro5, lead-like chemical space (*i.e.* criteria for lead-likeness, defined by the Rule of four Ro4), fragment chemical space (*i.e.* criteria for fragments, called the Rule of three Ro3 and building blocks Rule of two, Ro2) have been proposed [10].

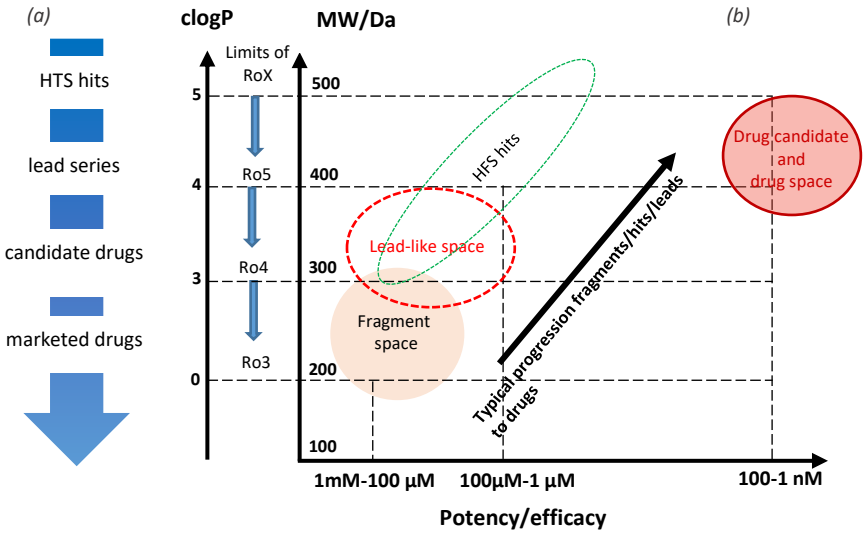


Figure 2: (a) Drug discovery paradigm and (b) progression of the optimization from fragments/hits/leads to drugs; picture taken from Mignani's paper [17].

In recent years, computational models have been increasingly sought to get information on structural features related to physico-chemical and biopharmaceutical properties: for instance, since 1980s correlation between lipophilicity and target affinity or permeation has been extensively explored for drugs with both peripheral and central effects [16]. Molecular descriptors (e.g. partition coefficient log *P*, polarizability, dipolar moment) have been integrated with constitutional descriptors (e.g. counts of heteroatoms, rings, hydrogen bond donors and acceptors) in chemometric models that are supposed to return useful guidelines for the synthesis of active compounds.

In the context of holistic approach, recent literature has been reporting some improved criteria of drug-likeness that a potential candidate should meet in order to access the following stage of selection. Common filters regard potency, selectivity, hydrophobicity, solubility, MW, number of aromatic rings or heteroatoms and parameters of permeability from Caco-2 cell assays; metabolic descriptors related to interaction with cytochrome P450 and toxicity indications derived from interaction with hERG [17].

Waring and co-authors examined data regarding drug attrition for > 800 oral small-molecule drug candidates from four major pharmaceutical companies and stressed how calculation of physico-chemical properties during optimization plays a major role in prioritizing most promising drugs [10].

Since Lipinski's Rule-of-five was proposed, many other rules of thumb arose based on the analysis of different data set and, beyond that, a comparison of the distribution of different properties for drug-like and non drug-like molecules has been attempted. By examining marketed compounds, privileged scaffolds and substructures have been identified and selected as suitable for new candidates, but it is apparent that following the principle of similarity novel scaffolds may be lost in the chemical space [18].

In that sense, substructures of brand new chemical entities (NCEs) or their descriptors may not be properly represented by training sets of software for *in silico* calculation of physico-chemical properties, a condition that often leads to high errors in prediction and, in turn, to a poor drug-likeness.

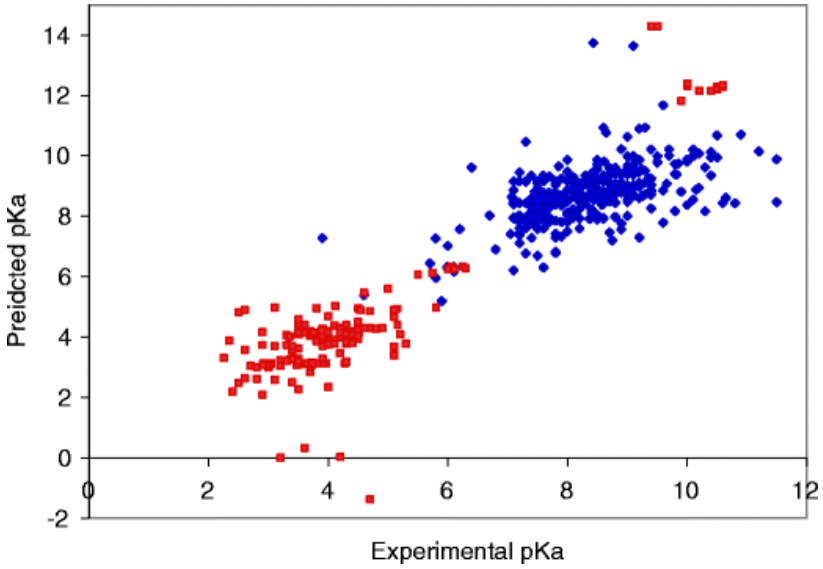


Figure 3: Correlation between computed and experimental pKa data from Settimo's paper [19]. Red and blue squares represent acidic and basic compounds respectively.

Settimo and co-authors [19] built a chemically diverse and drug-like dataset from high-throughput UV–vis spectrophotometry pKa data and correlated such values with pKa predicted by Epik software (see Figure 3).
Although the correlation coefficient for the whole dataset is 0.84, R^2 was 0.83 for acids and only 0.35 for bases, which indeed represent the majority of currently developed drug candidates. The authors speculated that unpleasant accuracy in prediction should be ascribed to lack in diversity and chemical complexity in the database used to train Epik software.
Urgency for implementation of training sets has led, since year 2000, to the development of re-trainable software for physico-chemical property calculation, such as XLOGP3 developed by Cheng and co-authors for partition coefficient prediction [20]. The inclusion of drug discovery compounds in the calibrating set of a software is essential to develop expert systems and that is why experimental methods for measuring physico-chemical parameters are still warranted.
This Ph.D. thesis will focus on three physico-chemical properties: dissociation constant (pKa), lipophilicity (log *P* and log $D_{7.4}$) and solubility. Relevance in drug discovery and early development, as well as impact in biological systems and methods of assessment, will be discussed in the following paragraphs.

1.2 Relevant physico-chemical properties: dissociation constants (pKa)

1.2.1 Relevance in discovery and early development

Dissociation of an ionisable compound in solution can be described as a step-wise process, as summarized in Figure 4 for a weak acid (HA). In a generic solvent (S), four consecutive equilibria (parameterized by Kn) contribute to the overall dissociation:

$$HA + S \overset{K_1}{\rightleftharpoons} [HA\cdots S] \overset{K_2}{\rightleftharpoons} [A^-SH^+] \overset{K_3}{\rightleftharpoons} [A^-][SH^+] \overset{K_4}{\rightleftharpoons} A^- + SH^+$$

Figure 4: Scheme of dissociation process.

First, a complex between solvent and acid is promoted by electrostatic and dispersive interactions (step 1), as hydrogen bonding and Van der Waals forces. Then a primary ionization occurs by proton transfer and a contact-ion pair is formed (step 2). Depending on solvent and drug properties, interactions between ionic pair is weakened by the action of the solvent hence each ion is surrounded by a shell of medium (step 3) and solvation progresses until separated charges are liberated into the solvent (step 4).
Side reactions may be promoted by solvent properties (*i.e.* low solvation ability, low polarity): conjugation may occur between the ionized forms of the same compound (homoconjugation) or different compounds (eteroconjugation); similarly neutral, anionic or cationic species could cluster in self-association phenomena. Conjugation and association are more likely for multiprotic compounds with low charge-delocalization and easily accessible protonation sites.
Added to that, solvent properties widely affect dissociation progression: for instance, organic solvents (e.g. halogenated-alkanes) may not overcome interactions between charges so that the process stops at the intermediate step of ion-pairing. On the contrary, water is a polar solvent with high solvation capacity that results in complete dissociation of anions and cations, which is the common scenario found in biological systems.

Usually referred as pKa(s), dissociation constant(s) define the ratio between the charged and uncharged state of a drug molecule:

$$\mathrm{pKa} = -\log\frac{[A^-]}{[HA]} \qquad (1.1)$$

for a monoprotic acid

$$\mathrm{pKa} = -\log\frac{[B]}{[BH^+]} \qquad (1.2)$$

for a monoprotic base

Equation 1

Dissociation constants define the charged state of a NCE at physiologically relevant pHs, contributing to predict and understand its *in vitro* and *in vivo* behaviour: for instance, it deeply influences drug ADME behaviour and contributes to both its pharmacological affinity to enzymes or receptors and to unwanted off-target effects. Albeit underrated, acid/base behaviour could remarkably affect potency especially in the case of aforementioned ampholytes and zwitterions. In that sense, docking studies should take into account protomeric equilibria in modelling the ligand-target interaction: in fact, protein groups surrounding the binding site strongly impact on the local pH (also named micro-pH) which in turn determines the protonation state of the bound ligand.

To keep computation manageable, unreasonably high-charged structures should be excluded before modelling and only calculation or, better than that, experimental determination of pKas values should allow this selection [21].

Moreover, it has been suggested that ionic species (e.g. anions of carboxylic acids) are able to affect permeation and consequently pH-partition hypothesis according to which only neutral forms are able to drive permeation across lipophilic membranes by passive diffusion has been reconsidered [22].

(De)protonation could have a role in resistance mechanisms, as reported for antiangiogenic tyrosine-kinase inhibitor sunitinib. Fluorescence microscopy proved this anticancer agent to be sequestered into lysosomes and this founding could be rationalized considering acid-base equilibria in solution (weak base, pKa = 8.95) of such drug [23].
From a formulative perspective, salification is a common practice to increase solubility of hydrophobic active pharmaceutical ingredients (APIs) as described in the following chapters of the present work. The knowledge of pKas is a key factor in salt screening, which is a crucial step to select those crystallization conditions that could best optimize salification and thus solubility and dissolution rate of the new drug candidate.

1.2.2 pKa determination

1.2.2.1 Computational approach

Several algorithms have been developed over years with the purpose of predicting pKa of drugs before their synthesis. The crucial issue for predictive tools is their validation as it requires accurate experimental data preferably obtained by a consistent method (*i.e.* same experimental conditions). In a recent paper [19], Settimo and co-workers considered four literature datasets, made of results from alternative methods, and an additional drug-like dataset, experimentally compiled in their laboratory. They noticed that published database often show some errors in comparison to the publications they took results from, an issue that can be overcome by using in-house dataset.
Moreover, if multiprotic compounds are included in the training set, a site-specific attribution of each pKa value is mandatory to make calibration effective.
Among the algorithms available for pKa predictions, most of them are based on Hammett-Taft equation, such as ACD pKa DB (Advanced Chemistry Development, Inc., Toronto, Ontario, Canada) and Epik (Schrödinger, New York, USA).

Well established and complementary Hammett-Taft (HT) approach have been used for decades to predict pKa of aromatic and aliphatic molecules respectively. HT equation (reported below) defines pKa of an acidic or basic functional group (ABG) as a function of the attached substituent in comparison to hydrogen atom:

$$\mathrm{pKa_i} = \mathrm{pKa_0} + \mathrm{s_i} - \rho_o \sum_j \sigma_j$$

Equation 2

where pKa_0 represents a constant value and ρ is the response parameter, which indicates how sensitive the pKa for this ABG is to the attached substituents. For each substituent j, there are σ parameters that reflect how strongly it perturbs ABGs. pKa_0, ρ, and σ parameters are nearly always determined by fitting to experimental pKa data.

σ parameter depends on whether the substitution is on an aliphatic portion of the molecule (σ*) or on an aromatic ring system. In the latter case, three σ terms are used according to the site that brings the substituent (*i.e.* σ_o for ortho position, σ_m for meta position and σ_p for para position) in relation with the ABG.

s is a statistical factor which accounts for differences in the number of hydrogen atoms equivalent to the one involved in the ionization of this ABG in its acidic (n_a), and basic (n_b) forms:

$$s = -\log(n_a/n_b) \qquad [24]$$

Equation 3

Epik can also be employed to generate and predict different protonation states and tautomers for the drug.

Chemaxon (Chemaxon, Budapest, Hungary) predicts the microspecies distribution by additional increments from partial charge, polarizability and structure specific features of the molecule [19].

Program	Company	Method
ACD/pKa DB	Advanced Chemistry Development, Inc.	LFER
ADME boxes	Pharma Algorithms, Inc.	QSPR
ADMET Predictor	Simulations Plus, Inc.	QSPR
Epik	Schrödinger, LLC.	LFER
Jaguar	Schrödinger, LLC.	quantum-chemical method (DFT) with empirical correction
Marvin	ChemAxon Ltd.	QSPR
Pallas pKalc	CompuDrug International, Inc.	LFER
Pipeline Pilot	SciTegic, Inc.	QSPR
SPARC	University of Georgia/U.S. Environmental Protection Agency	blend of LFER and perturbed molecular orbital (PMO) method

Table 1: Main programs for pKa calculation as reported in Liao and Nicklaus's paper [25]. LFER linear free energy relationship; QSPR quantitative structure-property relationship.

Settimo's group compared measured pKa value with predictions from Epik, Chemaxon and ACD pKa DB software packages in terms of correlation coefficient and Median Absolute Deviation, as an index of prediction error with reduced sensitivity to random outliers. Differences between experimental and computed data was significant and mainly hinged on the lack of diversity and chemical complexity in the data set.

Dataset	# Unique compounds	# Acidic centres predicted	# Basic centres predicted
Vertex*	477[a]	167	330
Liao	105[b]	43	71

Avdeef	122[c]	49	86
Morgenthaler	174	0	174
Luan	67	0	67

*Table 2: Composition of the datasets. * in-house dataset made of high-throughput UV–metric data. (a) 20, (b) 9 and (c) 13 compounds in the Vertex, Liao and Avdeef datasets contained one basic and one acidic centre within the same molecule [19].*

In particular, higher errors in predictions were found for basic rather than acidic drugs because of the lower solubility and the wider range of pKa values.

Thus, for the sake of accuracy, it is advisable to consider the two classes separately during development and validation of improved predictive tools. Added to that Caron and Ermondi emphasized that pKas of structures with more than two 5-6-member rings and 4–5 nitrogen atoms were poorly predicted by any software [26].

Hence, the experimental way seems the only available one for an accurate determination of ionization equilibria of complex molecules.

1.2.2.2 Experimental approach: potentiometric "pH-metric" method by using Sirius instruments

2a Effect of ionic strength, temperature and carbon dioxide concentration on titration

Even if there are several methods for the experimental determination of dissociation constants, potentiometric titration (also known as pH-metric method) is regarded as the gold-standard for its repeatability and reliability [27].

In the 1990s, Avdeef and co-workers implemented an automated hardware-software titrating system called PCA 101 (Sirius, UK) for potentiometric determination of pKas and partition coefficients (log *P*) [28, 29, 30, 31, 32]. The updated version of the instrument, called first GLpKa™

and, more recently, T3™, is nowadays widespread for potentiometric measurements of such physico-chemical constants.

The main elements of the system consist of a combined glass electrode and automatic micro-dispensers for acidic (HCl 0.5 N) and basic (KOH 0.5 N) titrants. Titrations are performed in solution with constant ionic strength (*I* = 0.15 M for KCl) at room temperature (25 ± 0.5 °C) under a constant flow of an inert gas (typically nitrogen) in both the base reservoir and the sample.

Several experimental parameters of the potentiometric assay, such as temperature and ionic strength, affect electrode calibration and, in turn, pKa measurement. Ionic strength of the solution should be kept relatively high (in the order of 0.1 M) by using a so-called "common salt" or "background electrolyte". Hence, activity coefficients of sample species at low concentrations are kept stable. Added to that the junction potential between sample solution, medium and 3.5 M KCl bridge solution (that fills the electrode) is stabilized.

Different common salts could be used with different benefits: for instance, nitrates (KNO_3, $NaNO_3$) and perchlorates ($KClO_4$, $NaClO_4$) do not promote coordination that may affect measurements; potassium salts, such as KCl, reduce "sodium error" at high pH and are, nowadays, extensively used.

As any thermodynamic constant, Ka is affected by temperature that should be accurately controlled by means of a thermostatic bath. Room temperature has been widely used in literature and hence such condition has been conventionally adopted for the purpose of comparability.

Furthermore, potentiometric pKa results are affected by carbon dioxide content in the sample system.

$$CO_2\,(g) \rightleftharpoons CO_2\,(aq) \qquad (4.1)$$

$$CO_2\,(aq) + H_2O \rightleftharpoons H_2CO_3\,(aq) \qquad (4.2)$$

$$H_2CO_3\,(aq) \rightleftharpoons H^+ + HCO_3^-\,(aq) \qquad (4.3)$$

$$HCO_3^-\,(aq) \rightleftharpoons H^+ + CO_3^{2-}\,(aq) \qquad (4.4)$$

$$[CO_2\,(aq)] = K \cdot p(CO_2) \qquad (4.5)$$

Equation 4

Carbon dioxide dissolves in water according to Henry's law (see Equation 4.5, with K = 0.03 M/bar): with the partial pressure of carbon dioxide in the atmosphere ($p(CO_2)$) usually being $\sim 4 \cdot 10^{-4}$ bar, the equilibrium concentration of carbon dioxide in water would be around 10^{-5} M. If the concentration of neutral carbonic acid is greater than this, carbon dioxide gas gradually transfers to the atmosphere. Conversely, if the concentration is lower, carbon dioxide gas is progressively absorbed into solution (see Equation 4.2). Equilibrium between carbonic acid in solution and carbon dioxide in the gas phase affects the pH of the sample during titration: when the system is exposed to the atmosphere, dissociation equilibrium of carbonic acid shifts toward neutral or ionized form according to the pH of the solution, thus the concentration of dissolved neutral carbon dioxide changes. Consequently, there is a change in the rate at which the gas is absorbed from the atmosphere or evolved into it affecting pH measurements.
In order to prevent excessive carbonation of the sample solution and the basic titrant, the system is kept under a constant stream of nitrogen.

2b Calibration of the combined glass electrode

A combined glass electrode is employed in potentiometric titration and its calibration is performed by the so-called Four-Plus ™ procedure that Avdeef described in his papers [29-32]. First, an operational pH scale is established by calibrating the pH measuring circuit with a single aqueous phosphate buffer (pH 7.0) and assuming the Nernst's slope (*i.e.* a linear correlation between the reduction potential and pH):

$$E = E^0 + \frac{2.303RT}{F} \cdot pH$$

Equation 5

where E is the redox potential at the working temperature T in kelvin, E^0 is the standard potential, R is the gas constant and F is the Faraday constant.

Nernst slope is approximately equal to 59.16 mV at 25 °C and is sensitive to temperature that should be strictly controlled.
Then a strong acid-strong base titration (*i.e.* blank) is performed, from which proton concentration is calculated and used to relate the operative pH scale (pH) to the concentration pH scale ($p_cH = -\log[H^+]$) by the following equation:

$$pH = \alpha + S \cdot p_cH + j_H \cdot [H^+] + j_{OH} \cdot K_W/[H^+]$$

Equation 6

where α corresponds to the negative logarithm of the activity coefficient of protons at working temperature and ionic strength; *S* term represents the deviation of the actual slope from the ideal Nernst's one; the j_H accounts for acidic error, which is the non-linear pH response due to liquid junction and asymmetry potentials in moderately acidic solution (pH = 1.5–2.5); in a smiliar way, the j_{OH} parameter corrects for basic error (*i.e.* any non-linear effect at alkaline pH > 11).
Blank Refinement proceeds as follows: calculation starts from ideal Four Plus values, which are re-calculated iteratively. The HCl concentration factor and the carbon dioxide concentration are allowed to vary and during this procedure the pH of every point in the titration curve is calculated and compared with the measured pH. Eventually the differences between measured pH and calculated pH are minimised across the entire pH range.
Goodness of blank refinement is parameterized by Root Mean Square Deviation (RMSD) that is calculated after each iteration:

$$RMSD = \sqrt{\frac{\sum(R/BI)^2}{N}}$$

Equation 7

where R is the difference between observed and calculated pH, BI is the buffer index (*i.e.* the ratio between the added volume of the titrant and the

change in pH that derived from it) and N is the number of pH measurements.
The four terms are calculated by a weighed non-linear least square method importing constant for water autoprotolysis (K_W) by Sweeton's publication [33].
When RMSD value falls between 0.1 and 1 the fit is optimised and the Four-Plus values obtained are used to calculate p_cH values.

2c pKa determination

Information of titration curve (*i.e.* pH versus volume of titrant) are processed to return changes in drug ionization as a function of pH. The resulting graph, named Bjerrum or difference plot, represents the mean molecular charge that is the proportion of protonated molecules of the analyte as a function of pH. Data are fitted to Bjerrum plot to approximate equilibrium constants that serve as "seed" values for further optimization. Values are iteratively refined until sum of weighed squares of residuals (S, reported in Equation 8) is minimized:

$$S = \sum_{i}^{N_0} \frac{(pH_{i\,obs} - pH_{i\,calc})^2}{\sigma_i^2(pH)}$$

Equation 8

In Equation 8 N_0 is the number of experimental pH measures, σi^2 is the variance estimated for measured $pH_{i\ obs}$. The model equation $pH_{i\ calc}$ is a function of pKa and independent variables.
Weighting scheme used in Equation 8 in based on variances calculated in the following equation:

$$\sigma_i{}^2\,(pH) = \sigma_c{}^2 + (\sigma_v \cdot \frac{dpH}{dV})^2$$

Equation 9

where σ_c and and σ_v are fixed contribution related to measured pH and incremental volume of titrant, and are equal to $5 \cdot 10^{-3}$ units of pH and $2 \cdot 10^{-}$

^{4} mL respectively. Weighting system takes into account the higher variance of pH-measurements close to the end-point (*i.e.* dpH large for minimum dV).
At the end of each iterative cycle, progress of refinement is expressed as Goodness Of Fit (GOF) that is calculated as follows:

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

Equation 10

where Nr is the number of refined parameters (*i.e.* pKas). Ideal value for GOF equals 1 that means calculated and observed pH curve differs from about one standard deviation unit in pH (*i.e.* difference of about 0.005 pH unit in the buffer region). Convergence of non-linear equations is reached when solutions from two consecutive iterative cycles returned $\Delta GOF < 10^{-5}$ or when all refined parameters showed a calculated shift $< 10^{-4}$. Generally 15-50 iterations are required to match these goal values. Titrations are usually repeated three to five times and data are cumulatively refined by the polynomial function called MultiSet.

2d Solubility issue and use of co-solvent

Potentiometric method is suitable for concentrations down to 10^{-4} M, but generally more accurate results are obtained for more concentrated solutions within 1 to $5 \cdot 10^{-3}$ M. Solubility could be a hurdle in purely aqueous environment, so that co-solvent (*i.e.* MeOH) addition may be required.
Co-solvent method [29] is based on the abovementioned principles, but calibration of the electrode should account for percentage of organic solvent (R parameter in Equation 11) used in the assay. The following polynomial equation is then used:

$$P = \sum_{i=0}^{n} P_i \cdot R^i$$

Equation 11

where P is any of the four parameters in presence of co-solvent and n is the degree of the used polynomial (usually a good fitting is obtained for $n \leq 4$). In theory, calibration should be performed in each organic solvent at any percentage used, but the software-hardware package provided by Sirius is already equipped with a library of values for most common organic solvents (e.g. methanol, ethanol, dioxane, DMSO, ...) that are automatically imported in calculation.

In presence of co-solvent, an apparent pKa value called p_sKa is determined and needs to be related to purely aqueous condition to be meaningful.

Yasuda [34] and Shedlovsky [35] independently derived an extrapolation method based on linear interpolation of experimental p_sKa points obtained at three or more different percentages of co-solvent.

They related p_sKa to the dielectric constant (ε) of water/organic solvent mixture, including the molar concentration of water ($[H_2O]$) in Equation 12:

$$p_sKa + \log [H_2O] = A/\varepsilon + B$$

Equation 12

where A and B represent the slope and intercept of linear extrapolation, respectively.

Organic solvent decreases the dielectric constant of the mixture in comparison to pure water and, in turn, weakens acidic or basic strength of the ionisable site (*i.e.* $p_sKa < pKa$ for a base; $p_sKa > pKa$ for an acid). In that sense, value and sign of A slope should suggest the nature of the ionisable centre (*i.e.* $A > 0$ for acids and $A < 0$ for bases).

However, it is quite common to see confounding effects in the case of multiprotic compounds, for which it is hard to assign constants in a site-specific manner; or in large molecules like current NCEs, where the charge is often widely delocalized. Herein discrimination between acidic and basic moieties harvested from such slope is not always reliable.

Linear extrapolation at 0% co-solvent return the aqueous pKa value with an accuracy within ± 0.2 log units. In case of "long-distance extrapolation" (*i.e.* from 40% to 60% co-solvent-rich region) a bias correction has been

proposed [36], but it is a good practice to work at the lowest concentration of co-solvent allowed by solubility of the analyte.
From an experimental perspective, it is worth noticing that co-solvent negatively influences electrode performance, especially if relatively high percentages or very low polarity solvents are used. Moreover, some organic solvents, such as DMSO, act as carbon dioxide sponges thus negatively affecting the accuracy of the measurements.
Finally, in solvent-rich mixtures, solvation is strongly affected and deviation from linearity could be remarkable (e.g. for molecules with phenolic moieties) depending on the chemical properties of the solute [37].

1.2.2.3 Experimental approach: spectrophotometric method

UV-spectrophotometric approach was developed to increase sensitivity, reduce operative concentrations down to 10^{-6} M and improve determination of pKas at the extremes of the pH range of the glass electrode (pH < 3 or pH > 11)
A multi-wavelength spectrum, usually within 200 and 700 nm, is collected at each point of potentiometric titration and dissociation process is monitored as a change in the absorbance profile.
This method often allows to work in purely aqueous solution, but requires the analyte to exhibit at least one chromophore in proximity to the ionization centre(s) (that means n ≤ 3 bonds distance): hence different protonation species show a significant dissimilarity in their spectra.
The experimental apparatus includes a deuterium lamp that serves as light source, a UV-optic dip probe and a diode array detector and it is currently embodied in commercial instruments for potentiometric pKa measurements (Sirius Analytical, UK).

3a Target Factor Analysis (TFA) method

UV-signal from titration experiment is de-convoluted by means of factorial analysis, using an iterative procedure based on matrix calculation called Target Factorial Analysis [38]. Absorbance matrix (A) collects experimental

spectra (N_s) recorded at each wavelength (N_W). It is regarded as the product of concentration of each species at any pH tested (C) and absorptivity coefficient for each light absorbing species (E):

$$A = C \cdot E \quad (13.1)$$
$$A = N_s \times N_W \quad (13.2)$$
$$C = N_s \times N_c \quad (13.3)$$
$$E = N_c \times N_W \quad (13.4)$$

Equation 13

The number of independent absorbing species (*i.e.* principal component, N_c) should be lower than or equal to the number of wavelength (N_W) and the number of spectra (N_s) to prevent overfitting.
By using Principal Component Analysis (PCA) independent species that gave the most significant contribution to the total absorbance are selected. First, co-variance matrix (Z) is calculated by transposing A matrix (A^T):

$$Z = A^T \cdot A$$

Equation 14

Then diagonalization of the covariance matrix gives the eigenvector (Q) and eigenvalue (λ) ones:

$$Z \cdot Q = Q \cdot \lambda$$

Equation 15

Different methods have been applied to define the dimensionality of the factor space that is the selection of those eigenvalues providing the highest contribution to the observed data and the discard of the least relevant as noise. In this context, an example is provided by the empiric indicator (IND) function by Malinowsky, which reaches a minimum when the correct number of factors is employed and depends upon the knowledge of the experimental error associated with the method [39]. Malinowsky defined the real error (RE) as the difference between the error-free data and the

actual experimental data. If the experimental error is estimated with acceptable accuracy, it can be compared to the calculated RE to define the correct number of significant eigenvectors. In other words, the IND function defines the number of eigenvalues required to obtain a random and fairly uniform distribution of error throughout all the data in the A matrix; other options are the eigenvalue ratio [40] or the reduced eigenvalue ratio [41].

3b Target transformation procedure for data refinement

In the end, a goodness of fit (reported in Equation 10) method similar to the one described for pH-metric data refinement was also applied to UV-metric pKas refinement by Sirius Analytical software packages.
From eigenvector (Q_r) and eigenvalue (λ_r) matrices with selected principal components an abstract solution for the absorptivity matrix (E_{abs}) and concentration-pH profile matrix (C_{abs}) is proposed:

$$E_{abs} = Q^T_r \qquad (16.1)$$

$$C_{abs} = A \cdot Q_r \qquad (16.2)$$

Equation 16

The abstract solutions can be rotated to the ones with relevant physical significance (E_p and C_p respectively) by an iterative procedure known as target transformation procedure.
First, concentration test matrix (C_t) is calculated from abstract solutions for absorptivity and concentration-pH profile matrices (E_{abs} and C_{abs}), knowing starting concentration and predicted pKas for the analyte.
Then transformation matrix (T) is obtained by C_t:

$$T = \lambda_r^{-1} \cdot C^T_{abs} \cdot C_t$$

Equation 17

and abstract solutions are converted in ones with physical significance by using the calculated T matrix:

$$A \approx C_{abs} \cdot T \cdot T^{-1} \cdot E_{abs}$$

Equation 18

$$A \approx C_p \cdot E_p$$

Equation 19

$^{-1}$ notation in Equation 17 and Equation 18 denotes the inversion operation. For a particular A matrix, C_t is iteratively refined and effectiveness of refinement is evaluated through a SPOIL function [42]:

$$SPOIL = RET/REP$$

Equation 20

where RET is the RE in the target test vector (C_t) and REP is the RE in the predicted vector(C_p).

In general, optimization through target transformation procedure is regarded as converged to the optimal C_t matrix when a SPOIL function not greater than 3.0 is obtained: optimized C_t leads to the most suitable model for data fitting and, finally, calculation of pKas.

3c Ampholytes: micro- and macro-constants attribution

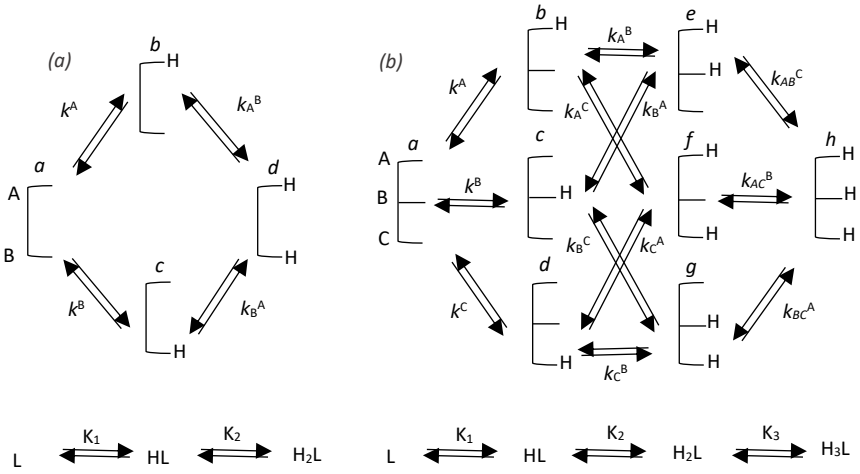


Figure 5: Protonation equilibrium of a bidentate (a) and a tridentate (b) molecule in solution [43].

Spectrophotometric approach proved useful to investigate dissociation equilibria of multiprotic drugs, which offer a prime example of modern research trend in the pharmaceutical field: the presence of several ionization sites leads to complex ionization equilibria in solution that are challenging to investigate.

Such a complex scenario (depicted in Figure 5*a*) could be described in terms of macro- and micro-dissociation constants [43]. The latter (K) represent an empirical concept and are straightforward measured by conventional experimental techniques; they refer to stoichiometric protonation of the analyte and are, therefore, structure-unrelated.

$$L + H^+ \rightleftharpoons HL \qquad K_1 = \frac{[HL]}{[L][H^+]} \qquad (21.1)$$

$$HL + H^+ \rightleftharpoons H_2L \qquad K_2 = \frac{[H_2L]}{[HL][H^+]} \qquad (21.2)$$

Equation 21

On the other hand, micro-constants (k) are a more theoretical concept and account for protonation of individual forms. Coexistence of multiple sites

could generate (anti)cooperative effects that hinder attribution of such constants in a group-specific way.

$$k^{A} = \frac{[b]}{[a][H^{+}]} \qquad (22.1)$$

$$k^{B} = \frac{[c]}{[a][H^{+}]} \qquad (22.2)$$

$$k_{B}^{A} = \frac{[d]}{[c][H^{+}]} \qquad (22.3)$$

$$k_{A}^{B} = \frac{[d]}{[b][H^{+}]} \qquad (22.4)$$

Equation 22

A special declination of multiprotic drugs is represented by ampholytes, which display both acidic and basic moieties. The simplest case is represented by those which show only an acidic and a basic site.

From a stoichiometric (or macroscopic) perspective, the two neutral macro-species, the charged and uncharged ones, are considered collectively as tautomers of a single neutral form and the ratio among them is parameterized by the so-called K^{Z}.

Each macro-constant is defined by the contribution of two micro-constants, as reported in Adam's equations (see Equation 23 [6] and Figure 6):

$$K_{a}^{acidic} = K_{a}^{A^{Z}} + K_{a}^{B^{N}} \qquad (23.1)$$

$$\frac{1}{K_{a}^{basic}} = \frac{1}{K_{a}^{B^{Z}}} + \frac{1}{K_{a}^{A^{N}}} \qquad (23.2)$$

$$K^{Z} = \frac{K_{a}^{A^{Z}}}{K_{a}^{B^{N}}} = \frac{K_{a}^{A^{N}}}{K_{a}^{B^{Z}}} \qquad (23.3)$$

Equation 23

however, there is no direct relation between macro-constants and tautomeric constant. Hence complete micro-speciation could be obtained only if at least one micro-constant and K^{Z} are determined in addition to macro-constants.

Complexity of ionization equilibria generated in solution increases exponentially with the number of ionisable sites (see Figure 5*b*): for a n-protic molecule, 2^n micro-species and $n^{2(n-1)}$ micro-constants could be defined.
While potentiometry is generally applied for investigating macro-speciation, several experimental strategies have been developed to monitor micro-speciation.

3d Determination of micro- and macro-constants by spectrophotometric approach

Amphoteric drugs that carry two ionisable sites of opposite nature are conventionally divided in two classes according to relative acidity of those groups. Ordinary ampholytes cannot ionize simultaneously on the two moieties since pKa of acidic group is higher than the one of the basic one (pKa ^{acid} > pKa ^{base}); on the contrary, zwitterionic ampholytes (also named zwitterions) are able to form internal salts (pKa ^{acid} < pKa ^{base}) [6].

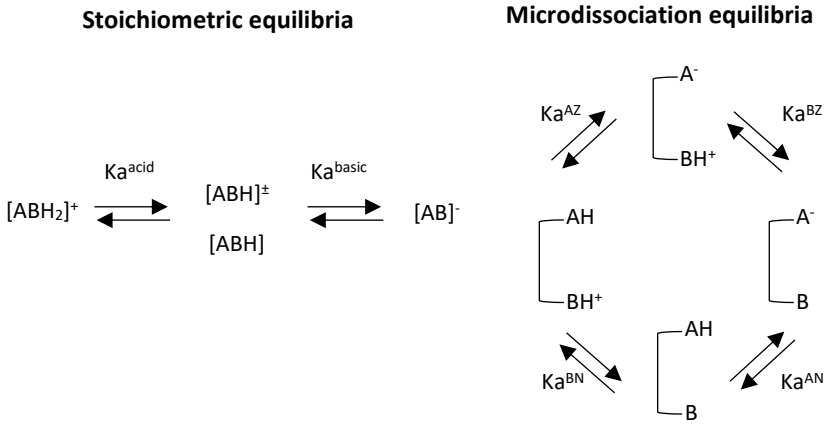


Figure 6: Dissociation equilibria of an ampholyte. Scheme taken from Pagliara's paper [6].

Spectophotometric approach is suitable for selectively monitoring one ionisable group if variation of absorbance profile could be ascribed almost totally to the ionization of that single centre. That requirement should be verified by spectral analysis, PCA and subsequent TFA calculation (see before).

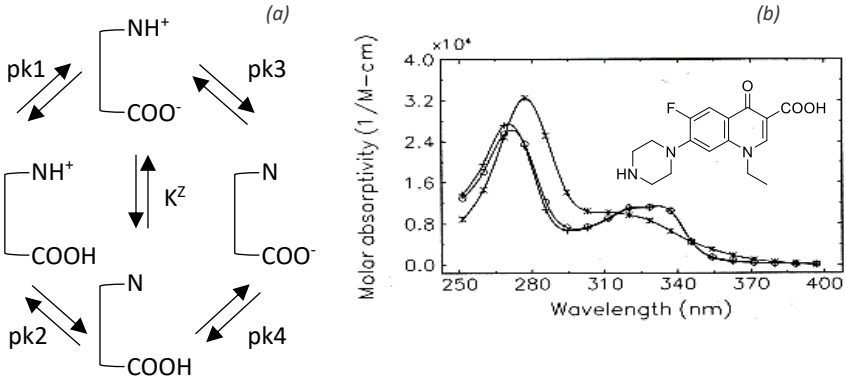


*Figure 7: (a) Ionization and (b) molar absorptivity profiles of norfloxacin, whose structure is reported. * H_2X^+, + HX (HX^0 and $HX^{\pm}$), o X^- [44].*

The neutral ampholyte norfloxacin proved the potential of integrating potentiometric and spectrophotometric approaches to profile micro-speciation of a multiprotic drug, as reported in Takacs-Novak and Kin's paper [44].

That fluoroquinolone displays the acidic site (*i.e.* a carboxylic acid) close to a UV-absorbing conjugated system, while the basic moiety (*i.e.* piperazine) is far enough from it not to affect absorbance profile in the analytical range. PCA on absorption matrix in the region 250-400 nm confirmed three independent light absorbing species to be involved in ionization equilibrium (*i.e.* the anion X^-, the cation H_2X^+ and the overall neutral form resulting from the contribution of the two tautomers HX^0 and $HX^{\pm}$). Absorptive coefficients were calculated from TFA and returned overlapped profiles for the monoprotic (XH) and anionic species (X^-) in the wavelength range 320-350 nm, proving the piperazinic moiety to be distant enough from the chromophore to not affect absorbance spectrum.

By monitoring absorbance changes in such a region, it was feasible to assign micro-constant to the carboxylic site (pk_1 = 6.28 ± 0.01). Macro-constants were then determined by potentiometric approach (pKa_1 = 6.25 ± 0.01; pKa_2 = 8.50 ± 0.03) and the results were combined to return micro-constants and the logarithm of the tautomeric constant (log K^Z, equal to 1.15).

Unfortunately, closeness of ionisable sites usually hampers the selective monitoring of one of them and alternative strategies should be considered. These methods take advantage from a simplification of the ionization profile by using close derivatives of the parent molecule which contain a reduced number of ionisable sites. Some of them apply semi-empirical corrective factors to reduce error in approximation.

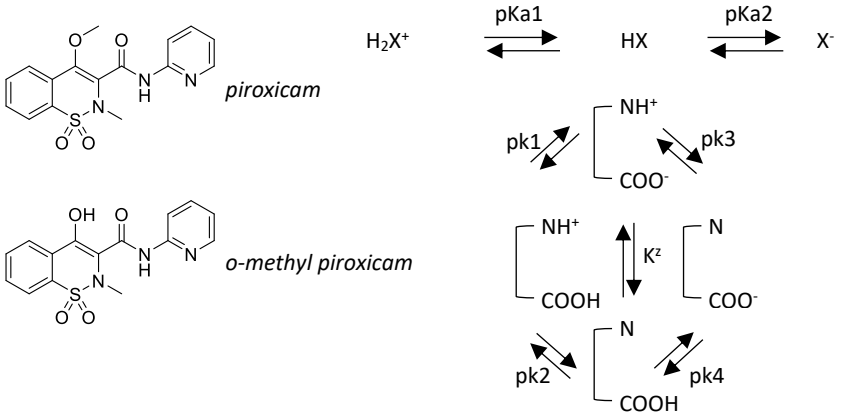


Figure 8: Piroxicam, O-methyl piroxicam and ionisation equilibrium of piroxicam [43].

Another option is provided by deductive approach that involves an ancillary compound in which one or more protonation sites of the parent molecule has/have been masked by chemical derivatisation. Such derivative can be used to assess at least one ionization micro-constant, assuming that: (i) masking functional groups would leave the acid-base properties of the others unaltered and (ii) sites are spaced enough to not generate intramolecular interactions due to conformational changes, a condition usually called “non-cooperativity”.

If assumptions are verified, macro-constant(s) of the derivative could be approximated as equal to the micro-constant(s) of the parent drug. This approach could involve, again, the combination of pH- and UV-metric methods that proves particularly useful for poorly soluble compounds, as piroxicam (see Figure 8). In this case, pyridinic pKa of ether analogue O-methyl piroxicam was measured spectrophotometrically and assumed equal to pk2 of the parent compound. Another example was provided by Pagliara and co-workers, who characterised macro- and micro-speciation of cetirizine by only using potentiometric method [7].

3e Determination of micro- and macro-constants by K^Z method

If both ionisable groups are linked to the chromophore (e.g. niflumic acid and pyridoxine in Figure 9), ionisation of each of those will result in a measurable UV-shift of the spectrum, hindering selective monitoring.
As a matter of fact, it is not possible to solve the spectral data from the zwitterion and the neutral species
because their concentrations are linearly dependent on each other (*i.e.* sum of those concentrations results in the one of the overall neutral macro-species HX).
In PCA, the spectral contributions from $HX^{\pm}$ and HX^0 degenerate to one principal component thus macro- instead of micro-constants are obtained in the TFA calculations. Taking advantage from this evidence, micro-constants can be determined through monitoring tautomeric equilibrium parameterized by K^Z [44].
From experimental and mathematical perspective, K^Z method recalls the abovementioned UV-metric and co-solvent approaches. Titration of the ampholyte drug is carried on in different mixtures of co-solvent (typically methanol) and water and absorbance spectrum is collected at any co-solvent ratio finally leading to K^Z extrapolation at 0% co-solvent.
Addition of an organic solvent to the aqueous medium decreases polarity of the environment which, in turn, induces a shift of tautomeric equilibrium towards the neutral rather than the zwitterionic species. Changes in the

relative concentration of the tautomers involve spectral changes that are monitored and deconvoluted by TFA calculation, assuming the spectrum of the neutral macro-species (HX) to be a linear combination of those of the overall neutral and the zwitterionic micro-species (HX^0 and $HX^{\pm}$ respectively).
Log K^Z is expressed as a linear function of the percentage of co-solvent in the mixture (wt.% co-solvent named R in Equation 24) and the intercept (S) is regarded as the tautomeric constant in purely aqueous condition (log K^Z (0%) obtained for R = 0).

$$\text{Log } K^Z (\%) = W \cdot R + S$$

Equation 24

Optical data are cast in a matrix:

$$a = E \cdot c$$

Equation 25

where *a* includes molar absorptivity spectra of the neutral macro-species (HX) collected at different co-solvent percentages and E and c represent the molar absorptivity spectra and the concentration-co-solvent content of the micro-species (HX^0 and $HX^{\pm}$) respectively.
First, it is proposed a test matrix c_t that contains the theoretical concentration-co-solvent content profiles of the micro-species:

$$c_t = \begin{bmatrix} c(HX^0) \\ c(HX^{\pm}) \end{bmatrix}$$

Equation 26

with:

$$c(HX^0) + c(HX^{\pm}) = 1$$

Equation 27

Concentration of the micro-species can be expressed as a function of co-solvent content:

$$c(HX^0) = \frac{1}{1 + 10^{(WR + S)}}$$

$$c(HX^{\pm}) = \frac{10^{(WR+S)}}{1 + 10^{(WR + S)}}$$

Equation 28

Then, approximated molar absorptivity matrices for HX^0 and $HX^{\pm}$ ($\bar{E}$) as well as for the neural species HX ($\bar{a}$) are calculated as follows:

$$\bar{E} = a \cdot c_t^T \cdot (c_t \cdot c_t^T)^{-1}$$

Equation 29

$$\bar{a} = \bar{E} \cdot c_t$$

Equation 30

An error function is adopted to evaluate the goodness of fit resulting from each step of the iterative optimisation:

$$err = (a - \bar{a})^2 \cdot \left(1 + \frac{\sum\sum|\bar{E}<0|}{\sum\sum|\bar{E}|}\right)$$

Equation 31

c_t depends only on W, S and R (wt.% methanol). That computation renders to an optimization of W and S for a minimum err value. Macro- and tautomeric constants are then used to calculate micro-pKa.
Relative abundance of the tautomers of an amphoteric drug could be qualitatively speculated from Yasuda-Shedlovsky extrapolation obtained from several experiments in co-solvent-water mixtures.

Slope of such curves highlights the impact of medium composition on the basic/acidic strength of the compound. If tautomeric equilibrium is shifted towards the zwitterionic species (e.g. ciprofloxacin; log K^Z > 1, *i.e.* ratio $[HX^{\pm}]/[HX^0]$ > 10), polarity of the medium significantly affects pKas values due to the polar nature of the analyte. On the contrary, if neutral form is predominant over the zwitterionic one (e.g. sulfathiazole), medium effect is less remarkable.

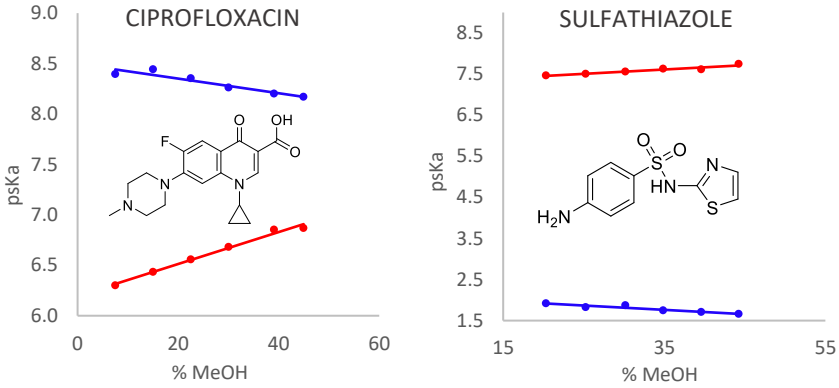


Figure 9: psKa versus methanol % plot for ciprofloxacin (zwitterionic ampholyte) and sulfathiazole (neutral ampholyte).

3f Dissociation profile of amphoteric drugs: pharmacological implications

Multiprotic compounds have been widely developed in pharmaceutical field: anti-diabetes candidates (e.g. salacinol and kotalanol) and their structural-modified derivatives and anti-inflammatory drugs (e.g. azapropazone) often show a zwitterionic nature that is crucial for their interaction with the biological target and the consequent pharmacological effect [45]. Zwitterions are widely represented in the antibiotic class (e.g. β-lactams, quinolones) as well.

For instance, levofloxacin (LVFX) and grepafloxacin (GPFX) harbour a carboxyle and a piperazine moieties (pKa 5.7 and pKa 7.9 respectively for LVFX; pKa 7.1 and pKa 8.8 respectively for GPFX) and inhibit the uptake of L-carnitine in the intestinal tract. The mechanism of action of these fluoroquinolones was proved to be highly dependent on the zwitterionic nature of the drugs.

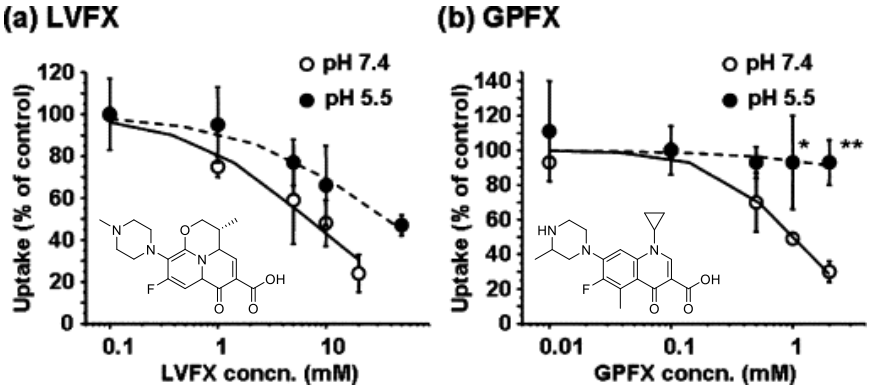


*Figure 10: (a) Inhibitory effects of various concentrations of LVFX (0.1-50 mM) and (b) GPFX (0.01-2 mM) on L-[^{3}H]-carnitine (5 nM) uptake at pH 5.5 or 7.4 was measured for *Significantly different from pH 7.4 at p<0.05, **p<0.01 [46].*

L-carnitine uptake mediated by Organic Cationic Transporter (OCTN2) was studied in permeation assays performed on Caco-2 cells in presence of antibiotic levofloxacin and grepafloxacin (see Figure 10). Hirano and co-authors [46] evaluated the impact of pH testing 5.5 and 7.4 conditions. Even if they did not unveil the mechanism of drug-transporter interaction, their results suggested the fluoroquinolones to be more active when the ratio of zwitterionic-over-neutral species was higher: for instance, LVFX IC_{50} decreased from 36.1 mM at pH 5.5, where the calculated abundance of zwitterion was 39%, to 6.8 mM at pH 7.4, where it was calculated the percentage of double charged species to increase up to 75%.

	pH	Ratio of % zwitterion	IC50 (mM)
VLFX	5.5	39	36.1

	7.4	75	6.8
GPFX	5.5	2	–
	7.4	64	1.0

Table 3: Ratio of zwitterionic form and IC50 value of quinolones on L-carnitine uptake [46].

Hirano's group speculated charge distribution over the molecule to be crucial, together with chemical structure, for recognition of the drug by the target and the consequent inhibition.

Since the amphoteric nature of a drug compound has a major impact on its biological behaviour, that relation could be ideally rationalized to allow predictions about newly synthesized compound without performing experimental assays. Added to that, these information could be parameterized by meaning of descriptors and employed in virtual designing of NCEs with target properties [47].

In order to understand structure-biological activity relationship for NCEs, and in turn obtain predictive models, computational system should be trained with high quality data. Thus determination of micro- and tautomeric constants proves relevant to gain site-specific and accurate information for implementation of expert systems.

1.3 Relevant physico-chemical properties: lipophilicity

1.3.1 Lipophilicity parameters: log *P* and log *D*

Lipophilicity (or hydrophobicity) screening is regarded of utmost importance in early discovery, especially since pharmaceutical industry has shown a tendency towards "drug-obesity".
Hydrophobicity is parameterized by the Log value of the partition coefficient (Log *P*, see Equation 32) for neutral compounds or the distribution coefficient measured at a given pH (Log D_{pH}, see Equation 33 and Equation 34) for ionisable compounds in a biphasic system made of an organic and an aqueous phase.
For a monoprotic chemical entity X, log *P* is expressed as:

$$\text{Log } P = \log \frac{[X]_{organic}}{[X]_{water}}$$

Equation 32

Log *D* of a monoprotic acid is:

$$\text{Log } D_{pH} = \log \frac{([X^-] + [HX])_{\ organic}}{([X^-] + [HX])_{\ water}}$$

Equation 33

while log *D* of a monoprotic base is:

$$\text{Log } D_{pH} = \log \frac{([X] + [HX^+])_{\ organic}}{([X] + [HX^+])_{\ water}}$$

Equation 34

Furthermore, for monoprotic acids and bases, log *D* can be estimated from pH-independent log *P* and pK_a, assuming the neutral form to be predominant in partitioning:

Log *D* ≅ log *P* – log $(1 + 10^{(pH - pKa)})$ *(35.1)* for a monoprotic acid

$$\text{Log } D \cong \log P - \log (1 + 10^{(pKa - pH)}) \quad (35.2)$$ for a monoprotic base

Equation 35

The difference between the two indexes is determined by the fraction of ionized compounds at a given pH: hence for an unionizable compound log *D* equals to log *P* [48].

Considering a monoprotic base at pH ≤ pKa-3, log *D* can be approximated to the log *P* of the cationic form; conversely, at pH ≥ pKa+3, log *D* can be approximated to the log *P* of the neutral form. The same speculation can be made for a weak acid: at low pH lipophilicity is controlled by the neutral species while at alkaline pH it is driven by the anionic species (see Figure 11).

While partition coefficient only considers one form, and typically the neutral one, distribution coefficient is more commonly used to parametrize lipophilicity of ionisable compounds, especially at physiological pH 7.4. Multiprotic molecules generate complex partition equilibria that are described by more complex equations.

For ionisable compounds, using log *D* over log *P* was proved to remarkably impact on selection of drug candidates on the basis of rule of thumb such as Lipinski's Rule of 5 [49].

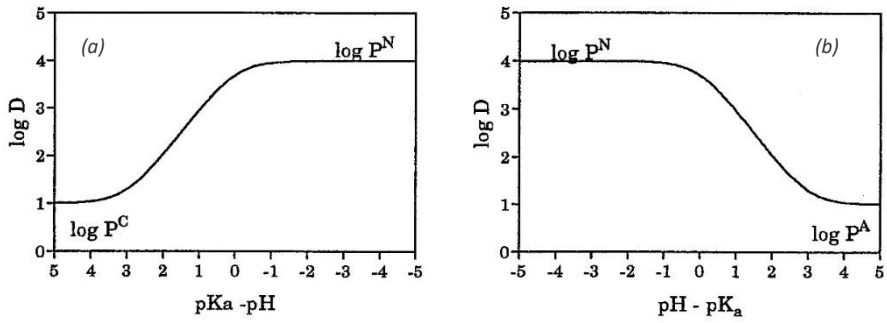


Figure 11: Lipophilicity profiled as a function of (pKa-pH) for (a) a weak monoprotic base and (b) a weak monoprotic acid [7].

1.3.2 Thermodynamic of partition mechanism

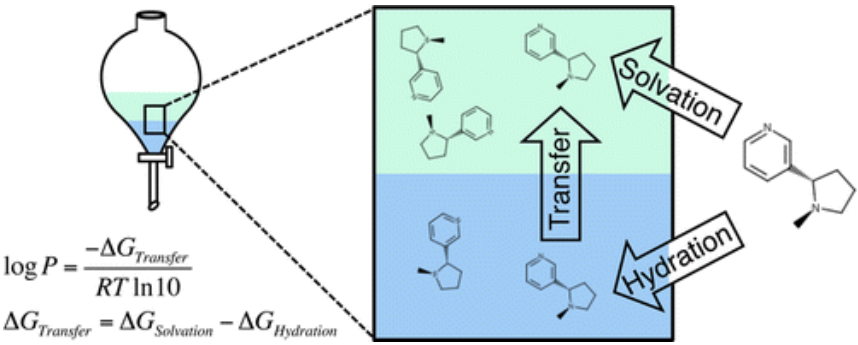


Figure 12: Scheme of partition equilibrium and free energy related to solute transfer between the phases of the biphasic system [50].

Lipophilicity relates to both *in vitro* and *in vivo* phenomena and should be carefully modulated to reach the maximum rate of success for a drug candidate.

Molecular properties as molecular weight (or size), hydrogen-bond donor and acceptor capacity, polarity and polarizability affect the overall hydrophobicity of a drug molecule, which results from intermolecular forces between solute molecules as well as between solute and the two-phase system in which partition occurs.

The process has been extensively investigated from a thermodynamic perspective during the XX century. Such approach considers molecular transfer from the compound bulk to each solvent of the partition system and, finally, adds up the factors relative to both phases [51].

As any thermodynamic equilibrium, free energy (ΔG) related to partition accounts for an enthalpic (ΔH) and an entropic (ΔS) contribution. For instance, transfer of hydrocarbons in water has been studied at 298 K showing a negative enthalpic contribution for the system (*i.e.* a favourable enthalpic change) by the evolution of heat. Hydrocarbons, however, are notoriously insoluble in water: this reluctance to mix in that polar solvent must be ascribed to a large and predominant entropic contribution to the process.

A small enthalpic benefit results from the large energy expense required for re-ordering both the hydrocarbon solute and the water solvent molecules when placed together, thus separation of the two phases is promoted. The same phenomenon can be considered regulating the distribution of apolar molecules in an apolar solvent-water system.

	T (K)	ΔS_u	ΔH	ΔG^0_u
propane C_3H_6 (l) → C_3H_6 in H_2O	298	-23	-1800	+5050
butane C_4H_8 (l) → C_4H_8 in H_2O	298	-23	-1000	+5850
benzene C_6H_6 (l) → C_6H_6 in H_2O	291	-14	0	+4070
toluene C_7H_8 (l) → C_7H_8 in H_2O	291	-16	0	+4650
ethylbenzene C_8H_{19} (l) → C_8H_{19} in H_2O	291	-19	0	+5500

Table 4: Thermodynamic changes in hydrocarbon transfer. Data are mutuated from Leo and Hansch's paper. ΔS_u is unitary entropy, ΔG^0_u is reported in kcal/mol [51].

On the other hand, the origin of the large negative entropic change involved in partitioning between aqueous and non-aqueous phases could be ascribed to arrangement of water molecules around apolar portion of the organic compound. Water polar molecules form unstable, flickering clusters that induce high randomization of the aqueous phase and hence in high entropic factor.

Hydrogen bond has paramount importance in determining the characteristics of both solute and organic solvent phase. As regards the analyte, the number of hydrogen bond acceptors and donors showed by the polar portion accounts for its hydrophilicity. When alcohols, esthers and ketons are used as organic phase, hydrogen bonding between compound and solvent determines an enthalpic benefit that drives the overall process.

It must be kept in mind that rather large amounts of water are present in those oxygen-containing solvents when saturated during the partitioning process: for instance, *n*-octanol dissolves in water only to the extent of 4.5

mM, but the molar concentration of water in octanol is 2.30 M. Thus it is likely that polar functions of the solute would be more or less solvated by both water molecules and/or the hydroxyl function of the alcohol.
Dispersion forces could ideally impact on the position of partition equilibrium, but it is more likable their contribution to be largely cancelled out, when a solute molecule leaves one phase and enters the other one.
The energy required for transfer from the aqueous phase to the organic phase for any charged or ionized solute heavily depends on the dielectric constant of the organic phase in question. Most of the water-immiscible organic solvents have lower dielectric constants in comparison to water, and thus charged solutes must contain rather large hydrocarbon residues to be prone to access the organic solvent, unless other phenomena occur (e.g. ion pairing, see in the following paragraph) [52].
To unveil, quantify and rationalize several interacting solvents effects on diverse physico-chemical properties of the solute (*i.e.* partitioning, solubility, blood-brain distribution and human intestinal absorption) Kamlet and Abraham developed the so called solvatochromic equation or linear solvation energy relationship:

$$XYZ = XYZ0 + mV/100 + s(\pi^* + d\delta) + a\alpha + b\beta$$

Equation 36

in which the solute property of interest (XYZ) is described in terms of five solute descriptors that have been defined by averaging multiple normalized solvent effects on a variety of properties involving many diverse types of indicators [53].
(i) A cavity term (mV/100) measures the endoergic process of locating the solute in a suitably sized cavity that is generated by separation of solvent molecules. Focusing on partitioning model, that term is measured as solute liquid molar volume [54].
(ii) π* parameterizes solvent dipolarity/polarizability, that is the ability of the solvent to stabilize a charge or a dipole exploiting its dielectric effect. Values of π* for non-chlorinated non-protonic aliphatic solvents with a

single dominant bond dipole have been shown to be generally proportional to molecular dipole moments.
(iii) δ is a polarizability correction term equal to 0.0 for non-chlorinated aliphatic, 0.5 for poly-chlorinated aliphatic, and 1.0 for aromatic solvents.
The values accounts for the fact that differences in solvent polarizability are significantly greater between these classes of solvents than within the others.
(iv) α scale of solvent HBD (hydrogen-bond donor) acidity defines propensity of the solvent to donate a proton in a hydrogen bond with a solute molecule.
(v) Conversely, ß scale of HBA (hydrogen-bond acceptor) basicity parameterizes solvent ability to accept a proton (*i.e.* propensity to donate an electron pair) in a solute-to-solvent hydrogen bond.
The m, s, d, *a*, b, coefficients in Abraham's equation express the relative susceptibilities of the physico-chemical property XYZ of the solute to the indicated solvent property scales. The five descriptors have been roughly normalized to cover a range from 0.0 to 1.0 hence ratios between different coefficients should provide straightforward measures of the relative contributions of the indicated solvent properties.

1.3.3 Relevance of lipophilicity in discovery and early development

Estimation of hydrophobicity is recommended at early discovery stage for fragments and hits to guide design towards minimization of attrition in clinical studies. In this context the triad potency, lipophilicity and size is crucial to sort further modification in the molecular structure of potentially active compounds.
Lipophilicity assessment returns potentially useful information for a better understanding of solid state characteristics of drug candidates [55]. Lipophilic compounds with rigid, highly aromatic structures and large flat exposed areas are characterised by high melting points and enthalpy of fusion. These factors account for high stability of the crystal lattice, which impacts on solvation and, in turn, solubility. Therefore, the knowledge of the lipophilicity of a NCE gives highlights to pre-formulation scientists on

the most suitable strategy to optimize drug *in vivo* administration. Moreover, the need for formulating highly lipophilic drugs opened new frontiers in drug-delivery area (e.g. Self Emulsifying Drug Delivery Systems, SEDDS [56] in all their declinations).

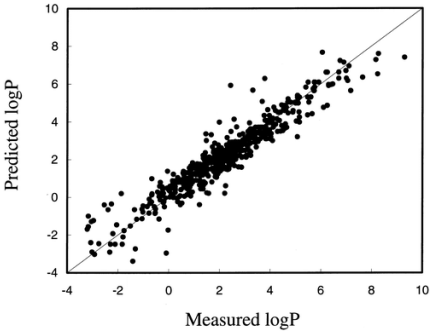
As regards the correlation to the physico-chemical properties, the close relation between lipophilicity and solubility has been extensively discussed in the literature, leading to the general rule that the higher the lipophilicity the lower the solubility. This condition could negatively affect absorption of those compounds for which solubilisation is the limiting step (e.g. oral drugs belonging to class II of the Biopharmaceutics Classification System, BCS [57]).

In addition, hydrophobicity guides passive diffusion through biological barriers, as the intestinal and brain blood ones, that influences distribution, metabolism, pharmacokinetics (PK) and toxicity, since highly lipophilic drugs often show off-target effects through unspecific and promiscuous binding [58].

1.3.4 Lipophilicity Assessment

1.3.4.1 Computational approaches

As reported for dissociation constants, several computational methods have been developed to predict lipophilicity of drug candidates in order to save time and substance usually required for experimental assays.



$\text{Log } P_{\text{predicted}} = 1.00 \text{ Log } P_{\text{measured}} + 0.0$

$R^2 = 0.893$ s = 0.653

F = 1631 n = 592

Figure 13: Correlation between measured and calculated log P for a training set of 592 molecules in Xing and Glen's [63].

Table 5: Equation of regression model and associated statistical parameters as reported in the paper [63].

Since 1980s many models have been developed to correlate log *P* in *n*-octanol/ water system to solubility [59, 60, 61] and all of those are affected by the quality of equations and solubility data.

Added to that, since lipophilicity plays a key role in ligand-protein mechanism of recognition, an early investigation of the relation between hydrophobicity and target binding is recommended to optimize molecular features and avoid target promiscuity.

Different parameters have been proposed for such purpose, like the Ligand Lipophilicity Efficiency (LLE) and its latest modification (LLE_{AT}) introduced by Mortensen and Murray [62]. Xing and Glen [63] calculated log *P* using three molecular descriptors (*i.e.* polarizability and partial atomic charge of nitrogen and oxygen) finding good agreement between results and log *P* values from the training set (see Figure 13 and Table 5 below).

All these indexes are based on the free energy change associated with the binding, to which a negative contribution is given by transferring drug from aqueous environment to the locally hydrophobic pocket offered by the target protein.

LLE and derived parameters are useful to determine suitable potency and log *P* for fragments and allow ranking of efficiency within homologous that are decorated with different substituents.

The main drawback of these computational approaches is the fact that ionization is ignored since pKas and consequent log *D* predictions are affected by excessive uncertainty to return reliable results. In that sense, the need for experimental log *D* values is still undeniable.

Added to that, several software packages that calculate partition coefficient are available and among them ClogP is the most widespread in drug discovery. ClogP (Pomona College) is a substructure-based design software that assumes additivity of fragmental contributions. The analyte is broken down into fragments and the log *P* value of each of those is computed; a correction factor (F_j in Equation 37) that accounts for interaction within functional groups is included:

$$\mathrm{Log}\, P = \sum_{i=1}^{n} a_i \cdot f_i + \sum_{j=1}^{m} b_j \cdot F_j$$

Equation 37

where f_i is the fragment constant, F_j is the interaction factor, a_i is the occurrence of the fragment and b_j is frequency of the interaction. ClogP was calibrated by using Starlist that is a high-quality dataset including over 10^4 diverse compounds for which log *P* was experimentally measured. The additive-constitutive approach makes the calculator comprehensible and suitable for implementation that means is should be regarded as a rule-based empirical model. However, since fragment constants and correction factors in ClogP are largely derived from 600 small-sized molecules, it lacks of generality: despite its great predictive performance on Starlist compounds, goodness of the output was significantly decreased when applied on different database (e.g. Pfizer's set of 95,809 compounds and Nycomed's set of 882 compounds).

Diverse log *P* calculators have been developed, including fragmental methods (ACD/log *P*, KLOGP), atom-based methods (AlogP, XLOGP), and a variety of property-based methods. These models were trained on datasets composed of around 200,000 of compounds on the basis of a wide spectrum of molecular descriptors, ranging from molecular dynamics simulations to quantum mechanical parameters [48].

Even if accurate for neutral molecules, calculators return high error in prediction associated with multiprotic molecules displaying interactions between functional groups. Moreover, they often fail in prediction for complex compounds whose fragments are not included in calibrating set [64]. Machatha and Yalkowsky [65] compared different types of software for log *P* prediction based on group contribution. They evaluate accuracy in prediction by mean of the average absolute error (AAE):

$$\mathrm{AAE} = \frac{\sum \lvert \mathrm{observed} - \mathrm{predicted} \rvert}{n}$$

Equation 38

where *n* is the number of compounds considered. To evaluate difference between two data-sets *t*-tests were performed on the logarithmic data and *P*-value was determined setting significance level at 0.05 (*i.e.* if *P*-value is < 0.05 than the two data sets are considered to be significantly different).

Group	n of compounds	AAE	
		ACD/logP	ClogP
all compounds	103	0.327	0.265
without tautomers and zwitterions	84	0.27	0.248
tautomers and zwitterions	19	0.58	0.31

Table 6: Average absolute errors (AAE) and P-values without the five outliers out of the 108 compounds initially considered [65].

Testing the models on a dataset of 108 drugs they found out computed lipophilicity of zwitterionic, tautomeric and strongly hydrogen-bonding compounds poorly correlated with experimental values (see Table 6 and Figure 14).

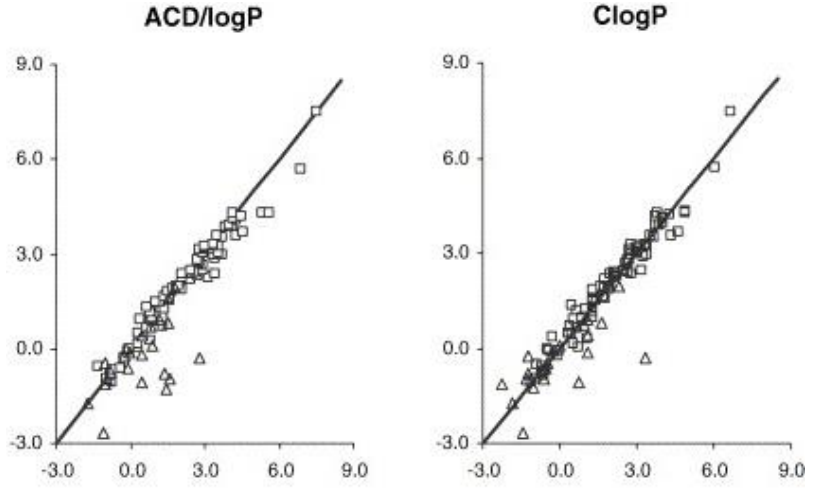


Figure 14: Plots of experimental against predicted log P values for ACD/logP, ClogP. Squares are non-tautomers and non-zwitterions; triangles represent tautomeric or zwitterionic compounds [65].

The need for improving predictivity has been leading to update calibrating libraries since year 2000. Schroeter and co-workers [66] developed a predictive tool based on over 14,000 drugs from Bayer Schering Pharma; the resulting model proved more effective than the four commercial software packages taken as reference.

Moreover, developing XLOGP3 software, Cheng and co-authors [67] allowed including new experimental lipophilicity values in the dataset to re-train the software. That strategy maximizes the mutually advantageous relation between computational and experimental method: the former gives hints about experimental set-up; the latter improves predictions providing concrete and reliable data.

Shroeter's and Cheng's studies [66] proved the need to improve dataset of any commercially available software, usually based on small neutral molecules and fragments that are far from the complexity of actual discovery molecules.

To meet this need of pharmaceutical industry high quality experimental data are still needed.

Name	Company (Website)	Method/comments/reference
ACD/log *D*	Advanced Chemistry Development Inc. (www.acdlabs.com)	Fragment-based
ADMET Predictor™	SimulationPlus Inc. (www.simulations-plus.com)	Neural network
AlogP	Virtual Computation Chemistry Laboratory (www.vcclab.org)	Neural network [68, 69] ; displays results of other programs for comparison.
Bio-Loom	Bio-Byte Corp. (www.biobyte.com)	Fragment-based; replaces ClogP
Chemaxon	ChemAxon Ltd. (www.chemaxon.com)	Atom-additive method with correction factors [70]; ionic strength effects

HyperChem 7.0.	Hypercube Inc. (www.hyper.com)	Atom-additive method [71]
KowWin 1.67 (EPI SuiteTMpackage)	US EPA/ Syracuse Research Corp. (www.epa.gov)	Fragment-based
MlogP (DRAGON 3.0 package)	Talete srl. (www.talete.mi.it)	Counting atoms, bonds, fragments or functional groups [72, 73]
MOLPRO package	ChemDBsoft (www.chemdbsoft.com)	Fragment-based
PrologD (Pallas package)	CompuDrugChemistry Ltd. (www.compudrug.com)	Three algorithms: 2 linear and 1 neural network [74]
SPARC	University of Georgia (http://ibmlc2.chem.uga.edu/sparc/)	[75]; ionic strength effects
CSlogDTM	Chemsilico LLC. (www.chemsilico.com)	Neural networks [76]
XlogP	Virtual Computation Chemistry Laboratory (www.vcclab.org)	Atom-additive method with correction factors [77]

Table 7: Several software for calculation of log P; table from Kah and Brown's paper [78].

1.3.4.2 Experimental assessment of lipophilicity

Experimental methods for lipophilicity determination are classified in direct and indirect approaches according to the parameter they allow to quantify. The former permit a direct measure of partition or distribution coefficients and include the shake-flask approach and potentiometric method that are still the most widely used in pharmaceutical industry. Moreover, shake-flask approach is recommended by the Organization for Economic Cooperation and Development (OECD) [79].

Indirect methods determine a hydrophobicity parameter that can be related to log *P* or log *D* through specific equations. Although description of indirect methods is beyond the purpose of this thesis, some references

to those will be provided focusing in particular on chromatographic approaches.

2a RP-HPLC methods

The general purpose of indirect methods is to return information about lipophilicity in a quicker way in comparison to the direct ones. Most of them is based on separation techniques, such as reversed-phase high-performance liquid chromatography (RP-HPLC), counter-current chromatography (CCC) [80], immobilized artificial membrane (IAM) chromatography [81], reversed-phase thin-layer chromatography (RP-TLC), immobilized liposome chromatography (ILC), micellar LC (MLC) [82] and bio-partitioning micellar chromatography (BMC) [83].
For reproducibility, minimization of liquid handling and on-line detection, RP-HPLC has gained increasing attention so that it has been included in the OECD accepted approaches.
Chromatographic methods rely on the assumption that retention and the *n*-octanol/water partitioning could be regarded as homoenergetic processes characterised by the same free-energy changes.
Direct information returned by chromatographic analysis consists in the retention time (RT) and it is evident how the ease of the method has made it so attractive for pharmaceutical screening.
The theoretical base of RP-HPLC methods is provided by Collander Equations [79] (see Equation 39 and Equation 40), in which the logarithm of retention factor (log *k*) and its relation to log *P* are described.

$$\text{Log } k = \log ((t_R - t_0)/t_0)$$

Equation 39

Equation 39 defines retention factor as a function of retention time of the analyte (t_R) referred to retention time of an un-retained compound, such as a solvent, e.g. methanol, or a salt, e.g. sodium nitrate (t_0).
Elution conditions impact on the measured t_R, thus general ranking of retention factors for diverse compounds analysed by using different mobile

phases is allowed only by considering completely aqueous eluting medium (log k_W). For such purpose Snyder-Soczewinski Equation (Equation 40) can be applied:

$$\text{Log } k_W = \log k - S\varphi$$

Equation 40

where φ represents the percentage of organic modifier in the mobile phase and S is a solute-dependent parameter derived from linear regression analysis.
Reaching totally aqueous conditions is hampered by the operative limit of reversed-phase column and quite long retention times that would be required for lipophilic analytes, hence log k_W was often extrapolated by at least 4 isocratic conditions, with good approximation when log k_W values are within 0 and 1. Finally, a linear regression could be found between log *P* and log k_W:

$$\text{Log } P = a\log k_W + b$$

Equation 41

Building of relative scales in the same chromatographic conditions [84], effect of the organic modifier and the stationary phase employed have been discussed in literature [85, 86].
To speed up experimental procedures gradient-based approaches have been proposed, such as Valko's Chromatographic Hydrophobicity Index (CHI) [87] and Krass's Kg [88].
It should be pointed out that Equation 41 only refers to a hypothetical neutral drug. For ionisable compounds, included ampholytes, retention factor results from the contribution of both unionized and ionic species of the molecule. In such case it is better to correlate log k_W to log *D* instead of log *P* and proper equations have been provided for this purpose [79].
For the ease of preparation, low-cost and wide feasibility, these techniques have turned out to be useful to estimate hydrophobicity of NCEs in early drug discovery screening. Effective calibration represents the main limit for

a reliable application of indirect methods in late drug discovery and pre-development phases, in which conventional in-depth methods such as shake-flask and potentiometry are generally preferred.

2b Shake-flask method

Despite the growing interest towards more biomimetic conditions (e.g. partition between artificial membranes/water, liposomes and chromatographic *ad hoc* methods) [89], organic solvent/aqueous buffer system is still the most convenient model to adopt in lipophilicity investigations.

For the sake of completeness, an amphiprotic, an inert, a hydrogen-bond donor and a hydrogen-bond acceptor isotropic systems should be assayed to cover the range of biophysical properties of biological membranes (e.g. *n*-octanol/water, alkanes/water, chloroform/water, and dibutyl ether/water). However, *n*-octanol is the most used solvent in lipophilicity assays of pharmaceutical compounds because of the uncountable published data obtained in that medium. The main advantages are represented by practical aspects, such as (i) the low vapour pressure (8.7 Pa at 20 °C); (ii) the low UV-absorbance within a wide range of wavelengths and (iii) its relative density if compared to water (0.83).

Moreover, the chemical properties of *n*-octanol make it suitable for lipophilicity studies: it is prone to H-bonding and shows a relatively high content of water at room temperature. That reduces structural variations observed for apolar solvents such as hexadecane that may promote molecular association and subsequent precipitation of the analyte [90].

Many techniques have been reported over the years to determine lipophilicity, among which conventional shake-flask method is still regarded as the gold-standard. It simply consists in the partition of the analyte (introduced as powder or stock solution) between the two phases of the system using a flask or a tube.

After partition, the phases are separated, properly diluted and quantified by using off-line UV-vis absorbance, high- (HPLC) or ultra-high- pressure chromatography (UHPLC) coupled to mass spectrometry (MS) [91]. The

dynamic range of the technique depends on the detection method, but normally covers on average a range of 4 log units. Albeit allowing the direct determination of log *D*/ log *P*, conventional shake-flask is tedious, time-consuming and substance demanding. To meet the needs of early screening and reduce amount of analyte, partition times and the need for manual liquid-handling, medium and high-throughput shake-flask assays have been developed.

The major innovation of these approaches is the use of automation for sample preparation and processing; in addition, the starting material is usually a DMSO stock solution of the drug, which is more available than the solid form at the stage of early discovery.

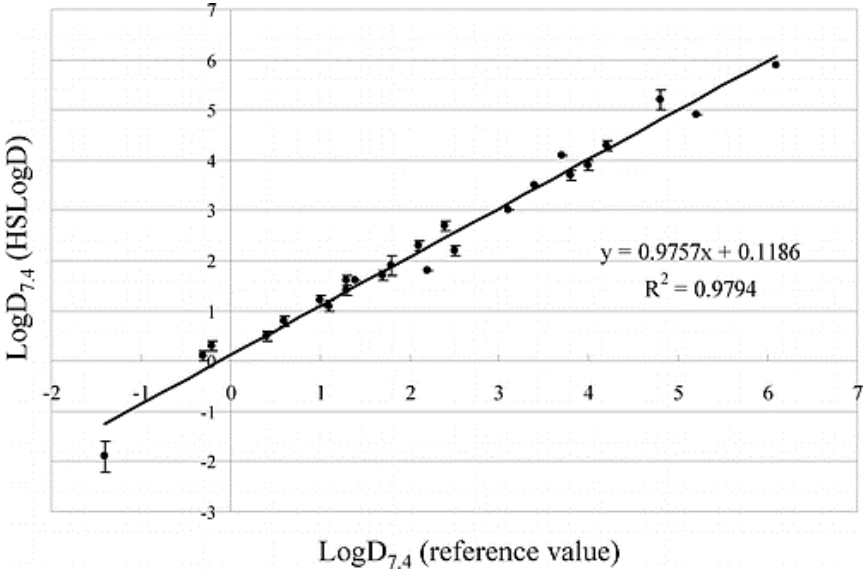


Figure 15: Correlation of High Speed LogD versus reference values in Dohta's paper [92]. The results for the 25 commercial drugs are reported as the means ± SD of triplicates.

Dohta and co-authors [92] reported an elegant approach in which miniaturized shake-flask is combined with LC-MS detection resulting in a high-throughput parameter (High Speed Log *D*, HSLogD at pH 7.4).

Authors focused on mutual contamination of *n*-octanol and aqueous phase that is the main problem encountered at the stage of separation by robotic tool. To overcome that issue, they set up a water-plug in system in which

the needle draws a minimum volume of water (water-plug) before lowering in the biphasic system to sample aqueous phase. The water in the needle was expected to repel *n*-octanol, preventing contamination. The method was successfully validated on 25 commercial drugs spanning between 0 and 5 log units, by correlating the HSlog *D* values with reference log *D* reported in the literature.

Another concern in miniaturized shake-flask protocols is the potential co-solvent action provided by DMSO when starting from stock solution of the drug, which could be avoided by evaporating the solvent under vacuum condition before adding *n*-octanol and aqueous buffer [93]. It is worthy noticing that a percentage of DMSO stock solution ≤ 1% v/v is generally regarded as safe to prevent biased drug solubility.

Finally, buffer system should be carefully chosen to avoid apparent increase in measured lipophilicity. Hydrophilic species of hydrophobic molecules (cation or anion) are able to generate less polar, more lipophilic ion-pairs with counter-ions in solution. Ion-pairs could be regarded as thermodynamically stable species characterised by a proper log *P* [94]. Relationships between the structures of both drugs and counter-ions have been investigated for quaternary ammonium salts and a series of acidic counter-ions [95]. The study reported flexibility and size to be the most relevant factors affecting ion-pairing propensity, together with the solvent accessible area displayed by the counter-ion. Large, rigid and hydrophobic counter-ions have low rate of probability to find a favourable orientation towards an ionized drug displaying the same characteristics. Added to that, dual charged counter-ions (such as maleate acid at pH 7.4) generate still charged species that are not prone to partition. Taking advantage from this evidence, biorelevant zwitterionic buffers (e.g. 3-(N-morpholino)propanesulfonic acid, MOPS) are preferred over anionic ones (e.g. phosphate buffer) in shake-flask assays. Because of the aromatic moiety and the dual-charge nature, they are supposed to reduce ion-pairing and consequent over-estimation of lipophilicity.

2c Potentiometric method

Another gold-standard for lipophilicity measurement is the potentiometric (also known as pH-metric) method developed by Avdeef during the 1990s [28, 31, 96, 97] and implemented in hardware-software package by Sirius Analytical, UK. The biphasic system of choice is *n*-octanol/ISA water 0.15 M KCl, in which the phases are mutually saturated. Using potentiometry, the method takes advantage from the effect of partition solvent on ionization equilibrium and is suitable for log *P* determination of the neutral form of ionisable compounds.

First, electrode calibration is performed as reported above (see **2b Calibration of the combined glass electrode** section) and dissociation constants are determined in aqueous or co-solvent/water mixtures.

Then titrations are repeated in different partition solvent/ISA water volume ratios, chosen taking into account the expected lipophilicity of the analyte (*i.e.* the higher the expected hydrophobicity, the lower the ratio). If Yasuda-Shedlovsky extrapolation was previously required, error in calculation is higher in comparison to the one obtained starting from purely aqueous pKa data.

The neutral rather than the ionized form tends to enter the *n*-octanol phase involving a shift of the flexes of the titration curves (apparent pKas in presence of octanol, p_oKas) in comparison to the ones recorded in purely aqueous conditions (pKas). That difference is visualized through Bjerrum plot: ionization equilibrium is weakened by partitioning and pKa shifts towards higher (acidic pKa) or lower (basic pKa) values according to the nature of the site.

For a monoprotic weak acid HA, the following relation between the partition coefficient P, p_oKa and pKa was assumed:

$$P_{HA} = \frac{10^{(p_oKa - pKa)} - 1}{r}$$

Equation 42

where r is the ratio of the volume of partitioning coefficient over the ISA water phase.
For a monoprotic weak base B, the corresponding equation is:

$$P_B = \frac{10^{(pKa - p_oKa)} - 1}{r}$$

Equation 43

If ion-partitioning were significant, partitioning of both forms of a monoprotic drug should be taken into account (XH and X). Considering pKa, apparent dissociation constant at two different phase ratios (p_oKa_1 and p_oKa_2 respectively) and such ratios (r_1 and r_2), three constants are calculated (*i.e.* pKa, log P_{XH} and log P_X) by using the following equations:

$$P_{XH} = \frac{r_2 10^{(p_oKa_2 - pKa)} - r_1 10^{(p_oKa_1 - pKa)} - (r_2 - r_1) 10^{(p_oKa_1 + p_oKa_2 - 2pKa)}}{r_1 r_2 (10^{(p_oKa_1 - pKa)} - 10^{(p_oKa_2 - pKa)})}$$

Equation 44

$$P_X = \frac{r_1 10^{(p_oKa_2 - pKa)} - r_2 10^{(p_oKa_1 - pKa)} + r_2 - r_1}{r_1 r_2 (10^{(p_oKa_1 - pKa)} - 10^{(p_oKa_2 - pKa)})}$$

Equation 45

Referring to a generic multiprotic compounds XH_n, the apparent dissociation constant for the *n*-proton (p_oKa_n), the pKa for the (n-1)-proton (pKa_{n-1}) and the partition coefficients (PXH_n and PXH_{n-1}) for the respective species are related as follows:

$$10^{(p_oKa_n - pKa_{(n-1)})} = \frac{rP_{XHn} - 1}{rP_{XH(n-1)} - 1}$$

Equation 46

Refinement of log *P* values of multiprotic analytes is provided by a non-linear least-squares approach based on the same principles described for pKa data fitting.

Bjerrum difference plots are analysed to approximate equilibrium constants that serve as “seed values” for iterative refinement of log *P*. Refinement is iterated until minimization of sum of weighed squares of residuals is reached (S in Equation 8) to obtain optimized values.

Potentiometric log *P* determination was validated by Avdeef et al. on 61 structurally diverse compounds (e.g. amino acids, peptides, ampholytes, barbiturates, ß-blockers, herbicides, phenols) by comparing pH-metric with shake-flask data [98, 99, 100].

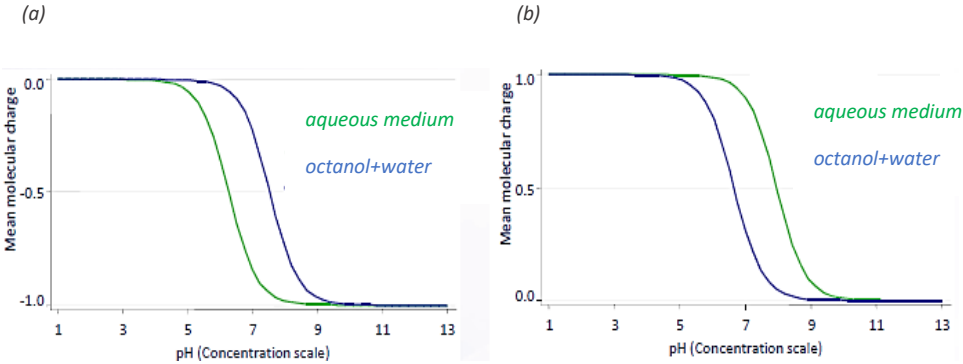
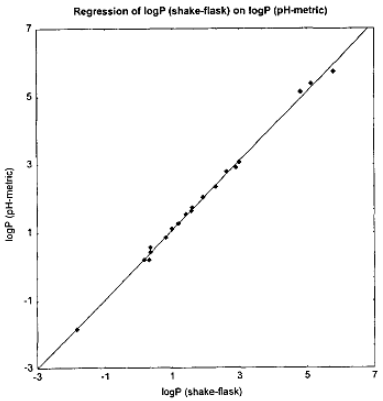


Figure 17: Bjerrum plot for partitioning of (a) the weak acid flumequine and (b) the weak base diacetylmorphine.



$\log P_{\text{pH-metric}} = 0.9794 \log P_{\text{shake-flask}} - 0.0397$

$R^2 = 0.9987$ s = 0.091

F= 8153 n=23

Figure 16: Linear regression model of log P values obtained by shake-flask and pH-metric methods for 23 structurally different compounds in Takács-Novák and Avdeef’s paper [100].

The abovementioned topic of ion-partitioning and the impact of background electrolyte was investigated by studying prostaglandins (Pg) E1 and E2 in different ionic strength conditions [96].
After determining log *P* values in regular conditions (0.15 M KCl, 25 °C), Avdeef's group repeated the same experiment in 0.15 M NaCl, 0.10 M NaCl, and 0.0003 M KCl. Results proved the neutral form of prostaglandines not to be affected by the nature and the concentration of the positive counter-ion in the medium. On the contrary, partition of the anionic species varied within 1 log unit being maximized in presence of a low concentration of KCl, a condition which, however, is far from physiological conditions.

2d Lipophilicity of ampholytes: micro- and macro-partition equilibria

Lipophilicity profile of multiprotic molecules is often hard to predict. As a matter of fact, complex ionization equilibria generate different species in solution and all of them could ideally affect partitioning, thus accurate determination of dissociation constants is crucial to reliably describe partition behaviour by using potentiometric approach.
As discussed for dissociation, both a micro- and a macroscopic perspective can be adopted to investigate partitioning. Referring to an ampholytic drug and focusing on its overall neutral forms, the scheme in Figure 19 should be used to depict the partition scenario.
As in purely aqueous conditions, a tautomeric constant K^Z can be defined in the biphasic system:

$$K^Z = \frac{p^0}{p^{\pm}}$$

Equation 47

Distribution coefficients consider intrinsic lipophilicity of all different species in solution:

$\log D = \log [f^N P^N + \Sigma(f^i P^i)]$

Equation 48

which is proportional to the molar fraction of each form (f). A rule of thumb is that log *P* of neutral species is from 2 to 5 log units larger than that of ionized forms. Approximation are allowed when one species is

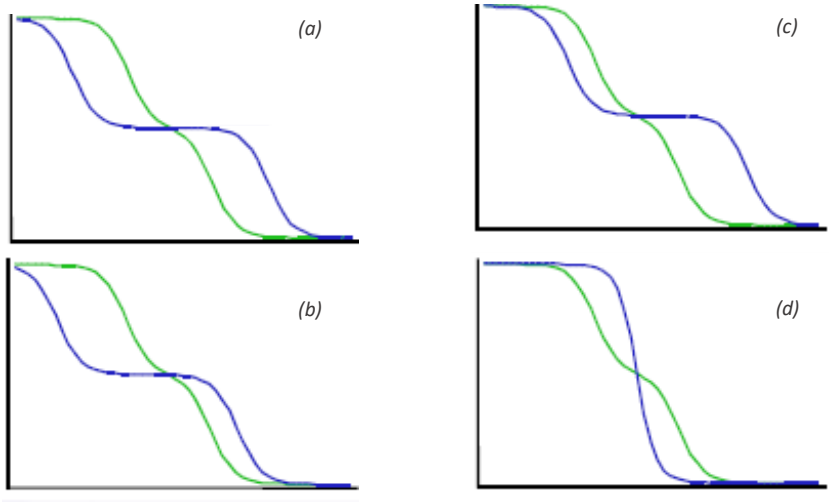


Figure 18: (a, b, c) Theoretical Bjerrum plots for partitioning of an ampholyte and (d) zwitterion. (a) XH partitions; (b) XH and X^- partition; (c) XH and XH_2^+ partition; (d) X^- and XH_2^+ partition.

predominant over the others [7]. Ordinary ampholytes (*i.e.* the ones for which $pKa^{basic} < pKa^{acidic}$) with $\Delta pKa \geq 3$ have only one group charged at a time so that partition will involve only specific micro-species.

For ampholytes with $\Delta pKa \leq 3$ and for zwitterionic ampholytes (*i.e.* the ones for which $pKa^{basic} > pKa^{acidic}$) contribution of different ionized forms should be further investigated.

While distribution (D) is the same whether the species are described in terms of macro- or micro-constants, partition value (P) can be described in terms of macro-log *P* (referred as log *P*) and micro-log *P* (referred as log *p*) as schematised in Figure 19. The former is also named apparent partition coefficient (P_{app}) since it is a cumulative parameter derived from true partition coefficients of tautomeric micro-species. For a typical zwitterion, macro-partition results from the contribution of two micro-species (*i.e.* tautomers):

$$P = p^0 \frac{k_1^0}{K_1} + p^{\pm} \frac{k_1^{\pm}}{K_1}$$

Equation 49

where p^0 refers exclusively to the neutral micro-species ($[XH^0]oct/[XH^0]wat$) and $p^{\pm}$ to the zwitterionic one ($[XH^{\pm}]oct/[XH^{\pm}]wat$). If partition of the neutral is predominant over the zwitterionic form ($p^0 >> p^{\pm}$), log p^0 can be expressed as a function of hydrogen ion concentration ($[H]^+ = h$), micro-constants (k_1 and k_2) and tautomeric constant (K^Z) and related to log *D*:

$$\log p^0 \sim \delta + \log D + \log\left(1 + K^Z + \frac{1}{hk_1^0} + hk_2^0\right)$$

Equation 50

with

$\delta = K^Z p^{\pm}/(2.303\, p^0)$

Equation 51

Partition of micro-species was fully characterised for niflumic acid [101, 102]. Micro-constants were derived from K^Z method (see **3e Determination of micro- and macro-constants by KZ method** section) and used together with pH-metric macro-constants to calculate site specific pKas for carboxylic and aminopyridinyl moieties. Then apparent log *P* was determined by both potentiometric and shake-flask methods, the latter used to experimentally draw lipophilicity over pH profile. Contrary to the common assumption that the neutral form is the only one to undergo partitioning, the authors found zwitterionic and anionic micro-species to enter *n*-octanol. So they highlighted the importance of assessing species-specific log *p* to reveal anomalous behaviour and ion-pair partitioning.
This concept was investigated for tri-dentate molecules, such as tyroxidine that displays phenolate, amino and carboxylate moieties. By integrating

different approaches log *p* was assigned to each micro-species; data proved zwitterionic form to remarkably partition due to its overwhelming dominance over the neutral species. Added to that, cation showed a log *p* close to the one of the neutral form (3.04 vs. 2.98).

It should be highlighted that attribution of lipophilicity parameters on microscopic scale is challenging, time-demanding and relies on the integration of different experimental approaches according to the complexity of the molecule (e.g. UV-spectroscopy, Nuclear Magnetic Resonance, shake-flask, potentiometry). That implies true partition coefficients are seldom investigated and log *D* is preferred to express the overall lipophilicity behaviour of drugs at discovery stage.

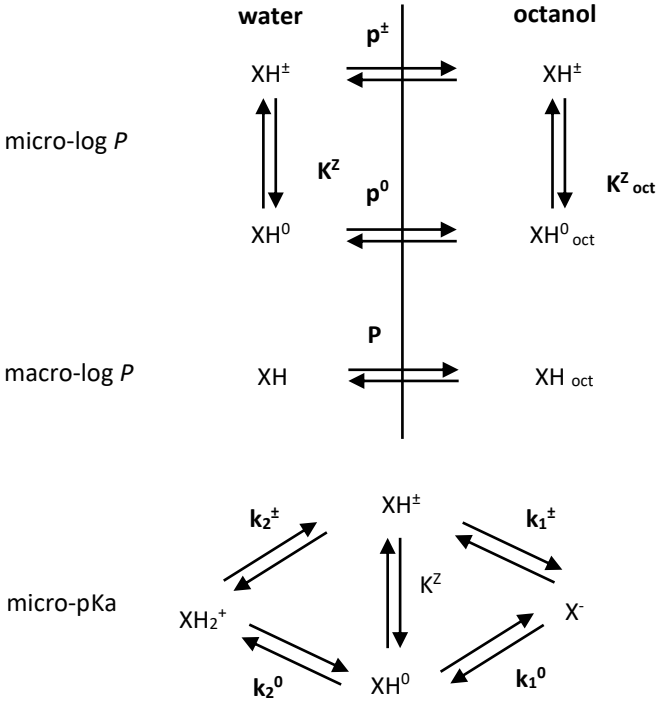


Figure 19: Micro- and macro-partition equilibria.

1.3.5 Lipophilicity of ampholytes: biological implications

As hinted above, the widespread belief that the more the ionization, the less the lipophilicity of drug molecules should be reconsidered for complex structures that often display peculiar features.

Lipophilicity of zwitterions is almost invariant around isoelectric point that is the pH range within pI ± 1. That is why in literature they are sometimes referred as "lipophilicity-buffer" (*i.e.* lipophilicity almost constant in physiological pH region). Such topic has gained increasing attention in molecular pharmacology: when interacting with its biological target, an active compound may encounter micro-pH environments. In this context, the ability of the drug to maintain its hydrophobicity may positively or negatively affect ligand-receptor interaction during long-distance recognition and docking [45]. In case zwitterionic and neutral tautomers coexist in appreciable proportion around pI (*i.e.* small K^Z), they could both contribute to membrane permeation resulting in a better adaptation to biological barriers.

Moreover, the lipophilicity profile is known to correlate to absorption, distribution and metabolism.

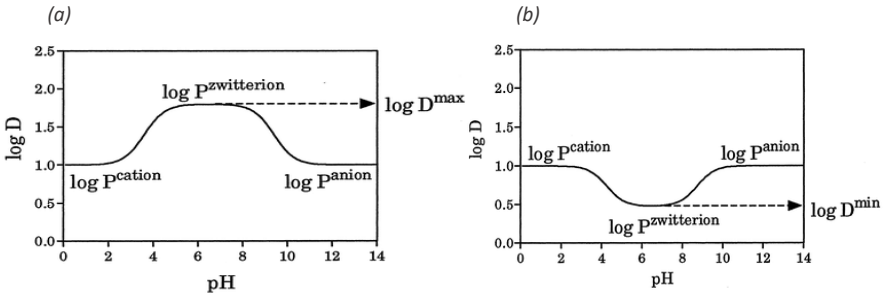


Figure 20: Two extreme lipophilicity profiles for zwitterionic compounds: (a) bell-shape and (b) U-shape [7].

Considering ampholytic drugs, two log *D* profiles are usually observed according to which protonation isomer is favoured by tautomeric equilibrium [7]: on one hand, log *D* of zwitterions with $K^Z \geq 10^4$ lowered to a minimum value at pI, assuming a U-shaped profile. This is the case of β-

lactamic nucleus that does not allow appreciable intra-molecular neutralization and determines in turn a highly polar zwitterion (*e.g.* cephalexine).

On the other hand, zwitterions that show small k^Z values usually reach a maximum log *D* at pI (bell-shaped profile). This behaviour is shared by neutral ampholytes or some zwitterionic drugs with large K^Z whose lipophilicity is magnified by intramolecular effects (*i.e.* molecular proximity, internal ionic bonds, and charge delocalization) [7].

Cetirizine provides a well investigated case in that context: Pagliara et al. studied its properties and compared them to the ones of cationic antihistaminic (anti-H_1) drugs [6].

Zwitterionic tautomer of cetirizine is favoured over the neutral one and gives a major contribution to partition and related phenomena. Anti-H1 drugs with the same lipophilicity profile as cetirizine display lower sedative and cardiotoxic effects in comparison to cationic analogues, due to longer plasmatic persistence and minimal distribution. Intramolecular interactions between opposite charged moieties impact on binding, inter-molecular interaction and site-specific recognition towards plasmatic proteins. As regards metabolism, zwitterionic ampholytes generally exhibit lower affinity towards CYP450 and enter epatocytes less readily than cationic analogues.

1.4 Relevant physico-chemical properties: solubility

Since 1990s, the introduction of high-throughput screening (HTS) and combinatorial chemistry has driven drug-like molecules towards higher molecular weights, increased lipophilicity and reduced solubility: as a consequence of this trend, it has been estimated that 90% of current NCEs suffer from low solubility according to Biopharmaceutics Classification System (BCS) [103]. It is undeniable that the relevance of solubility in preclinical research of potential drug candidates is outstanding and its knowledge is functional to a wide range of needs.
Solubility results from structural properties (*i.e.* molecular weight, aromaticity, flexibility) and, in turn, affects *in vitro* and *in vivo* phenomena investigated at multiple stages of discovery and development.
In the early phases, it is useful for ranking compounds within homologous series; as the candidate enters pre-development, solubility allows to assess absorption, distribution and metabolism; finally, that property is essential for development of formulation for safety studies, preclinical and early clinical trials.

1.4.1 Definition of solubility from a physico-chemical perspective

The concept of solubility can be declined in different ways that involve several approaches for experimental determination. Moreover, special attention should be payed to ionisable over neutral molecules.

1.4.1.1 Thermodynamic and intrinsic solubility

Thermodynamic solubility is defined as the maximum concentration of the most stable crystalline form of a compound that remains in a certain solvent at a given temperature and pressure at equilibrium.
Such equilibrium is reached when interactions between solvent and solid interacting with themselves balance interactions of solvent and solid interacting with each other, *i.e.* balance of forces between solution and solid form.

To reach thermodynamic equilibrium, the system seeks for the most stable conditions: the higher the solid-state energy stabilisation, the higher the energy required to break crystal-lattice and accommodate a molecule of solid into solution by means of solvent-solute interaction.

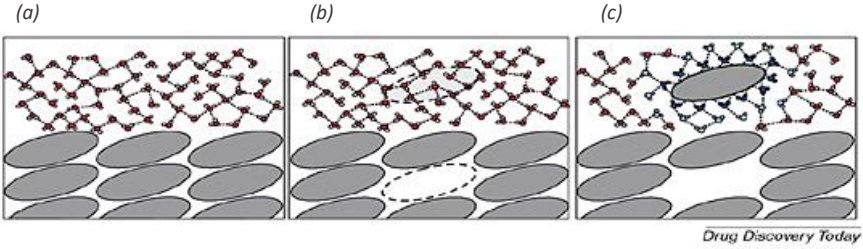


Figure 21: Intermolecular forces involved in thermodynamic solubility [104].

In other words, energy returned by solvent-solute interactions, which represents the entropic contribution to the free energy of the system, is overwhelmed by crystal lattice forces, which determine the enthalpic factor, so that the process is unfavoured.

Thermodynamic solubility is controlled by inter-molecular forces as lattice, cavitation and solvation energies and should be discussed independently for neutral and ionisable compounds.

Considering a general drug solid in water (see Figure 21), in the beginning solid and solvent are segregated and each interacts with molecule of its own kind (Figure 21*a*). Then a solute molecule can be moved into solution breaking crystal lattice forces (lattice energy) and interactions among solvent molecules in a cavity that can host the solute (cavitation energy) (Figure 21*b*). Entropy of the system slightly decreases since the ordered network of hydrogen bonds is locally disrupted, but when the solute molecule is completely surrounded by solvent molecules new stabilizing interactions are developed between solute and solvent (solvation energy). Entropy is increased by mixing of the two species, but is also locally decreased by short-range order owed to the presence of the solute [104] (Figure 21*c*).

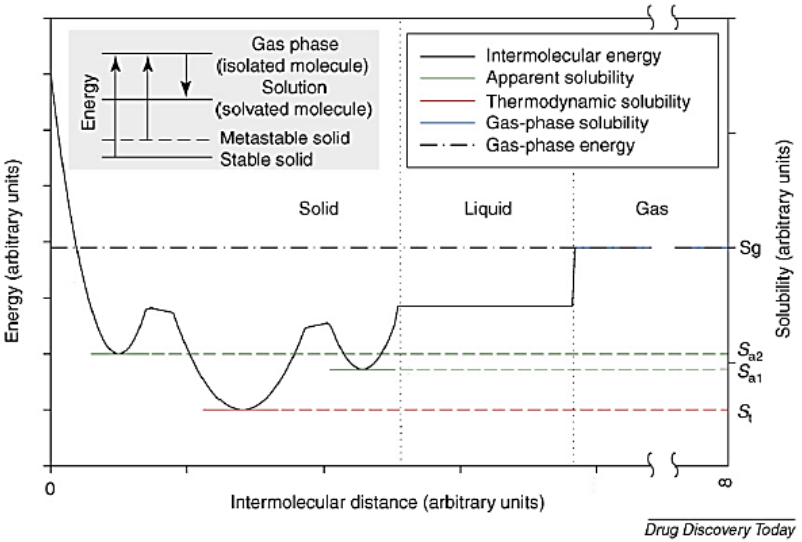


Figure 22: Relation between energy and solubility [104].

Polymorphism is often exhibited by pharmaceutical molecules and it is the property to exist in different crystalline solid forms due to one or two different orientations (*i.e.* lattice structure and/or molecular orientation). Solvent molecules can be included in the crystal lattice generating different crystals known as solvates, also called hydrates when the solvent is represented by water; both solvates and hydrates are regarded as pseudo-polymorphs. Amorphous does not show long-range ordered packing, but only short-distance order. (Pseudo)polymorphs and amorphous exhibit different physico chemical properties, such as density, melting point, hardness, crystal morphology and, of course, solubility [105].

Speaking in terms of free energy landscape (see Figure 22), solid forms are found in minima that represent favourable intermolecular interactions holding the crystal form. The most stable form is located in the deepest trough and gives rise to the lowest thermodynamic solubility (S_t). Metastable solid forms occupy local minima and they are characterised by different degrees of intramolecular energy stabilization. They provide different energy barriers that must be overcome to reach dissolution thus yielding different apparent solubilities (S_{a1} and S_{a2}). If supplied with

activation energy, metastable forms will eventually convert to the one with the lowest energy and return thermodynamic solubility. In the gas phase, solubility reflects only the interaction between solute and solvent (S_g).
Most drugs harbour one or more ionisable sites that involve a more complex definition of solubility. For a neutral drug, thermodynamic solubility is parameterized by a single value that accounts for balance between molar free energy of the solute and the one of the solute interacting with the solvent. In case of an ionisable drug in an aqueous polar environment, ionisability of both solvent and solute should be considered and thermodynamic solubility must be always referred to the pH at equilibrium and the pKas of the drug compound.
Solubility can be profiled as a function of pH according to Henderson-Hasselback equation:

$$S = S_0 [1 + 10^{(pH - pKa)}] \quad (52.1)$$ for a mono-acidic compound

$$S = S_0 [1 + 10^{(pKa - pH)}] \quad (52.2)$$ for a mono-basic compound

Equation 52

where solubility at any pH (S) is a function of solubility of the completely neutral form (intrinsic solubility, S_0) and the dissociation constant of the drug.
Focusing on the case of a monobasic compound of pKa = 5 for example, pH-solubility profile displays four main regions (see Figure 23):
(i) Intrinsic solubility region (pH >7), defined as the pH range in which the drug is completely unionized in solution and displays the lower solubility. In this region, any compound precipitates from solution in the free neutral species despite the starting salt form.
(ii) Ionizing portion (within pH 4.5 and 5) shows the steepest slope and is around the pKa of the solute. Moving of one log unit above or below pKa, the concentration of the neutral or ionized form increases ten folds, respectively. Precipitate formed in this pH range can be in either the free form or the salt form, depending on the strength of the solid state

interactions. According to the number and nature of ionisable sites in the drug, the ionizing portion of pH-solubility curve shows variable complexity. (iii) pH max (∼4) is the point at which maximum solubility is reached (*i.e.* ionizing portion meets the salt plateau).

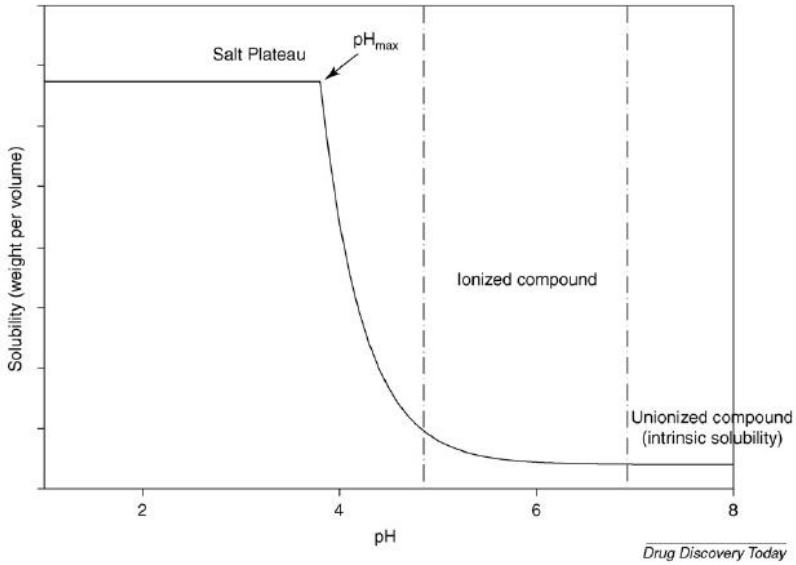


Figure 23: pH-solublity profile of a weak monobasic compound with pKa = 5 [104].

At this pH point the equilibrium state is represented by a salt form in which the completely ionized species of the drug interact with an oppositely charged counter-ion by coulomb interaction.
(iv) The salt plateau (pH < 4) is dominated by solubility of the salt form and it is almost constant. This value depends upon the strength of solid-state interaction within the salt and is parameterized by the solubility product K_{sp}:

$$\text{drug} \cdot \text{salt} \overset{\text{dissolution}}{\rightleftharpoons} \text{drug ion} + \text{salt counter-ion}$$

$$K_{sp} = [\text{drug ion}] \cdot [\text{salt counter-ion}]$$

$$S = \sqrt{K_{sp}}$$

Equation 53

Solubility product is a constant value for a given ionisable drug, thus the maximum concentration reached in saturated solution depends on the concentration of the counter-ion and can be modulated through it.
Moreover, pH-solubility profiles allow to investigate how to improve solubility on a molecular level. Considering a monoprotic compound, a simple interaction between S_0, pKa and Ksp can be defined to assess the pH at which solubility is maximized:

$$pH_{max} = pKa + \log\left(\frac{S_0}{\sqrt{K_{sp}}}\right)$$

Equation 54

Since intrinsic solubility and pKa are drug properties, manipulation of the molecule could improve solubility of the compound. Referring to the monoprotic base which gives the profile in Figure 23, an increase in basic strength (*i.e.* larger pKa value) extends salt plateau and shifts all the curve. This could be useful to shift pH_{max} in the physiologically relevant pH or to make solubility of a solution formulation determined by the desired form, either the free neutral or the salt one. Increasing S_0 would provide higher solubility over the whole pH range, but structural changes that lead to such condition are not easy to harvest.
In general, intervention is made on Ksp to raise salt plateau and solubility at pH_{max}. Being a counter-ion dependent parameter, no chemical modifications on drug structure is required.
Salification represents the key strategy in this context and thus the rational screen of salt selection with low amount of API, a practice known as salt screening, is a crucial step at discovery stage to push solubility of NCE to a suitable level. Several examples of successful salification have been reported in literature: for instance, the aqueous solubility of delveridine

mesylate (320 mg/mL) was improved 2000-fold in comparison to its free base (intrinsic solubility of 143 mg/ mL at pH 6.0). Similarly chlorothiazide potassium dihydrate showed a 400-fold increase in aqueous solubility than the free form of the API [106].
It should be pointed out that an opposite effect could be obtained if ionized drug encounters counter-ions with which it forms poorly soluble salts. Added to that, the modulation of counter-ion concentration is valuable for the subsequent *in vivo* scenario: for instance, the active concentration achieved by poorly soluble hydrochlorides of API is limited by chloride concentration in the gastrointestinal tract, a phenomenon known as "common ion effect" (*i.e.* suppression of the degree of dissociation of the ionisable drug [107], generally followed by precipitation of the solute).

1.4.1.2 Kinetic solubility

Information regarding solubility are crucial at early stages of discovery to guide optimization of lead series, as well as to set operative concentrations for further bioassays.
In preliminary phases of research, determining thermodynamic solubility is not feasible because of limited amount of substance and lack of information about polymorphism of the NCEs. Hence a kinetic parameter is preferably assessed. Kinetic solubility is regarded as the concentration of the drug in solution when precipitation starts and provides information on precipitation rate rather than actual solubility.
In fact, it is common to find poor agreement between kinetic and thermodynamic solubility values owing to the different experimental assays used to assess them. Kinetic solubility is determined from starting material that often carries impurities or is not crystallized, hence resulting concentration is higher than the true value.
Furthermore, kinetic assays, as it will be discussed further in this thesis, involve the use of organic solvent that can affect solubilisation and induce supersaturation of the aqueous solvent through precipitation of amorphous or metastable crystalline form.

Even if the tendency to use kinetic solubility data to assess structure-activity relationships and to take decisions about pre-formulations is widespread, caution should be used: misleading solubility values could mask issues that will appear as soon as the compound will be crystallized [104].

1.4.2 Molecular properties and target-driven effects on solubility

Solubility is determined by two steps each of those could be the limiting one: dissociation of the crystal lattice and solvation of free molecules. Compounds with multiple interaction points and strong intermolecular bonds often show a limited capacity to dissociate from the solid form. Those are sometimes referred as "brick dust" molecules to denote the poor solubility of a stone-like solid, and are characterised by high melting points (Tm > 200 °C).

As soon as the molecule dissociates from the crystal structure, the surrounding solvent needs to prepare a cavity to host it. High molecular weight involves a larger cavity to be formed for incorporating the molecule, resulting in a larger energy penalty associated with the process.

As a final step, the free molecule needs to be inserted into the solvent cavity, but hydrophobic compounds have limited capacity to interact with the aqueous phase. Molecules exhibiting solubility restriction owed to limited hydration are known in the scientific jargon as "grease-balls". When log *P* of a neutral drug generally exceeds the cut-off value of 2-3, hydration becomes a significant limitation for solubility; for ionisable drugs the corresponding log *D* value at the pH of interest should be considered. Compounds with high melting points as well as log *P* (*D*) display both solid-state and solvation-limited solubility and are commonly regarded as "anything-phobic" [108].

General decrease in solubility of drug candidates results from molecular properties that have been selected to hit specific therapeutic targets. In recent years, it has been observed a switch from aminergic G protein-coupled receptors (GPCRs) and enzymes to kinases, ion channels, nuclear receptors and protein–protein interactions. To reach suitable affinity for

these targets, higher lipophilicity or strong intramolecular interactions are required. Those properties depend on molecular features such as an increase in aromatic regions and hydrogen bonding network that in turn increase crystal lattice stabilisation with a negative impact on solubility. It is worth to point out that the relevance of solubility issue depends on the therapeutic area addressed by the NCEs. For instance, solubility is not regarded as crucial for drugs meant to target central nervous system (CNS): in neuroscience, a high potency is required to minimize off-target effects and it is usually addressed by increasing lipophilicity. Moreover, many candidates show amino moieties that are protonated to increase solubility. On the contrary, low solubility is a hurdle for NCEs meant for treatment of oncologic diseases since such compounds usually require high doses that should be completely dissolved in injection volumes for intravenous administration. Current targets of cancer treatment are represented by protein kinases whose inhibitors tend to have highly planar backbones made of several aromatic rings. Such structures account for extensive intermolecular interactions through π-electrons and hydrogen bonding that result in tight packing of crystal lattice. High energy is required to disintegrate intermolecular network and consequently solubility is rather poor.

Hence it is undeniable the importance of profiling solubility during lead optimization and candidate selection stages to unveil potential issues for further development [109].

1.4.3 Relevance of solubility for *in vitro* assays

Solubility affects *in vitro* characterisation regarding ADMET studies and bioassays. Insoluble compounds could have incorrect starting concentration, if poorly soluble in DMSO, or unpredictable final concentration *in vitro* due to precipitation in aqueous media. If the drug precipitates before reaching its cellular target, that will be exposed to a lower concentration than the one expected in experimental design, causing a diminished or undetectable response. Sometimes, low solubility accounts for false negatives that confound the project team: for instance, a clean

profile in a counter assay for off-target activities can be returned by a compound only because slightly solubilised in the medium. Such data can be misinterpreted as a proof of high selectivity thus leading to the selection of a sub-optimal candidate for further advancement [110].
An early knowledge of solubility could be helpful for redefining laboratory practices: for instance, Faria et al. optimised an high-throughput assays in 96-well plates to distinguish binders from substrates of PepT1, a transporter promising for increasing oral absorption of drug-like compounds [111]. In order to reach a concentration in plate that was compatible with solubility of analytes in stock solution, they minimized dilution steps by halving operative volumes and then validated the revisited protocol.
Choosing candidates on the base of solubility values allows a more effective selection for further resource-consuming studies such as structure-activity relationships screening or characterisation of ligand-protein interactions by X-ray or Nuclear Magnetic Resonance (NMR) [109].

1.4.4 Use of biorelevant fluids for *in vitro* solubility assessment

Over the last two decades, development of biosimilar media for studying solubility has drawn the attention of pharmaceutical research [112]. In particular, a large number of modified media have been proposed to better predict gastric and intestinal solubility of NCEs *in vivo*, being the oral route the most exploited. Moreover, experimental data obtained in simulated fluid are sought since potentially useful to improve predictivity of solubility models [109].
For instance, solubility of basic drugs is maximized in gastric environment, but could be negatively affected by the alkaline conditions of the intestinal lumen. The extent at which this phenomenon is balanced by solubilizing properties of biological surfactants should be investigated by using complex media instead of purely aqueous buffers; it results particularly relevant for oral drugs whose solubility is dissolution-limited (BSC II).
To reflect the entire human intestine, several media mimicking gastric juice in pre- and post-prandial conditions (e.g. Fasted State Simulated Gastric

Fluid, FaSSGF developed by Vertzoni [113] and Fed State Simulated Gastric Fluid, FeSSGF [114] respectively) have been introduced. Added to that, small intestinal fluids (e.g. Fasted State Simulated Intestinal Fluid, FaSSIF, and Fed State Simulated Intestinal Fluid, FeSSIF [115]) and colonic media (e.g. Fasted State Simulated Colonic Fluid, FaSSCoF and Fed State Simulated Colonic Fluid, FeSSCoF [116]) have been designed.

Development of such media required a deep knowledge of physiological composition of fluid in different tracts of the gut lumen as well as in different nutritional states. Hence new versions of simulated fluids arose in parallel with availability of updated data collected from human intestinal samples. In addition, physico-chemical properties such as pH, superficial tension, ionic strength and osmolarity should be taken into account [117].

In that context, first intestinal fluids developed by Dressman [118] and Galia [115] were further updated in 2008 [114] and named FaSSIF-V2 and FeSSIF-V2. Improvement regarded not only a better reflection of small intestinal fluid composition, but also higher stability during dissolution assays and short-time storage that is, however, no longer than 24 hours. Furthermore, reproducibility in preparation is essential to introduce these media in laboratory routine and a balance between cost and benefit for scientific investigations was reached by the advent of commercially lyophilized media (e.g. Biorelevant.com Ltd, UK).

All those media are made of aqueous buffer at suitable pH and a mixture of solubilizing additives, such as bile salts and phospholipids. As reported in Table 8 [119], differences between these fluids essentially regard bile acid compositions. Taurocholate, glycocholate, and glycochenodeoxycholate are commonly used as biliar component and cholesterol or other free fatty acids could also be added as additional intestinal surfactants. Despite the large use of phosphatidylcholine as main phospholipid, recent reviews suggested lysolecithin should be preferred. As a matter of fact, that is the predominant phospholipid species in fasted HIF because of the enzymatic degradation of lecithin [117].

Medium	FaSSIF	FaSSIF-V2	FaSSIF-V2 plus	Copenhagen Fasted	SEIF

BS (mM)	GC					1
	GDC					0.7
	GCDC					1
	TC	3	3	3		0.5
	TDC					0.3
	TCDC					0.5
	Crude*				2.5	
PL (mM)	PC	0.75	0.2	0.2	0.625	
	LPC					1
Ratio BS/PL		4/1	15/1	15/1	4/1	4/1
SO				0.5		
Chol				0.2		
pH		6.5	6.5		6.5	6.5
Buffer		Phosph.	Maleate	Maleate	Trizma Maleate	Phosph.
Osmolarity		270	180	181.2**	270	289.25**
Stabilizer	NaN_3					6

Table 8: Composition of several proposed to simulate fasted state in the small intestine reported in Fuch's paper [119].
*SEIF is for Simulated Endogenous Intestinal Fluid; Copenhagen fasted. BS bile salts; PL phospholipids; SO sodium oleate; Chol cholesterol; GC glycocholate; GDC glycodeoxycholate; GCDC glycochenodeoxycholate; TC taurocholate; TDC taurodeoxycholate; TCDC taurochenodeoxycholate; * crude porcine bile extract or taurocholate; ** calculated.*

The surface tension is regarded as a surrogate parameter for the qualitative composition in terms of bile salts and phospholipid content. Indeed, it reflects physiological wetting behaviour of drug substances and its impact on dissolution rate.

Solubilizing ability of simulated and human intestinal fluids (SIF and HIF respectively) is due to the formation of different micellar structure that avoid lipophilic drugs to precipitate out of the system [119].
Recently, a further step towards mimicking actual *in vivo* conditions was made by investigating inter- and intra-individual variability of intestinal fluids in both fasted and fed state [120, 121, 122]. Relatively different compositions are expected to impact on solubility of orally administered drugs and, in the end, to result in a certain variability of their absorption in vivo.

1.4.5 Relevance of solubility for pre-formulation and *in vivo* pre-clinical studies

Following the current trend in pharmaceutical discovery, property-based optimization of lead candidates requires animal studies that assay PK, efficacy and toxicity of NCEs. In that sense, solubility issues result especially challenging for pre-formulation scientists that should optimize solid state (*i.e.* salt vs. neutral, amorphous vs. crystal form) as well as formulation of drug candidates.
Crystalline drugs are usually regarded as less soluble in comparison to amorphous ones since breaking of crystal-lattice forces is energy-demanding. Added to that, polymorphism is often shown by pharmaceutical drugs: polymorphs, hydrates, solvates and co-crystals display different solubility and could interconvert in solution. On the other hand, amorphous form of any chemical entity lacks of long-distance order and thus exists at a higher free energy state than the crystalline counterpart. That involves higher solubility, faster dissolution and potentially enhanced bioavailability. It should be noted that advantages in solubility of amorphous forms remain to be investigated, since the need for rational characterization, prediction of bioavailability and strategies for stabilisation is still unmet [123]. The main concern is that amorphous is intrinsically unstable and can change into a crystalline phase in solution inducing a dramatic drop in solubility.

In vivo low solubility of desired dosage could cause incomplete absorption and thus reduced bioavailability of the drug. Oral route is the most convenient one and general formulations are employed, including 0.5% methylcellulose in water and 0.1–2% Tween 80 with 0.5% methylcellulose in water, but they may turn out inappropriate since they are not optimized on the basis of physico-chemical properties.
Sometimes, efforts put in drug delivery result in promoting an otherwise dismissed drug to the next phase of discovery program, as Palucki reported for a Merck lead candidate meant for subcutaneous administration [124]. Because of small amount of material and short timelines, specific formulation strategies (e.g. nanoparticles instead of DMSO/ aqueous hydroxylpropylcyclodextrin suspension) and vehicles (e.g. non-ionic surfactants, water insoluble lipids, cyclodextrins and phospholipids) [125] can be implemented only if solubility is profiled at early phases of discovery reducing attrition downstream in the development phase.

1.4.6 Solubility determination

1.4.6.1 Computational methods to predict solubility

Predictive *in silico* models for solubility of NCEs are useful for medicinal chemists to study solubility-activity relationships. Several solubility models have been proposed over the years with different level of accuracy, calculation time and physico-chemical interpretation.
Quantitative Structure Activity Relationships (QSAR) and Quantitative Structure Property Relationships (QSPR) are extensively used in this field, since providing good predictive results at a reasonable computational cost. Application of these models, however, is limited to compounds similar to those used in the training set and, furthermore, they lack a complete physical interpretation.
Several fitted or derived general equations involving only few empirical parameters have been proposed: one of the most successful is the General Solubility Equation (GSE) established by Jain and Yalkowsky in 2001 [126].
The GSE states that thermodynamic solubility of the neutral form (intrinsic solubility, log S_0) is related to both melting point (T_m) and lipophilicity:

$$\log S_0 = 0.5 - 0.01(T_m - 25) - \log P$$

Equation 55

$$\log S_{pH} = 0.5 - 0.01(T_m - 25) - \log D_{pH}$$

Equation 56

Moreover, automated data-driven design protocols have been defined, such as Matched Molecular Pair Analysis (MMPA). MMPA explores how a single point change in the molecular structure can affect the property of interest and allowing to acquire previously unknown data from existing datasets. Physics-based models ranging from classical simulations to quantum chemical calculations have been applied to solubility prediction, covering a wide range of complexity (e.g. Molecular Dynamics).

As a matter of fact ionisable molecules are predominant in pharmaceutical industry, thus an increasing effort has been put in studying solvation of ions (e.g. Atomic Multipole Optimised Energetics for Biomolecular Applications, AMOEBA) or amino acids with small water clusters (quantum chemical topology force field, QCTFF) [127].

All the above mentioned solubility models share a common weakness represented by the questionable quality of the training sets. Non-drug-like compounds are often employed and the chemical space defined by the dataset is often poorly representative of actual NCEs, yielding to inaccurate predictions [109].

As summarized in the statement "garbage in, garbage out" the experimental reliability of the training data is another impactful factor on the predictivity of the model. For instance, one of the most repeatedly used dataset includes solubility values spanning between -11.62 to 2.77 on a log molar scale, corresponding to 2 pM and 589 M respectively [128]. A higher uncertainty is expected for the lowest value of the range as well as for the highest one, which is greater than that of water (*i.e.* 55 M). If such data are employed in computational modelling, a good measure is to weight the influence of observation by the accuracy of experimental data to minimize the impact of poor quality results on the output [108]. Furthermore, global

models calibrated by using a wide range of compounds (*i.e.* large chemical space) are usually less predictive than local models, trained by using series of compounds; although the former are more useful for global ranking, the latter return more accurate predictions, even if requiring more resources for development. In any case both approaches are usually employed in drug discovery according to available resources and project needs.
A general hinder in predicting solubility is represented by the complexity of the process since both dissociation of molecules from the solid state and solvation should be taken into account; other difficulties concern the uncertainty in the prediction of pKa(s) for ionisable drugs and, above all, crystal packing. Thus, quality datasets are ideally composed of solubility data aligned with solid state characterisation [109].

1.4.6.2 Experimental assessment of solubility

	Solubility	Method	Format	Compound	Solvents	Dispensing	Incubation/ Mixing	Workup	Analysis
DISCOVERY	kinetic	HT	micro plates	DMSO soution	aqueous	robot	seconds to minutes	none	light scattering
								filtration	absorbance
DEVELOPMENT	thermod.	small scale	single tubes	solid of known form	aqueous organic excipient formulation	manual	hours to days	filtration	HPLC or pH titration
									solid state analysis

Table 9: Solubility determination in drug discovery and development and assay workflow [132]. Thermod. = thermodynamic; HT = high throughput.

In drug discovery and development pathway solubility of a drug candidate is experimentally determined several times.

As soon as synthetic protocol and purification are implemented, NCE changes and so do the tools utilized for solubility characterization in terms of speed and accuracy.
In the following paragraphs, classical approaches for thermodynamic and kinetic solubility determination will be discussed and special attention will be paid to turbidimetric approach.

2a Experimental determination of thermodynamic solubility: shake-flask approach

From a practical perspective, equilibrium solubility is the concentration of a saturated solution in equilibrium with undissolved solid. Classical protocols take advantage from shake-flask method [123]: an excess of solid is put in a certain volume of the solvent of interest (generally an aqueous buffer or a biorelevant medium) and stirred for a time period assumed to be sufficiently long for equilibrium to be established (from 48 hours to 2 weeks). The required amount of drug depends on the estimated solubility for the analyte and the set-up of the assay, but is usually in the range from 100 µg to 20 mg.
In the classical shake-flask approach, a sedimentation phase lasting from 24 hours to 3 days is interposed between shaking and analysis and represents the safest way to separate undissolved solid. In a review by Baka and co-authors, sedimentation step was proved even more crucial than the stirring one [129]; alternative techniques for separation are centrifugation and filtration. The latter has been questioned to affect resulting concentration since small particles could be unretained by filters and thus increase the apparent solubility value. Nonetheless this procedure is preferred over stirring especially in miniaturized approaches [130]. Finally, concentration in the supernatant is assessed by suitable techniques (e.g. UV detection).
Classical thermodynamic assays are based on a single measure, but in order to be sure equilibrium was achieved (*i.e.* invariant concentration over time) sampling at multiple time points is recommended. Furthermore,

characterization of the remaining excess solid is essential to unveil any on-going transformation, hence confirming the nature of the equilibrium state. It is apparent that the shake-flask assay is substance- and time-demanding so that miniaturized approaches in 96-well plates have been developed to meet the needs of discovery. Protocols recall the classic shake-flask approach, but volumes are miniaturized from several milli- to micro-liters to save drug material. Instead of directly using powder, a solution of the analyte could be spiked in the well at the beginning of the assay and then evaporated to obtain solid material as starting form [131]. Since evaporation of DMSO may lead to loose low molecular weight compounds (MW ≤ 250), acetonitrile or methanol/dimethoxyethane mixture 50/50 can be used instead [132].

Thermodynamic solubility refers to a specific crystalline form and precisely the most stable one. Thus experiment should start from such form at a high degree of purity to have the best chance to reach equilibrium after a lab-compatible amount of time (*i. e.* form several hours to several days). However, at early discovery stages solid state of a drug candidate is seldom characterised and short incubation time would be un-sufficient for metastable forms to convert in the most stable one and the resulting solubility should be apparent for a different crystal form.

Moreover, pKa of the compound should be measured and pH of the solution should be checked at the beginning as well as at the end of the assay: in fact, if the measured solubility lays on the steepest region of pH-solubility profile (ionizing region), small changes in pH could heavily affect the result.

Other experimental features should be considered, such as buffer that is usually a mixture of different ions, and the solid in equilibrium with the solution at the end of the assay. Finally, undesired phenomena like precipitation of insoluble salts (often hydrochlorides) and common ion effect could impacts on the reliability of the final result [104].

2b Experimental determination of kinetic solubility: miniaturized shake-flask approach

Even if of utmost importance, experimental determination of thermodynamic solubility during early discovery is hampered by substance availability (from 3 to 10 mg or even more are required for in-depth measures), impurities or presence of amorphous form, hence it is usually reserved to development stage.

To overcome these issues, kinetic solubility is preferably assessed in the preliminary phases of research by using medium-high-throughput methods that take advantage from anti-solvent precipitation.

Kinetic assays are designed to minimize the request for substance to sub-milligram amount and to facilitate screening of hundreds of compounds per week by applying solubilisation processes that start from stock solution of the drug. Typically, the compound is dissolved in an organic solvent to prepare a solution of known concentration that is added gradually to the aqueous solvent of interest until the anti-solvent properties of the aqueous medium trigger precipitation of the drug from solution. Such phenomenon is optically detected and kinetic solubility is measured at the point at which water can no longer provide solvation of the drug.

Two approaches are usually adopted to quantify analyte in solution: one involves removal of the precipitate by filtration or centrifugation after incubation time (*i.e.* several hours) and then quantification of the analyte by HPLC or plate readers. On the other hand, formation of the precipitate can be monitored indirectly by an increase of UV-absorbance (*i.e.* base line scatter) due to particles hindering light pathway to detector. Another option is to monitor precipitation directly by detecting light scattering signal through a nephelometric turbidity detector.

DMSO is employed as standard solvent owing to its ability to solubilize a wide variety of drugs and its widespread application in multiple studies (e.g. *in vitro* and *in vivo* biological studies) [132]. DMSO could affect intrinsic solubility of the analyte acting as co-solvent: while reducing dielectric constant of the solution, it may promote solvation of lipophilic drugs, but generally percentages below 2% are regarded as safe. Moreover, it may

promote precipitation of the drug into the stock solution due to freeze-thaw cycles. Despite other co-solvents have been proposed, such as ethanol and dimethoxyethane, DMSO is *de facto* the most common solvent employed to store NCEs and distribute them to pre-clinical assays.
Aqueous buffers like phosphate are the most commonly adopted in screening assays and in the last decade more biorelevant media have been proposed to mimic conditions closer to biological environment (see paragraph **1.4.4 Use of biorelevant fluids for *in vitro* solubility assessment**).
Miniaturization and automation support the use of 96 to 384-well microtiter plate reducing the need for manual-liquid handling and minimizing substance and media requirements.
The information provided by kinetic solubility values does not account for crystal-lattice break because no equilibration takes place in the aqueous phase: hence these data should be valuable only to define the feasibility of biological assays. There is the widespread tendency to classify NCEs according to their kinetic solubility, but such ranking must be considered with caution and validated by comparison to thermodynamic data as soon as they are available [104].

2c Potentiometric CheqSol method: a medium throughput assay for solubility in drug discovery

2c.1 Experimental assay

Avdeef was a pioneer in the application of potentiometric approach for solubility profiling of ionisable drugs in the 1990s [133]. The method takes advantage from monitoring perturbation of dissociation equilibrium, as described in the previous paragraphs for log *P* assays: while in that case shift of pH-distribution profiles was due to partition, in solubility assays it is owed to loss of compound from solution through precipitation.
The original potentiometric procedure set by Avdeef was tedious and time consuming (from 8 to 24 hours per determination), but had the major advantage of drawing the whole pH-solubility profile of an analyte by

performing just one titration. The method was regarded comparable to shake-flask and results returned from it proved more reliable than pH-solubility curves acquired in a series of different buffers.

Box and Stuart further developed a method to assess intrinsic solubility [134, 135] called "Chasing Equilibrium" (CheqSol) that they integrated in the commercially available apparatus by Sirius Analytical UK (*i.e.* GLpKa™ and, more recently, T3™).

"Chasing equilibrium" refers to the practice of flipping solution from supersaturated to under-saturated state by changing the pH [123]. This assay is based on the evidence that alternative dissolution and precipitation of an ionisable compound could be induced by controlled addition of basic or acidic titrants and the resulting effects on pH can be monitored over time (pH gradient).

$$\text{pH gradient} = \frac{\text{dpH}}{\text{dt}}$$

Equation 57

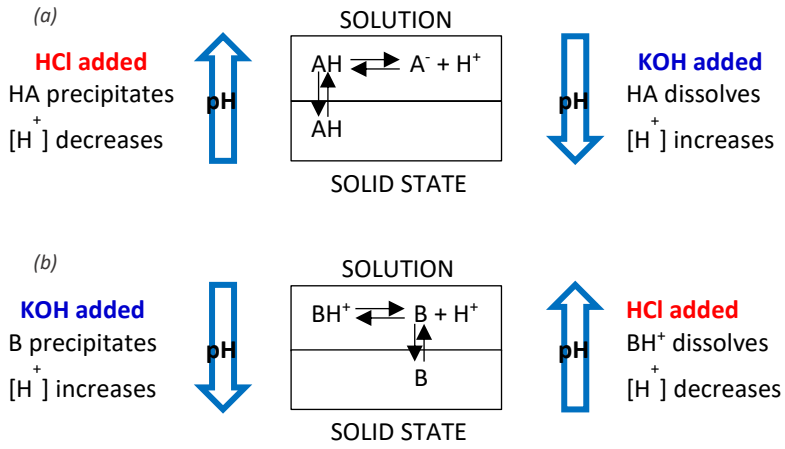


Figure 24: Changes of pH-gradient depicted for (a) a weak acid and (b) a weak base.

For a weak acid (case (*a*) in Figure 24), addition of acidic titrant (*i.e.* HCl) induces precipitation of the neutral species that leads the solution to a sub-saturated state: protons are subtracted from solution producing a pH

increase (positive gradient, dpH/dt > 0). On the contrary, when base (*i.e.* KOH) is dropped into the system, solid precipitate re-dissolves generating a super-saturated solution (*i.e.* concentration is higher than thermodynamic solubility) in which dissociation is promoted, generating an increase of hydrogen-ion concentration and hence a pH decrease (negative gradient, dpH/dt < 0). The same considerations could be made for a weak base (case (*b*) in Figure 24*).*

The experimental apparatus for CheqSol assay includes an optical dip probe connected to a source of light and it is integrated into a titration system: so a UV-spectrum is collected at any pH-point of titration, at a wavelength at which the analyte is not absorbing.

A certain amount of drug is completely dissolved at a pH where it exists predominantly in its ionized form (e.g. pH ≥ pKa + 2 log unit for a monoprotic acid as diclofenac, reported in Figure 25), and then precipitation of the neutral species is induced by adding acidic titrant. The occurrence of precipitation is detected as a scatter of the base-line of UV-spectrum and at that point concentration of the drug in solution equals kinetic solubility. That is also the pH point at which oscillation between supersaturated and under-saturated states is induced: after detection of precipitation, addition of acidic titrant is stopped and pH gradient is monitored each second.

Linear fit of experimental points is performed and when correlation is acceptable (R^2 > 0.9 for at least n = 30 points) mean slope and average pH are noted. Two more slopes are recorded before reversing the gradient by adding the counter-titrant (*i.e.* base in this example) and measurements are performed in the other direction.

At equilibrium dissolution and precipitation phenomena are balanced and pH gradient is invariant (*i.e.* dpH/dt = 0): thus the intrinsic solubility of the neutral form can be calculated as follows.

Some information are needed before starting the assay, that are (i) the formula weight of the compound, (ii) the weight of the solid sample, (iii) the concentration of the acid and base titrants and (iv) the ionization constant(s) (pKas) of the sample, both the value and type (acid or base). At

each point of titration, the volume of added water, acid, and base titrant and also pH is known.

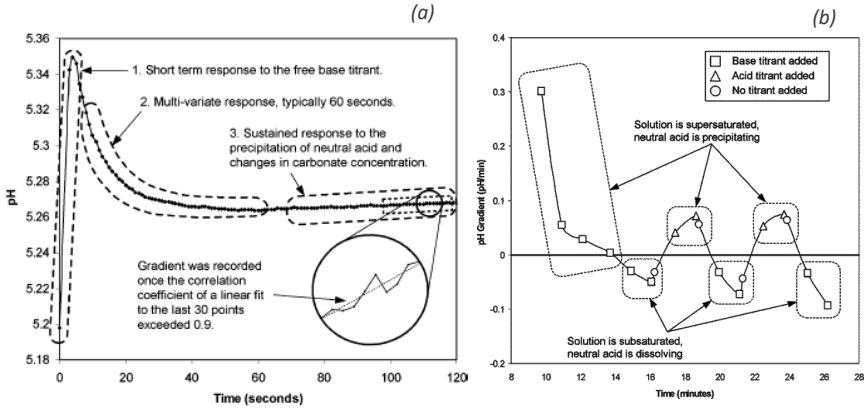


Figure 25: (a) pH response in a solution of diclofenac after addition of base titrant. The neutral form precipitates and a positive slope is recorded.
(b) Change of pH gradient while chasing equilibrium of diclofenac [134].

The concentration of free hydrogen ions is determined from the p[H] (see **2b Calibration of the combined glass electrode** section for calibration of the titrating system):

$$[H^+] = 10^{-p[H]}$$

Equation 58

The total volume (V_t) of the assay is calculated from the additive contributions of water (V_w), acid (V_a) and base (V_b):

$$V_t = V_w + V_a + V_b$$

Equation 59

The concentration of free hydroxide ions is derived from water self-dissociation constant (K_W):

$[OH^-] = K_w/[H^+]$

Equation 60

correcting K_w value for temperature and pH.
The resulting concentration of free positive ions (e.g. K^+) is determined from the amount of base titrant (e.g. KOH) plus any positive ions in the original sample if it was in salt form:

$$[K^+] = \frac{V_b \cdot C_b + (m_s \cdot z_{s+}/FW)}{V_t}$$

Equation 61

where C_b is the concentration of base titrant, m_s is the sample weight of test compound in grams, z_{s+} is the charge of any positive salt counter-ions and equals to zero if the sample was introduced as free neutral form) and FW is the formula weight of the test compound.
In a similar way, concentration of free negative ions (e.g. Cl^-) is obtained from the amount of acid titrant (e.g. HCl) plus any negative ions in the original sample if it was a salt:

$$[Cl^-] = \frac{V_a \cdot C_a + (m_s \cdot z_{s-}/FW)}{V_t}$$

Equation 62

with C_a as the concentration of acidic titrant and z_{s-} is the charge of any negative salt counter-ions; $z_{s-} = 0$ if the analyte is in its free neutral form).
The concentration of all the ionized species of the sample is then determined from a charge balance equation, which results more or less complex according to the number of ionisable sites displayed by the analyte.
For instance, for a monoprotic acid HA, this is simply the concentration anion A^-:

$$[A^-] = [H^+] - [OH^-] + [K^+] - [Cl^-]$$

Equation 63

Knowing the free hydrogen ion concentration $[H^+]$ and the sample ionization constant (pKa), the concentration of neutral species in solution is calculated from the concentration of the ionized species. Considering a monoprotic acid:

$$[\mathrm{HA}] = \frac{[\mathrm{A}^-][\mathrm{H}^+]}{10^{-\mathrm{pKa}}}$$

Equation 64

Accuracy of pKa determination strongly affects the one of the solubility measurement: it is estimated that each unit of error in the value used for the pKa closest to the neutral species gives an error of one log unit in the intrinsic solubility result.

Each time the pH-gradient switches direction, the intrinsic solubility is delimited by the neutral species concentration at the points on either side of the change. Difference in concentration between those two points is small, and both return a value close to the intrinsic solubility.

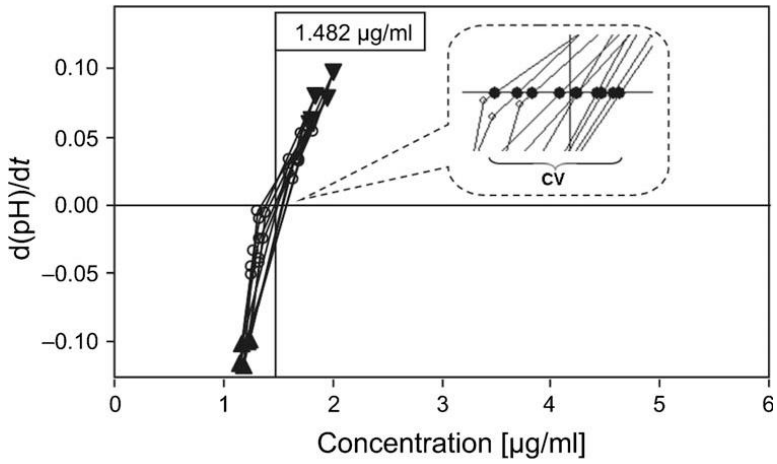


Figure 26: Crossover plot for the weak acid diclofenac [135].

Each value of mean pH-gradient can be plotted as a function of the corresponding calculated concentration to obtain the "Crossover Concentration Plot" and the points lying on y = 0 are averaged to return intrinsic (thermodynamic) solubility for the analyte.
Kinetic solubility can be calculated as well and represents the concentration of the neutral form of the drug at the pH of precipitation.
The authors recommend to use this value only as additional information about how much the solution remains supersaturated shortly after precipitation begins. This concept is parameterized by degree of supersaturation that is the ratio between kinetic and intrinsic solubility. However kinetic solubility values returned by CheqSol method are not expected to be comparable to the ones from other kinetic assays [134] since highly time-dependent.

2c.3 Bjerrum plots: solubility and precipitation curve

Bjerrum curve visualizes changes in protonation of the sample state caused by ionization. As long as the drug is in solution, the graph shows the average number of bound protons (*i.e.* molecular mean charge, Bj) versus pH as described in the **2c pKa determination** section of this introduction.
Once precipitation occurred, the chasing equilibrium procedure quickly brings the solution close to equilibrium with the solid precipitate and then oscillation between supersaturation and under-saturation is induced, with very small changes of the neutral species concentration.
In theory, all the data points collected during chasing should lie close to a Bjerrum function (named precipitation curve) that can be calculated from the known data and the intrinsic solubility result, assuming that all the solid is in the form of the neutral species.
Suitable equations have been proposed for calculating precipitation curves for monoprotic drugs as well as for ampholytes:

$$B_j = 1 - \frac{S_0 Ka}{[X_{total}][H^+]}$$

Equation 65

for a monoprotic acid;

$$B_j = \frac{S_0[H^+]}{[X_{total}][Ka]}$$

Equation 66

for a monoprotic base;

where S_0 is the intrinsic solubility of the drug (M), $[X_{total}]$ is the sample concentration (M), $[H^+]$ is the concentration of hydrogen ions (M), and Ka is the dissociation constant of the sample (M).

As depicted in Figure 27, drug precipitation from solution induces a shift of the curve. Distance between the curves depends on intrinsic solubility of the analyte: at a certain concentration, the lower the solubility the larger the distance [134].

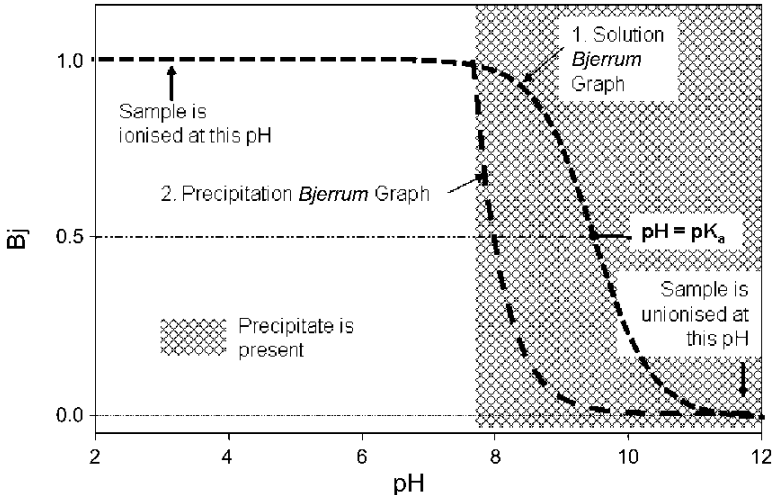


Figure 27: Theoretical Bjerrum curves for a weak base.

2c.4 Chaser and non-chaser: an empirical classification of drug molecules

Box and co-authors proposed a system of classification for ionisable drugs based on their experimental behaviour in solubility assays. If after

precipitation all the experimental points fall on the same Bjerrum precipitation curve (e.g. imipramine [135]), the compound is regarded as non-chaser which means it is not able to supersaturate. Hence intrinsic solubility can be approximated to kinetic solubility.
On the contrary, if the sample (e.g. ibuprofen [134]) is able to supersaturate (*i.e.* solution can be repeatedly switched from a supersaturated to an under-saturated condition in CheqSol assay), the compound is classified as a chaser that exhibits kinetic larger than intrinsic solubility.
These two classes display different physico-chemical properties: for instance, chasers include basic, acidic and ampholytic compounds while non-chasers are predominantly bases; the former undergo slow precipitation and dissolution that makes "chasing equilibrium" feasible, on the contrary non-chasers precipitate immediately as soon as critical pH is reached and do not re-dissolve on the time scale of solubility assay.
Chasers free form are usually crystalline thus showing high melting points and high hydrogen bonding capacity. On the other hand non-chasers are often amorphous or meta-stable, with low melting points (usually below room temperature) and low hydrogen bonding propensity.
It has been estimated 70% of current commercial drugs are chasers so that investigation of their property has been gaining increasing attention in the landscape of pharmaceutical research. As a matter of fact supersaturation can be modulated by different additives or by using biorelevant media and represents a potential strategy to increase concentration of the drug at the absorption site (see paragraph **1.4.7 Strategies to overcome poor solubility of drug candidates**).
Another advantage of potentiometric solubility assay is that it reveals solid-form changes that may be pH or time dependent as reported for sulindac. Such drug displayed a dual behaviour: after precipitation occurred, it initially acted as a non-chaser (S = 71 µg/ mL) but, after spending 20 minutes around pH 6, it behaved as a non-chaser with lower solubility (S = 11 µg/ mL). The chaser form was proved to be a different polymorph from the non-chaser one [135]. In that sense, this technique opens to

preliminary investigation of solid form stability in different pH or medium environments.

1.4.7 Strategies to overcome poor solubility of drug candidates

Solubility proves critical for oral absorption that could be regarded as a two-step process: first dissolution is required for the drug to pass in solution as molecular dispersion; then permeation through biological membranes addresses it to the target site.

Poor oral absorption could derive from low-solubility as well as from low-dissolution rate. In the latter case, orally administered drugs may miss the time window for absorption in the small intestine (*i.e.* 1.5-4 hours according to the species). Solubility, in turn, could be dissolution- (*i.e.* the case of brick-dust molecules) or solvation-limited (*i.e.* the case of grease-ball molecules) so that different strategies have been described to overcome each of those issues.

For brick-dust molecules, manipulation targeting the solid crystal form represents a useful strategy: for instance, salt formation is very attractive for pharmaceutical industry since leads to increased dissolution rate, higher apparent solubility in physiologically relevant pH range and, finally, more effective absorption *in vivo* [108]. From a physico-chemical point of view, the coulombic attraction within the drug ion and the counter-ion changes the energetic landscape of the solid state leading to stronger interaction between solute and polar aqueous solvent [104].

Another option is offered by prodrugs that have successful history in medicinal chemistry (e.g. fosfluconazole, fosphenytoin and fosamprenavir) mainly as regards phosphate esters.

Such prodrugs undergo enzymatic cleavage by alkaline phosphatase in the gut lumen or blood to release the parent drug. Other types of water-soluble prodrugs include the use of amino esters, polyethyleneglycol (PEG) esters and others. Release of the active compound, however, may lead to

potential toxicity liability caused by the prodrug-solubilizing auxiliary that should be carefully evaluated in advance [109].

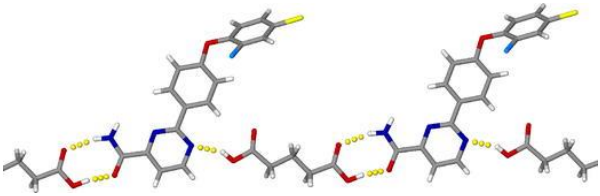


Figure 28: Hydrogen bonding in NaV1.2 blocker and glutaric acid co-crystal reported in [136]. Interactions between the carboxylic acid groups of the glutaric acid molecule and the amide and pyridine groups on the drug molecule are shown.

Co-crystals represent an uprising approach to overcome poor solubility. That consist of a solid form option for non-ionizable APIs that show oily or amorphous nature, stability issues or low solubility/poor dissolution rate. Co-crystals are formed through specific non-covalent interactions between the NCE and one or more neutral solid ligands, named co-formers. Co-crystallization proved successful in several cases, such as for NaV1.2 blocker co-crystal with glutaric acid [136].

Amorphous Solid Dispersions (ASD) allow to increase the apparent solubility over that of the corresponding crystalline form. ASD include a drug and a soluble component (e.g. polymers like polyvinylpyrrolidone, PVP) in mixture, formed though co-precipitation from a common solvent, induced by evaporation. Other than promoting solubility, the polymeric component reduces the mobility in the amorphous drug and thus stabilizes its solid state [137].

Moreover, in order to decrease energy associated with cavity formation into the solvent, the tight structure of water can be loosen up by adding co-solvents, such as ethanol or polyethyleneglycol. In the case of grease-ball APIs (log $P \geq 3$), solvation problem can be solved by means of lipid-based formulations that include lipids, surfactants and/or co-solvents.

The most troubled case is the one of drugs showing high melting point, molecular weight and lipophilicity: indeed such compounds are hard to manipulate even by working on formulation and are likely to generate too

low concentration *in vivo* to be active [108]. Special techniques are used to formulate such insoluble compounds, such as sprayed dried dispersions (SDDs), self-emulsifying drug delivery systems (SEDDS), micronization and nanoparticle approaches [109].

1.4.8 Supersaturation and classical nucleation theory

Supersaturated Drug Delivery Systems (SDDS) are a useful strategy to increase the amount of the drug at absorption site and, consequently, its bioavailability.

Supersaturation is a metastable condition in which concentration of the drug in solution exceeds thermodynamic solubility value. Being intrinsically unstable, the system will spontaneously revert to a more energetically favourable state represented by equilibrium solubility by means of precipitation.

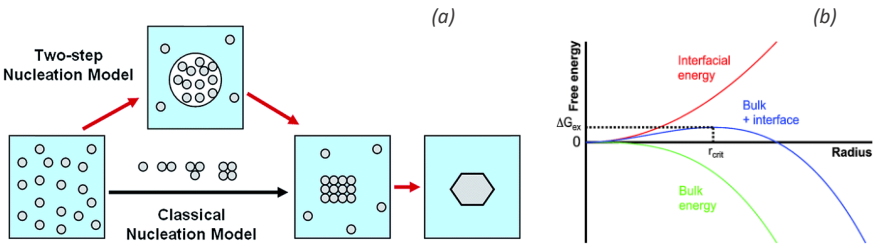


Figure 29: (a) Classical nucleation theory and (b) energy landscape related to CNT nucleation [140].

Supersaturated solutions are often characterised in terms of degree of supersaturation (DS):

$$DS = C_{supersation}/C_{equilibrium}$$

Equation 67

that is the ratio between the concentration (C) of the supersaturated system over the thermodynamic equilibrium solubility of the solute.
In terms of formulation strategy, SDDS are effective only if the drug could maintain an appropriate degree of supersaturation for a reasonable time period to allow a suitable amount of drug to be absorbed; that ability, however, is strongly compound-dependent.
Recalling the free energy landscape (see Figure 22), a supersaturated system is characterised by an increased chemical potential (μ) in comparison to the corresponding saturated or under-saturated one. In a supersaturated solution:

$$\Delta\mu = \mu_{supersaturation} - \mu_{equilibrium} > 0$$

Equation 68

that makes the system thermodynamically unstable and prone to precipitation.
Precipitation is conventionally described as a step-wise process that includes two main events: nuclei formation and crystal growth. Even if the physical mechanism through which activation energy is overcome is still unknown, it is speculated that nuclei-precursors are continuously forming and breaking until a critical radius (r*) is reached.
Critical radius is defined by Gibbs-Thomson equation:

$$r^* = (2\Upsilon\upsilon)/ \Delta\mu$$

Equation 69

where ϒ is the specific interfacial energy between the media and the nuclei, and υ is the molecular volume.
At the critical radius (r*), the nucleus-embryo has the highest interfacial energy towards the solvent and can either grow larger to form a nucleus or disintegrate.
For dimension above the critical radius, energetic barrier for the second step have been exceeded and crystal will grow to form the final poymorph until equilibrium has been reached (*i.e.* $\Delta\mu = 0$).

Nucleation rate (J) is defined by using Classical Nucleation Theory (CNT, see Figure 29), which refers to the ideal condition of a spherical nuclear system:

$$J = A\exp\left(\frac{16\pi\vartheta^2\gamma^3}{3k^3T^3\ln(DS)^2}\right)$$

Equation 70

where k is the Boltzmann constant, T is the absolute temperature, and A is a pre-exponential factor influenced by several conditions, such as growth rate and stability of the nuclei-precursors.
Assuming induction time (t_{ind}) for precipitation being inversely proportional to nucleation rate, the following linear relationship can be defined:

$$\ln(t_{ind}) = \alpha + \beta\ln(DS)^{-2}$$

Equation 71

Classical nucleation theory has been applied to study supersaturation-precipitation behaviour of poorly soluble drugs [138, 139] both in buffer and physiologically relevant media. However, its applicability to amorphous precipitation still needs to be investigated.
Alternative theories have been developed based on experimental evidences, like the two-step nucleation theory that was initially proposed for protein crystallization [140, 141]. This model involves formation of pre-critical amorphous clusters of solute molecules that grow into sufficient-size mesoscopic clusters. Then they re-organized into an ordered structure generating the final crystal form.
Applicability of the two-step mechanism to both macromolecules and small organic molecules have been proved by theoretical as well as experimental studies, suggesting that this mechanism may underline most of crystallization processes from solutions.

2. Aim of the work and research steps

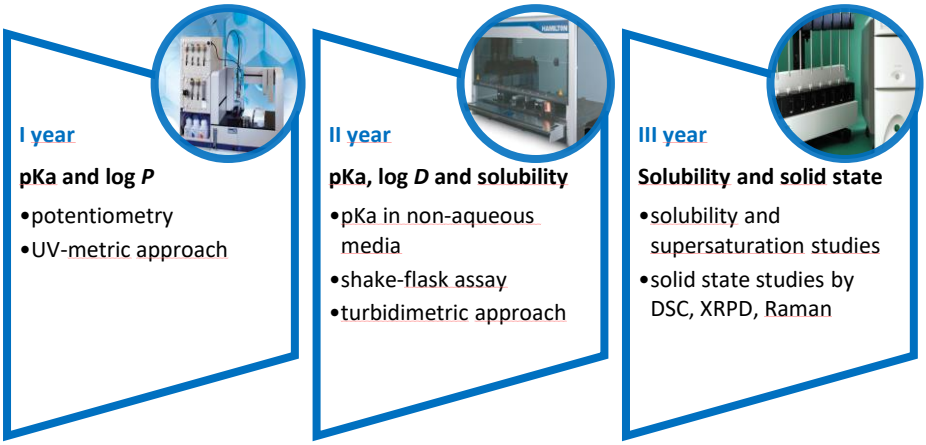


Figure 30: Scheme of the research steps.

The aim of the present Ph.D. project was to develop and define an operative workflow to efficiently profile some relevant physico-chemical properties (*i.e.* ionization constants, lipophilicity and solubility) of small-molecule drug candidates at discovery and pre-development stage.

The work was developed over three years involving three main steps: the first milestone was to investigate applicability of potentiometric and spectrophotometric approaches to pKas and lipophilicity determination.

Taking advantage from the knowledge of the limitations of such approaches, a miniaturized shake-flask operating procedure was designed to assess log $D_{7.4}$ of drug candidates that were not suitable for potentiometric assay; moreover, acid-base equilibria in non-aqueous solvents were studied by using an *ad hoc* UV-metric set-up.

Finally, solubility was measured by means of a turbidimetric technique both in common aqueous buffer and biorelevant intestinal media and evaluation of solution-mediated interconversion between solid forms of the drug molecules was attempted. A critical approach to the employed experimental methods was applied to point out advantages and drawbacks

of each technique thus selecting in each single case the most suitable protocol for industry practical needs.

Added to experimental work, the Ph.D. project focused on theoretical aspects related to physico-chemical profiling, such as a deep understanding of the complex dissociation behaviour in solution of ampholytes and its influence on pharmacokinetics phenomena.

Furthermore, experimental data collected during the project were compared to predicted values and employed to support the in-house training of an expert system for physico-chemical properties prediction, with a particular reference for pKa and solubility.

The strict confidentiality of produced results prevented their publication in respect to Chiesi Farmaceutici S.p.A. legal policy.

3. Dissociation constants

3.1 Assessment of macroscopic constants by potentiometric and UV-spectrophotometric approaches

3.1.1 Aim of the work

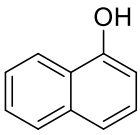
In the introductive chapter of this thesis relevance of pKa experimental determination was widely discussed: in that context, the goal of the first part of this research project was to evaluate the suitability of Sirius Analyical hardware-software packages (GLpKa™ and T3™) to a research and development center needs.

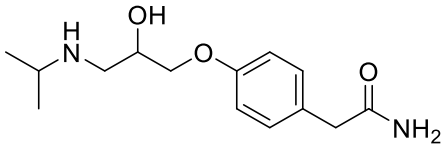
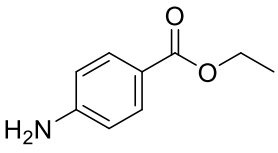
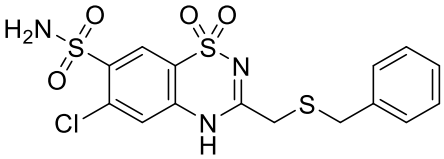
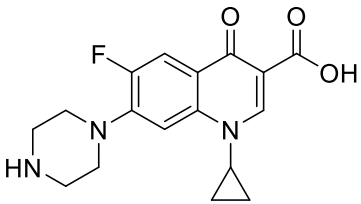
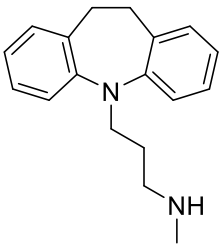
Such techniques were expected to be applied to drug-like molecules that have passed early drug discovery phases (*i.e.* optimization of the synthesis) being accessible in solid form thus allowing in-depht physico-chemical characterisation.

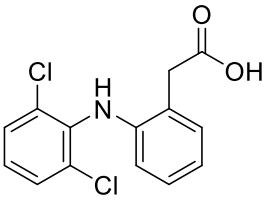
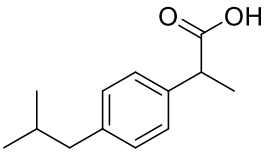
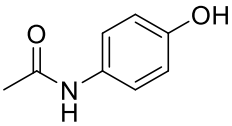
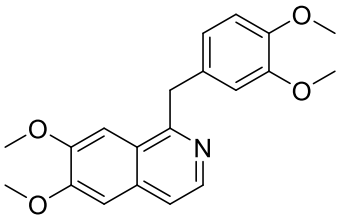
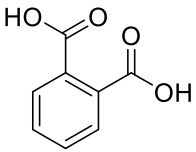
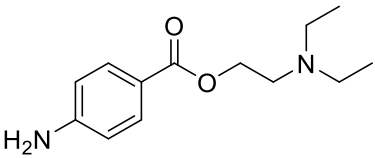
For this purpose, a set of commercially available standards was selected and tested to practice the methods that were then transferred to proprietary NCEs. Correlation of experimental data to computed values by ACD Lab Percepta 2012 v. 8 (Advanced Chemistry Development, Inc., Toronto, Ontario, Canada) was calculated for drug candidates to evaluate predictivity of the predictive software.

3.1.2 Experimental part

3.1.2.1 Materials

Name	Structure	MW
1-naphtol		144.2

atenolol		266.3
benzocaine		165.2
benzthiazide		431.9
ciprofloxacin		331.3
desipramine		266.4

diclofenac		296.1
ibuprofen		206.3
paracetamol		151.2
papaverine		339.4
phtalic acid		166.1
procaine		236.3

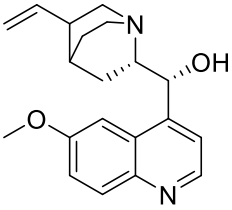
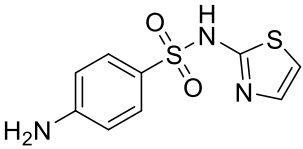
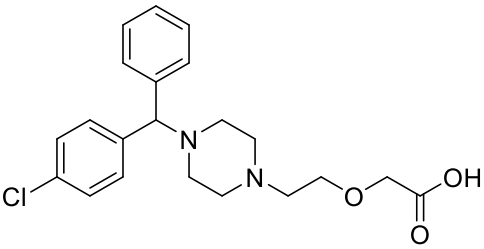
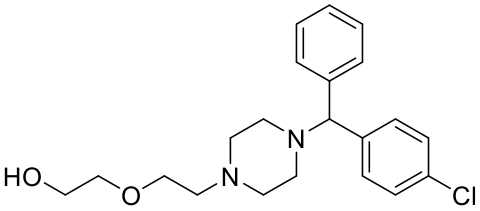
quinine		324.4
sulfathiazole		255.3
cetirizine		388.9
hydroxyzine		374.9

Table 10: Commercial standards employed in pKa determination by Sirius T3. Each compound is reported with its chemical structure and molecular weight (MW).

A selection of bases (B), acids (HA) and ampholytes (HX) was chosen as reference set (n = 16; see Table 10

Procaine hydrochloride (B), benzocaine (B), quinine hydrochloride (B), 1-naphtol (HA), paracetamol (HA), ibuprofen (HA), diclofenac sodium salt

(HA) and sulfathiazole (HX) were supplied by Food and Drug Department of University of Parma (Parma, Italy); desipramine (B), atenolol (B), papaverine (B), phtalic acid (HX), benzthiazide (HA), ciprofloxacin (HX), cetirizine (HX) and hydroxyzine (HX) were purchased from Sigma Aldrich (Milan, Italy). Proprietary molecules were kindly donated by Chiesi Farmaceutici S.p.A. (Parma, Italy).

Potassium chloride (KCl), potassium hydrogen phthalate (KHP) and potassium dihydrogen orthophosphate (KH_2PO_4), analytical grade, were purchased from Acef SpA (Piacenza, Italy).

Methanol (MeOH), dimethylsulfoxide (DMSO) and isopropanol HPLC grade were bought from Sigma Aldrich (Milan, Italy). De-ionized water was bi-distilled by using Sanyo Gallenkamp PLC (Loughborough, UK) equipment and then employed to prepare all reagents and solutions. Orion™ pH 7 buffer and KCl 3M solution were purchased by Thermo Fisher Scientific (Massachusetts, USA) and used to stock and refill the electrode respectively.

Ionic Strength Adjusted water (ISA water) was prepared every week by dissolving KCl in freshly bidistilled water to a final concentration of 0.15 M. Mid-range pH buffer consisted of a 2 g/L solution of KH_2PO_4 in ISA water to a final concentration of ~ 15 mM.

Methanol 80% (MeOH 80%) was made by dissolving 11.18 ± 0.1 g of KCl in 200 mL of bi-distilled water and making up the volume to 1 L with MeOH. Titrisol® ampoules for making HCl and KOH titrants were provided by Merck KGaA (Darmstadt, Germany). The content of an ampoule was diluted to 2 L with bi-distilled water to reach a final concentration of 0.5 M for both titrants and used up to three months.

Isopropanol was mixed with bi-distilled water to a concentration of 50% w/w and used to clean the probes when needed.

1a Apparatus

pKa measurements were performed by using Sirius GLpKa™ coupled with D-PAS module and the upgraded version of that instrument, which is called

Sirius T3. Both instruments were purchased by Sirius Analytical Ltd (Forest Row, UK).

GLpKa™: GLpKa™ system is an accurate potentiometric titrator integrated with the software package RefinementPro 2. The hardware module consists of a moving arm with the assay probes attached that are a glass combined pH electrode (Ag-AgCl by Sirius), a UV dip probe, an overhead stirrer, a tube for dispensing inert gas (*i.e.* N_2) and a temperature sensor. The probe set allows working in 10 mL sample volumes.
A fluidic system dispense reagents during the assays and includes five reservoirs (see Figure 31*a*) for ISA water, acid (HCl), base (KOH), co-solvent (MeOH 80%) and partition solvent (*n*-octanol saturated with ISA water). ISA water dispenser is fitted with a 25 mL syringe, which can accurately dispense water with a minimum increment volume of 2.1 µL. The other dispensers are fitted with 5 mL syringes, which can accurately dispense volumes with a minimum increment volume of 420 nL. The reagents are pipetted into the sample through a MultiTip which contains five narrow polyimide clad quartz capillary tubes. An autosampler module interfaced with the titrator offers a worktable with four 48-position vial trays. The worktable is connected to a thermostatic bath to ensure temperature control within 25 ± 0.5 °C.
Before running pH-metric assay, a strong acid-strong base titration was performed to calibrate the pH-metric system as detailed in **2b Calibration of the combined glass electrode** section.

D-PAS: spectrophotometric pKa were determined by using the Dip Probe Absorption Spectroscopy technique developed in the 1990s [38, 142, 143, 144] and implemented by Sirius in the D-PAS™ module.
That consists of a bi-furcated fibre optics dip probe with 1 cm optical pathlength, a UV-light source (deuterium pulsed lamp) and a photodiode array (PDA) detector (see Figure 31*b*) and is supposed to be interfaced with GLpKA™ titrator. The light was directed into the sample vessel and

absorption spectra of the solution were recorded for each pH point of potentiometric titration between 200 and 700 nm of wavelength.
A full spectrum was collected within 25 to 50 ms and an average of 10 to 20 spectra was performed at every titration point to minimize noise. A blank spectrum in pure water was recorded to set the scan time.

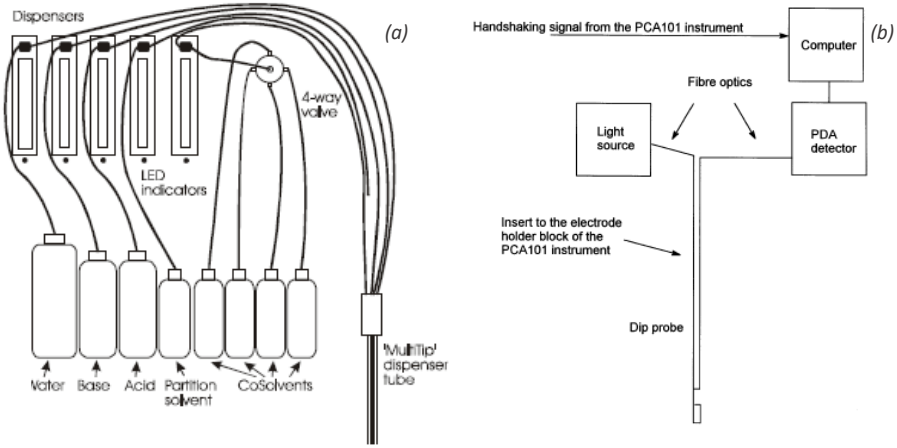


Figure 31: GLpKa (a) dispensing module and (b) optical set up utilized in D-PAS module [38].

T3 instrument: T3™ was purchased to replace the obsolete GLpKA™ during the first year of the research project. Sirius T3™ involves the same components described for the previous instrument with the main advantage of miniaturization, thus allowing to perform assays into 2.5 mL glass vials. In that sense moving arm includes a glass combined micro-pH electrode and a miniaturised bifurcated UV-optic dip probe with a 1 cm optical path length.
Dispenser module houses the precision micro-dispensers and also contains a UV-Vis spectrometer and light source connected to the dip probe; the sample position is temperature-controlled with a Peltier device at 25 ± 0.5 °C.

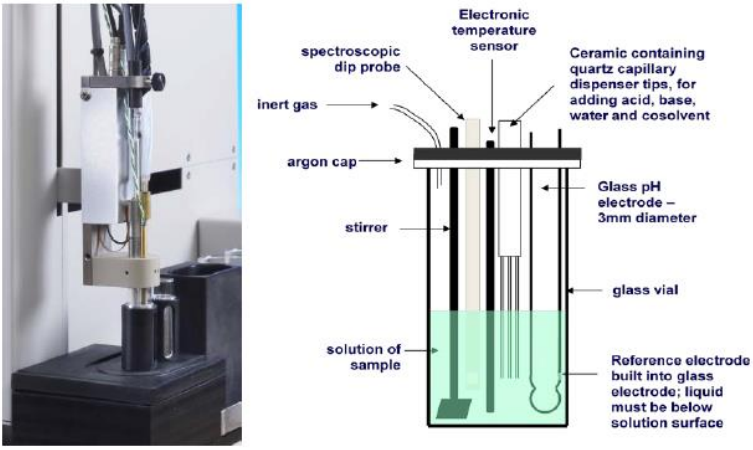


Figure 32: The probe set of Sirius T3.

3.1.2.2 Methods

2a Potentiometric pKa determination

The four-parameter procedure was used for electrode standardisation. A detailed description of the method was given in **2c pKa determination** paragraph.

Briefly, an HCl solution of known concentration was titrated with KOH -that was standardized by three repeated titrations against the primary standard potassium hydrogen phthalate at 25 ± 0.5 °C, at *I* = 0.15 M with KCl, under nitrogen atmosphere, in the pH interval 2–12. The operational pH reading was converted into the concentration p_cH values by the multi-parametric Equation 6:

$$pH = \alpha + S \cdot p_cH + j_H \cdot [H^+] + j_{OH} \cdot K_W/[H^+]$$

Equation 6

The four terms were determined by a weighed non-linear least squares procedure so that the intercept parameter α in aqueous solution mainly

corresponds to the negative logarithm of the activity coefficient of H^+ at the working temperature and ionic strength *I*.
The j_H term corrects pH readings for the non-linear pH response due to liquid junction and asymmetry potentials in low-pH solutions; on the other hand the j_{OH} term corrects for high-pH non-linear effect.
Factor S accounts for deviation of glass electrode from the Nernstian slope and K_w is the auto-dissociation constant of water. The actual calibration parameters of the semi-micro (for GLpKa) and micro- (for T3) combined glass electrode used in this research project were applied for the calculation of pKa and p_sKa values.

2b pH-metric titrations in aqueous media

A suitable amount of the analyte was weighed to obtain 10 mL of 1 mM - 10 mM aqueous solution.
The sample was titrated alkalimetrically (from low to high pH) or acidimetrically (from high to low pH) according to the solubility and the nature of the ionisable site(s) of the compound. The operative pH range was chosen to include pKa ± 2 log unit intervals.
The initial estimates of pKa values were obtained from Bjerrum difference plots and were then refined by a weighed non-linear least-squares procedure described in the **2c pKa determination** paragraph of the present work, by using RefinementPro v. 2.2.6.6 software by Sirius Analytical.
Shortly, seed values were iteratively refined to minimize the sum of weighed squares of residuals (see Equation 8):

$$S = \sum_{i}^{N_0} \frac{(pH_{i\,obs} - pH_{i\,calc})^2}{\sigma_i^2(pH)}$$

Equation 8

where N_0 is the number of experimental pH measures, σi^2 is the variance estimated for measured $pH_{i\ obs}$. The model equation $pH_{i\ calc}$ is a function of pKa and independent variables.
At the end of each iterative cycle, progress of refinement was expressed as Goodness Of Fit (GOF):

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

Equation 10

where Nr is the number of refined parameters (*i.e.* pKas). Convergence of non-linear equation led to GOF values close to 1 and was reached when solutions from two consecutive iterative cycles returned $\Delta GOF < 10^{-5}$ or when all refined parameters showed a calculated shift $< 10^{-4}$.
For each molecule three to five titrations were performed and the average pKa values along with the standard deviations were calculated. When allowed by the stability of the analyte in the sample solution, three titrations were repeatedly performed in the same vial to save substance.
Titrations on the T3™ instrument followed the same protocol, but the sample volume was reduced to 1.5 mL; an updated version of the software, called SiriusT3Refine v. 1.1.3.0, was employed for data refinement.
Titration in purely aqueous medium was conducted for 10 over 14 commercial standards (*i.e.* procaine, quinine hydrochloride, desipramine, atenolol, benzocaine, phtalic acid, 1-naphtol, paracetamol, ciprofloxacin and sulfathiazole).

2c pH-metric titrations in co-solvent/aqueous media

1 to 10 mM semi-aqueous solutions of the samples, containing 10–60 wt% methanol, were titrated under the same experimental conditions as in aqueous titrations. For each analyte 3 to 5 titrations were performed varying MeOH percentages and the Yasuda–Shedlovsky procedure was then applied to extrapolate the aqueous pKa values (*i.e.* p_sKa at 0% MeOH).

The present approach was applied to sparingly soluble compounds (*i.e.* papaverine hydrochloride, benzthiazide, ibuprofen and diclofenac sodium salt within the standard set and all the proprietary NCEs).

2d UV-metric pKa determination

As regards D-PAS™ set up, all experiments were performed in 10 mL solutions of 0.15 M KCl under nitrogen atmosphere at 25 ± 0.5 °C using 0.5 M HCl or standardized 0.5 M KOH titrants as reported for pH-metric experiments.
Sample solution were prepared by spiking 10 mM stock solution of the analyte in DMSO to reach a final concentration of 1-10 µM. The only exception was represented by ciprofloxacin: being insoluble in DMSO, a 10 mM solution in water at pH 2 was freshly prepared and spiked in the sample vial for the analysis.
Spectral data were recorded in the region of 200–700 nm after each pH measurement. The pH change per titrant addition was limited to about 0.2 pH units. Typically, 20 to 30 pH readings and absorption spectra were collected from each titration. pH stabilisation might require relatively long time in the unbuffered region and that would have negatively impacted on the throughput and quality of the assay. Hence when the analyte was expected to show pKa within 3 and 10, mid-range pH buffer was added to the sample in advance to reach a final concentration of 250 µM.
Before running spectrophotometric assays two spectra were acquired: (ii) a dark spectrum (*i.e.* a spectrum collected when deuterium lamp was off) that was assumed equal to noise due to the detector; (ii) a reference spectrum in a solution matrix that had a composition as close as possible to that of the sample.
Those were taken into account to return absorbance of the sample (A_λ) at each wavelength:

$$A_\lambda = \log_{10}\left(\frac{I_{ref} - I_{dark}}{I_{sample} - I_{dark}}\right)$$

Equation 72

with I_{ref}, I_{dark} and I_{sample} representing the intensity of the absorbance signal in the reference, dark and sample spectrum respectively.
In the case of Sirius T3, the abovementioned procedure was adopted except for lower operative volumes (1.5 instead of 10 mL).
For extremely insoluble compounds, titrations were performed in co-solvent/aqueous mixtures containing 10-30 wt% MeOH and Yasuda-Shedlovsky extrapolation was employed to extrapolate aqueous UV-pKas.

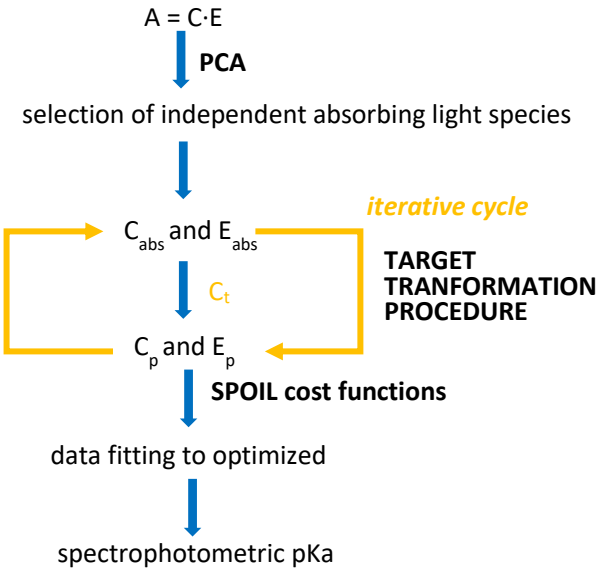


Figure 33: Scheme of Target Factor Analysis to deconvolute UV spectral data.

At least three UV-metric experiments were performed for each drug sample and data refinement was carried out by mean of Target Factor Analysis as fully reported in **3a Target Factor Analysis (TFA) method** paragraph of the present thesis. A scheme of the procedure is reported in Figure 33: to sum up, absorbance data from multi-wavelength spectra were cast into a matrix (A) regarded as the product of concentration-pH matrix (C) and absorption coefficients matrix (E). By applying PCA analysis the number of independent light absorbance species was defined (*i.e.* number

of pKas + 1) and abstract solutions were proposed for the concentration (C_{abs}) and the absorptivity matrix (E_{abs}).
Those solutions were subsequently converted in ones with physical significance and iteratively refined until the best fitting between the model and actual data was reached.
Effectiveness of refinement was evaluated through a SPOIL function (see Equation 20): target factor analysis converged when such function showed a value < 3.0. Then the model was regarded suitable for data fitting and, finally, for calculation of pKas.

3.1.2.3 Data analysis

Correlation studies between experimental and predicted values were performed for proprietary NCEs: plots and calculation were obtained with Microsoft Office Excel 2013 (Microsoft Corp., USA).

3.1.3 Results and discussion

Commercial standards

Compound	pKa $_{T3\ reference}$	pKa $_{GLpKA}$ ± SE	ΔpKa $_{T3\text{-}GLpKA}$
procaine	2.29	2.24 ± 0.03	0.05
	9.04	8.92 ± 0.01	0.12
quinine	4.24	4.24 ± 0.02	0.00
	8.55	8.48 ± 0.02	0.08
desipramine	10.08	9.98 ± 0.07	0.10
atenolol	9.54	9.28 ± 0.01	0.26
benzocaine	2.39	2.25 ± 0.04	0.14
phtalic acid	2.49	2.25 ± 0.01	0.24

	4.81	4.79 ± 0.01	0.02
1-naphtol	9.18	9.07 ± 0.01	0.11
paracetamol	9.52	9.39 ± 0.01	0.13
ciprofloxacin	6.23	6.14 ± 0.01	0.09
	8.58	8.42 ± 0.01	0.16
sulfathiazole	7.17	7.15 ± 0.04	0.02
	2.12	n/d	n/d

Table 11: Potentiometric pKa in aqueous medium reported with standard error (SE) and bias (ΔpKa). In-house determined values are compared to reference values. n/d not determined.

Both potentiometric and spectrophotometric methods were tested on commercial standards to ensure the reliability of the system (data shown in Table 11- Table 13) and highlight experimental limits of those approaches. Reference values were provided by Sirius Analytical (Forest Row, UK) and consisted of in-house experimental measurements by using Sirius T3^{TM}. Aqueous assays were run without any technical issue, showing a good agreement with reference values:

$$pKa_{GLpKa} = 0.99(\pm 0.01)pKa_{T3\ reference} - 0.79(\pm 0.06)$$

$$n = 14,\ s = 0.08,\ F = 17713,\ R^2 = 0.999$$

Equation 73

However, the assessment of extreme pK_a (e.g. sulfathiazole acidic pK_a) revealed more problematic, since an accurate determination was precluded by the operative pH range of the glass electrode. As regards data processing, fitting experimental values to the theoretical curves was reasonably straightforward, especially for monoprotic compounds, and the highest bias in pKa values was found at ± 0.26 log unit (see Table 11), in line with the observed bias (*i.e.* ΔpKa between reference and experimental value, which was assumed acceptable when within 0.2 log units).

Compound	Potentiometric pKa ± SE	UV-metric pKa ± SE	ΔpKa_{pot-UV}
benzocaine	2.25 ± 0.04	2.33 ± 0.01	-0.08
quinine	4.24 ± 0.02	4.21 ± 0.01	0.03
	8.48 ± 0.02	8.56 ± 0.01	-0.08
phtalic acid	2.49 ± 0.01	2.60 ± 0.01	-0.11
	4.84 ± 0.01	4.87 ± 0.01	-0.03
paracetamol	9.39 ± 0.01	9.45 ± 0.03	-0.04
ciprofloxacin	6.14 ± 0.01	6.25 ± 0.04	-0.11
	8.42 ± 0.01	8.46 ± 0.04	-0.04
sulfathiazole	n/d	2.04 ± 0.03	n/d
	7.15 ± 0.04	7.31 ± 0.02	-0.16

Table 12: Potentiometric and spectrophotometric pKas reported with respective standard error (SE) and bias (ΔpKa).

As expected, UV-metric titrations allowed to reduce drug concentrations of one order of magnitude reaching the lowest limit at around 0.1 µM and to access pK_a values at the extremes of pH range with a significant increase in accuracy and precision.

Experimental pKas assessed by both potentiometry and spectrophotometry showed no significant difference within each other, with the highest bias in pKa values of ± 0.16 log unit, in line with the observed experimental error (see Table 12).

Compound	$pKa_{T3\ reference}$	pKa_{GLpKA} ± SE	$\Delta pKa_{T3-GLpKA}$	% MeOH
papaverine	6.39	6.66 ± 0.06	-0.27	4.5 - 15
benzthiazide	6.67	6.74 ± 0.09	-0.07	12.6 - 50.5

	9.21	9.30 ± 0.04	-0.1	
ibuprofen	4.35	4.45 ± 0.03	-0.1	18.9 – 38.9
diclofenac	4.03	4.19 ± 0.03	-0.16	24.5 - 34.9

Table 13: Potentiometric pKa in co-solvent/water mixtures, reported with standard error (SE) and % co-solvent used in the assays. Experimental data are compared to reference values. In-house determined values are compared to reference values. n/d not determined. Co-solvent assays returned good data in comparison to references.

$$pKa_{GLpKa} = 0.99(\pm 0.02)pKa_{T3\ reference} - 0.18(\pm 0.14)$$

$$n = 5,\ s = 0.09,\ F = 2033,\ R^2 = 0.999$$

Equation 74

However, those experiments required an appropriate choice of methanol contents to meet solubility needs as well as to minimize the negative influence of the organic solvent on the electrode performance. In fact, extrapolation deviated from linearity at high co-solvent percentages (*i.e.* > 50 wt% MeOH).

In order to estimate the accuracy of prediction, a regression model between computed (ACDLab Percepta 2012 v. 8, Advanced Chemistry Development, Inc., Toronto, Ontario, Canada) and experimental values were calculated for commercial standards by using Microsoft Office Excel 2013:

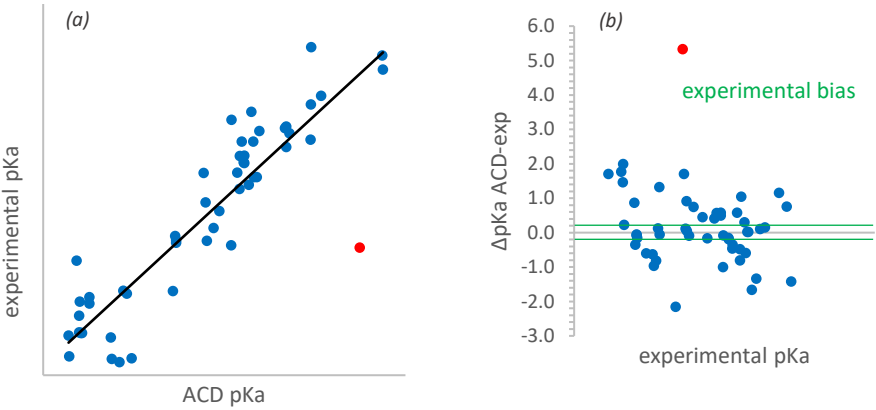
$$pKa_{experimental} = 0.97(\pm 0.04)pKa_{ACD} + 0.02(\pm 0.25)$$

$$n = 19,\ s = 0.42,\ F = 756,\ R^2 = 0.978$$

Equation 75

Equation 75 proved accuracy in prediction to be remarkable for standard drugs. This result was quite expectable, considering the low molecular complexity of such molecules (*i.e.* MW < 400, max. n = 2 protonation sites).

Proprietary NCEs



Graph 1: (a) Correlation between experimental and predicted pKa values; (b) residuals graph from experimental versus computed pKas. One evident outlier is marked in red.

The same experimental approach was subsequently employed for proprietary NCEs to test the compliance of such methods to actual research needs. Candidate chemotypes were generally characterised by a higher molecular complexity than that of the commercial standards as well as by a higher number of protonation sites, whose micro-dissociation constants could be hardly assigned to, even with the support of predicted values. For most of those drug candidates, only macroscopic pKas were required for practical reasons and spectrophotometric and potentiometric approaches were integrated to return macroconstants (n = 52).

Computed pKas (ACDLab Percepta 2012 v. 8, Advanced Chemistry Development, Inc., Toronto, Ontario, Canada) were taken as reference and correlation between experimental (experimental pKa) and predicted (ACD pKa values (see Graph 1*a*) was performed:

$$pKa_{experimental} = 0.90(\pm 0.06)pKa_{ACD} + 0.45(\pm 0.44)$$

$$n = 52,\ s = 1.13,\ F = 196,\ R^2 = 0.797$$

Equation 76

Although satisfactory for commercial standards, when it came to NCEs the value of the correlation coefficient was nearly acceptable (see Equation 76). Furthermore, the distance of some points from the regression line allowed to speculate a poor predictive power. In particular, one single pKa value displayed a deviation of over 5 log unit that was nearly 10-fold the mean bias observed for the other pKas (*i.e.* 0.7 log unit).
Excluding that outlier from the regression model, goodness of fit between experimental and computed data slightly increased, but was still significantly below that of commercial standards:

$$pKa_{experimental} = 0.97(\pm 0.05)pKa_{ACD} + 0.08(\pm 0.36)$$

$$n = 51,\ s = 0.90,\ F = 338,\ R^2 = 0.873$$

Equation 77

To further investigate predictivity of the regression model for proprietary molecules, the difference between predicted and experimental pKas (ΔpKa ACD-exp) was considered as a function of the experimental values to unveil the predictivity error.
As displayed in the residuals plot in Graph 1*b*, it turned out that computational method failed in the accurate prediction of pKa values for NCEs: in fact, 75% of the experimental bias exceeded the acceptable uncertainty of 0.2 log units that was assumed as reference. That could be regarded as a consequence of the increased molecular complexity displayed by proprietary compounds in comparison to the one of commercially available standards.
To strengthen this evidence the average bias was calculated for mono, di- and tri-protic NCEs:

$$bias = \Delta pKa = \left| pKa_{ACD} - pKa_{experimental} \right|$$

Equation 78

It was found that the uncertainty in prediction switching from a single to a multiprotic molecule nearly doubled (*i.e.* average bias of 0.42 for n = 1, 0.84 for n = 2 and 0.88 for n = 3 protonation sites respectively).
Comparison between regression models for commercial standards and for NCEs provided a concrete example of lack of accuracy of computational approaches for current drug discovery needs, which was discussed in the introduction of this thesis.
In that sense, experimental values returned by this study will be used to build an internal database and to train ACD Lab Percepta 2012 v.8 software in order to obtain more reliable predictions.

3.2 Investigation of macro- and micro-dissociation equilibria by potentiometric and UV-spectrophotometric approaches

3.2.1 Aim of the work

After practicing experimental methods for pKa determination, UV-spectrophotometry and potentiometry were combined to investigate micro-dissociation equilibria of amphoteric drugs to provide information about intramolecular interactions that could affect both solubility and lipophilicity (see **3c** and **3f** and sections of paragraph **1.2.2.3 Experimental approach: spectrophotometric method** for the entire discussion upon this topic).
As to ensure the feasibility and suitability of the deductive approach to complex ionization equilibria, the case study of cetirizine [6] was analysed. A similar protocol, that integrated spectrophotometric and potentiometric assays, was applied to a model NCE to assess site-specific dissociation constants.

3.2.2 Experimental part

3.2.2.1 Apparatus

The set-up adopted for potentiometric as well as UV-metric assays was the same reported in the **1a Apparatus** paragraph of the previous section. In particular, GLpKA^{TM} instrument and D-PAS^{TM} module were used to characterise cetirizine and NCE micro-speciation equilibria.

3.2.2.2 Methods

Potentiometric method was employed to determine macro- and microscopic pKa values of cetirizine, by means of deductive approach (see **3d Determination of micro- and macro-constants by spectrophotometric approach** section of the Introduction). That proved useful to overcome the

complexity related to the co-existence of multiple ionisable moieties: according to this approach, one or more protonation sites within the parent compound were masked by chemical derivatisation, and the analogue was used to assess at least one ionization micro-constant.
In the present case, cetirizine and its metabolite hydroxyzine served this purpose, as depicted in Figure 35.

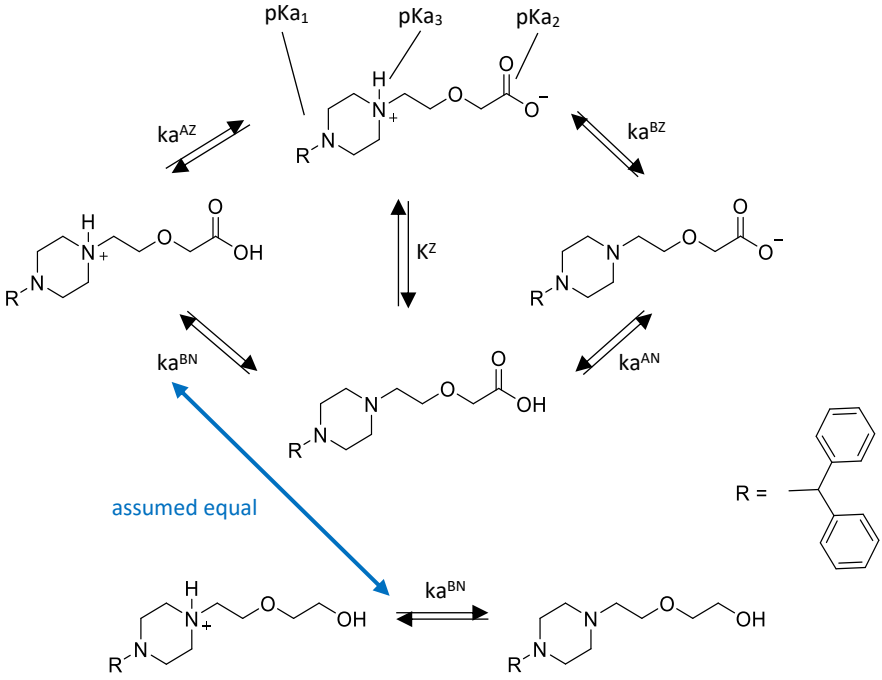


Figure 34: Cetirizine and hydroxizine dissociation equilibria as reported in Pagliara's paper [6].

Cetirizine harbours two basic sites on the piperazine ring (pKa_1 = 1.52 and pKa_3 = 8.27) and an acidic moiety represented by the carboxylic acid (pKa_2 = 2.92) (values from [145]).
Macro-dissociation involving the weakest base could be regarded as negligible above pH 3 thus simplifying experimental investigation. However, overall characterisation of micro-dissociation equilibria of cetirizine above pH 3 required to know at least one micro-pKa that was

chosen to be pka^{BN} (*i.e.* the microdissociation constant of the basic group when the acidic group is not ionized). That was assumed equal to the basic macro-pKa of hydroxyzine, the ethyl-ether analogue for cetirizine.
Due to the low aqueous solubility of cetirizine and hydroxyzine, potentiometric titrations were performed in different co-solvent/aqueous mixtures (n = 8, ranging from 10-40 wt% MeOH) to avoid precipitation of both analytes.
Spectrophotometric method was attempted to measure site-specific constant pKa_1 of cetirizine, taking advantage from the proximity to the aromatic rings.
However, absorption maximum was found in the region of solvent interference (∼220 nm) thus hampering collection of good quality data. Added to that, difference among the absorbance spectra of the neutral and ionized species (referring to the piperazine site) was insufficient for the application of UV-spectroscopic approach.

3.2.2.3 Calculation and data analysis

Calculation and plots were performed in Microsoft Office Excel 2013. Starting from macro-constants of cetirizine and hydroxizine, micro-pKa values and the tautomeric constant K^Z (Table 14) for the former were calculated according to Adam's equations (Equation 23):

$$K_a^{acidic} = K_a^{A^Z} + K_a^{B^N} \qquad (23.1)$$

$$\frac{1}{K_a^{basic}} = \frac{1}{K_a^{B^Z}} + \frac{1}{K_a^{A^N}} \qquad (23.2)$$

$$K^Z = \frac{K_a^{A^Z}}{K_a^{B^N}} = \frac{K_a^{A^N}}{K_a^{B^Z}} \qquad (23.3)$$

Equation 23

3.2.3 Results and discussion

Application to cetirizine

(a)

			VALUES DETERMINED		
			by potentiometry		through pka^{BN}
MACRO	(B)	pKa_1	2.19	0.06	
	(HA)	pKa_2	2.93	0.03	
	(B)	pKa_3	8.00	0.02	
MICRO		pka^{AZ}			2.93
		pka^{BZ}			8.00
		pka^{BN}	7.49		
		pka_A^N			3.41
Log K^Z					4.56

(b)

			VALUES DETERMINED		
			by potentiometry		through pka^{BN}
MACRO	(B)	pKa_1	n/d		
	(HA)	pKa_2	3.09	0.1	
	(B)	pKa_3	8.01	0.02	
MICRO		pka^{AZ}			3.09
		pka^{BZ}			7.85
		pka^{BN}	7.5	0.02	
		pka_A^N			3.44
Log K^Z					4.41

Table 14: (a) Reference values from Pagliara's work compared to (b) experimental values.

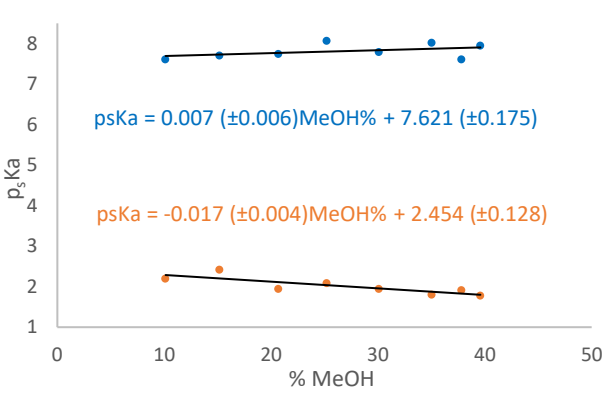
Experimental results were in good agreement with literature ones. In order to gain a better fitting, potentiometric data points were modelled by using a 2-pKa calculation. That was allowed since methanol presence weakened

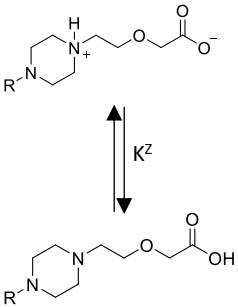
both acidic and basic sites, leading the lowest pKa values out of the operative pH range of the electrode.

Decreasing dielectric constant of the medium affected equilibria that led to an increase in the global charge at a larger extent in comparison to the ones where both sides of the reaction showed the same total charge. As reported by Pagliara et al., dissociation of the acidic site was found to be more influenced by co-solvent than the basic moiety, hence allowing unambiguous attribution of the pKa values to the sites.

However, it is worthy noticing such difference was so remarkable because of the large ΔpKa (∼5 log units) between the two groups and the delimited distribution of the charges (*i.e.* low delocalization) on cetirizine molecule.

The relatively flat slopes of p_sKa vs MeOH% curves, reported in Graph 2, suggested the tautomeric equilibrium to highly favour the zwitterionic over the neutral form around isoelectric pH; such speculation was confirmed by the high value of Log K^Z (4.41) that implies the zwitterion to be predominant in determining lipophilicity and related properties of cetirizine.





Graph 2: p_sKa versus co-solvent percentage in potentiometric assays for cetirizine. Data referring to the basic site are coloured in blue and data from the acidic sites are plotted in orange.

Taking advantage from that evidence and from the large ΔpKa, dissociation equilibrium of cetirizine could be further simplified to an un-branched system where only macro-constants are required to profile the actual speciation of the analyte in solution (*i.e.* $pka^{AZ} = pKa_2$ and $pka^{BZ} = pKa_3$):

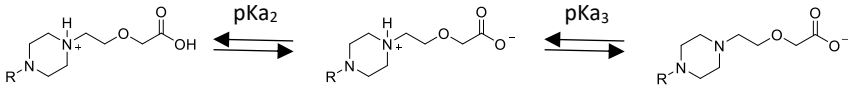


Figure 35: Simplification of dissociation equilibria of cetirizine.

Application to a proprietary NCE

The deductive approach was applied to profile the overall macro- and micro-dissociation equilibria of a proprietary compound displaying four ionisable groups. To this aim an analogue was synthesized, in which one site had been properly masked, and both spectrophotometric and potentiometric methods were performed. Through the comprehensive analysis of data from the parent compound and from the analogue, of the pKas of molecular fragments and of the estimated tautomeric ratio constant (K^Z), micro-pKas were reasonably assigned to each site. Those data were subsequently employed for salt-screening assays.

3.2.4 Conclusions

Dissociation equilibria were investigated by using potentiometric and UV-spectrophotometric methods, both implemented in GLpKA™ software-hardware package by Sirius Analytical (Forest Row, UK) and its updated version T3™.

After determining pKas of a series of diverse standards (n = 14), the same techniques were applied to NCEs at discovery or early development stage (n = 31). Experimental macro-constants were correlated to computed values by ACD Lab Percepta 2012 v. 8 (Advanced Chemistry Development, Inc., Toronto, Ontario, Canada) for both standards and proprietary

molecules. While prediction proved accurate for commercial drugs (R^2 = 0.978), goodness of computation remarkably decreased for NCEs (R^2 = 0.797 and R^2 = 0.873 including and excluding the outlier respectively). Moreover, analysis of the residuals (ΔpKa ACD-exp for proprietary molecules) proved the software to fail in most of the cases in which more than two ionisable sites were displayed. That proved the need for re-training the system by using experimental values from discovery compounds.

Micro-dissociation equilibria were investigated by using deductive approach and the case study of cetirizine was analysed as reported in literature. Then, an integrated spectrophotometric-potentiometric method was applied to a proprietary NCE to achieve complete micro-speciation.

Although elegant, deductive approach should be applied carefully: assuming one micro-constant of the analyte to be roughly similar to a macro-constant of a chemical analogue, choice of masking group affects uncertainty of such approximation. Hence synthesis of the analogue should consider electrostatic and steric proprieties of the masking moiety to minimize the impact on the overall properties of the parent structure.

Furthermore, absence of cooperativity within the protonation sites is the main requirement for feasibility and should be verified before adopting this approach.

In the context of industrial needs, it was also concluded that the investigation of micro-dissociation equilibria should be attempted only for the most interesting compounds entering the pre-development phase, as it was both time- and substance-consuming (about 20 mg). Nonetheless, in-depth characterisation of microscopic equilibria should provide complementary information to improve either formulation or pharmacokinetic properties of drug-like molecules.

4. Investigation of acid-base equilibria in non-aqueous systems

4.1 Aim of the work

Dissociation equilibria are commonly assayed in aqueous systems that represent the physiological condition of animal and human body, thus returning useful information for optimization and development of drug candidates.

However, studying acid-base equilibria in non-aqueous conditions may be useful for special needs, such as mimicking non-water based formulations (e.g. subcutaneous vehicles, skin patches, ointments or suppositories) to unveil the acid-base behaviour of the API in the final product.

In this part of the Ph.D. project, a UV-metric method was set up to investigate acid-base equilibria of the two commercial acids reported in Figure 36 in non-aqueous solvent mixture. Moreover, the pKas of the analytes were assessed in purely aqueous as well as in organic solvent/aqueous mixtures by pH-metric approach supplied by Sirius T3™.

The aim of these assays was to evaluate if relative acidities determined in a polar medium (*i.e.* ISA water, 0.15M KCl; $\varepsilon_r = 80.1$) decreased or even reversed in less polar or practically apolar (*i.e.* dielectric constant $\varepsilon_r << 15$) solvents.

1 *2*

Figure 36: Commercial acids assayed in organic solvent mixtures. Both acids display a chromophore (phenyl ring); added to that acid 1 has a hydrophilic residue (R^1) while acid 2 exhibits both an hydrophilic (R^2) and a hydrophobic group (R^3).

4.2 Experimental part

4.2.1 Materials

The two acids and the non-aqueous fluid were kindly donated by Chiesi Farmaceutici S.p.A (Parma, Italy); triethylamine (TEA), acetone, acetonitrile (ACN) and absolute ethanol (EtOH) were all of LC grade and were purchased by Sigma Aldrich (Milan, Italy); dimethylsulfoxide (DMSO) and tetrahydrofuran (THF) were provided by Acef SpA (Piacenza, Italy); ultra-pure Millipore water (Darmstadt, Germany) was employed in sample preparation.
Ethanol 60% w/w, DMSO 60% w/w, THF 60% w/w, acetone 60 % w/w and ACN 60% w/w were prepared by dissolving 11.18 ± 0.1 g of KCl in 400 mL of bi-distilled water and making up the volume to 1 L with the solvent of choice, thus ensuring *I* = 0.15 M. For reagents used in potentiometric titration please refer to **3.1.2.1 Materials** section of the previous chapter.

4.2.2 Apparatus

UV-metric assays were carried out using a Multiskan GO Spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) in 10 mm NIR Rectangular Cell (Quartz SUPRASIL 300, light path 10 mm, cell volume 3.5 mL) by Perkin Elmer (Massachusetts, USA).
Sirius T3^{TM} (Sirius Analytical, Forest Row, UK) was employed for potentiometric titrations in ISA water and co-solvent/aqueous mixtures. Details of the set up and procedure were provided in paragraph **3.2.2.1 Apparatus**.

4.2.3 Methods

4.2.3.1 UV-spectrophotometric investigation of acid-base equilibria in non-aqueous fluid mixture

The two analytes were individually titrated as follows. 3 mM solutions of each acid was prepared in non-aqueous fluid mixture with an increasing

concentration of TEA (*i.e.* 0, 0.5, 0.75, 1, 2 and 5 eq) to simulate titration step-wise process.
Each solution was transferred in the quartz cuvette and assayed using the Multiskan GO Spectrophotometer collecting spectra from 240 to 300 nm.
Since both acids showed identical spectra, benzoic acid was employed as indicator. Its absorbance profile was monitored as reported for the two acids, but, due to its higher absorptivity, 0.75 mM solutions were employed.
Then samples of each standard acid (3 mM) in addition to the indicator (0.75 mM) were prepared, containing increasing amounts of basic titrant (0, 0.5, 0.75, 1 and 5 eq in comparison to the total concentration of the two acids) and absorbance variation around 279 nm was monitored to qualitatively evaluate ion pairing propensity of each acid in organic solvent mixture; that was finally used to estimate relative acidic strength for the two compounds.
Each spectrum was the average of three readings and the analysis of individual acids and mixtures was repeated on two consecutive days.

4.2.3.2 Potentiometric pKa determination in co-solvent/aqueous mixtures

1.5 mL of 1 mM to 5 mM aqueous solutions of the samples were titrated alkalimetrically (from low to high pH) as triplicates at 25 °C and seed pKa values were estimated from Bjerrum difference plots.
Experimental points from the three titrations were refined independently by a weighed non-linear least-squares procedure implemented in SiriusT3Refine v. 1.1.3.0 software by Sirius Analytical, as reported in section **2b pH-metric titrations in aqueous media** from chapter **3**. An average of the three replicates returned the results.
The same operative volume, temperature and concentrations were employed to assay the acids in different co-solvent/water mixtures at three different percentages for each solvent, ranging within 22.11% and 69.19%. p_sKa points were linearly interpolated thus returning an estimation of acidity trend in each organic solvent.

4.2.4 Data analysis

Calculation, regression analysis and graphs were performed in Microsoft Office Excel 2013.

4.3 Results and discussion

Spectrophotometric approach

Sirius T3™ potentiometric approach was not suitable for studying acid-base equilibria in non-aqueous solvents.

$$HA + S \underset{}{\overset{K_1}{\rightleftharpoons}} [HA\cdots S] \overset{K_2}{\rightleftharpoons} [A^-SH^+] \overset{K_3}{\rightleftharpoons} [A^-][SH^+] \overset{K_4}{\rightleftharpoons} A^- + SH^+$$

Figure 37: Scheme of dissociation process.

As summarized in Figure 4 (recalled above), in polar solvents (*e.g.* if S is water) dissociation of a weak acid occurs at an extent that free anions and cations are liberated into solution, thus pH could be easily monitored by using a combined glass electrode sensitive to $[H_3O^+]$ ions (*i.e.* glass electrode of Sirius T3). In such cases, dissociation equilibria is complete and dissociation constants can be determined by mean of proper equations:

$$HA + S \rightleftharpoons A^- + HS^+ \qquad \mathrm{Ka} = \frac{[\mathrm{A}^-][\mathrm{HS}^+]}{[\mathrm{HA}]}$$

Equation 79

On the other hand, if an acid is titrated with a base in apolar media (*i.e.* halogenated solvents) dissociation is incomplete and only ion-pairing equilibrium could be observed:

$$HA + B \rightleftharpoons HB^+A^- \qquad \mathrm{Ki.\,p.} = \frac{[\mathrm{HB}^+\mathrm{A}^-]}{[\mathrm{HA}][\mathrm{B}]}$$

Equation 80

In non-aqueous fluid mixture potentiometric approach would have needed time-consuming set up [146], including pre-treatment of the electrode with the organic medium and validation of the procedure by repeated strong acid-strong base titrations. Furthermore, the relatively high operative concentration of the analyte might have induced intermolecular association and conjugation mediated by hydrogen bonding, which are parasite reactions perturbing dissociation equilibria:

$$2HA \rightleftharpoons (HA)_2 \qquad \text{Kass.} = \frac{[(HA)_2]}{[HA]^2}$$

Equation 81

Finally, potentiometry required presence of a certain percentage of water for the sake of feasibility (*i.e.* presence of H_3O^+ ions) thus precluding the use of purely organic solvent. That also accounted for the use of co-solvent mixtures in the above described Sirius T3™ potentiometric approach.
On the other hand, spectrophotometric approach allowed to reduce operative concentration minimising the risk for side reactions; moreover, simultaneous titration of two acids in the same solution avoided the need to measure pH, which was an issue in an organic environment.
The starting point for the set up was the work by Kaljurand and Kutt [147, 148, 149, 150, 151] in which the authors reported self-consistent scales for acids and bases strength in organic solvents as acetonitrile, dichloroethane and tetrahydrofurane. The authors pointed out some preliminary conditions to be verified: (i) the analytes should exhibit remarkable difference between the absorbance spectra of ionized and unionized species (ΔAbs) in the wavelength range considered. Since the analytes did not fulfil that requirement, benzoic acid was chosen as ion-pairing indicator. Moreover, (ii) the acids should ideally exhibit different analytical wavelength (*i.e.* the wavelength at which difference between the ionized and neutral forms was maximized). (iii) Basic titrant should be carefully chosen for monitoring ion-pairing: it had to be transparent in the explored UV-range and should not be prone to homoconjugation reactions. Triethylamine (TEA) was employed in the current study since it proved not to affect the spectra of the pure components.

The acid and the indicator together in solution competed for triethylamine, generating two ion-pairing equilibria:

$$HA + TEA \rightleftharpoons A^{-}TEAH^{+} \qquad Ki.\,p.\,(HA) = \frac{[A^{-}TEAH^{+}]}{[HA][TEA]}$$

Equation 82

where HA was acid 1 or acid 2;

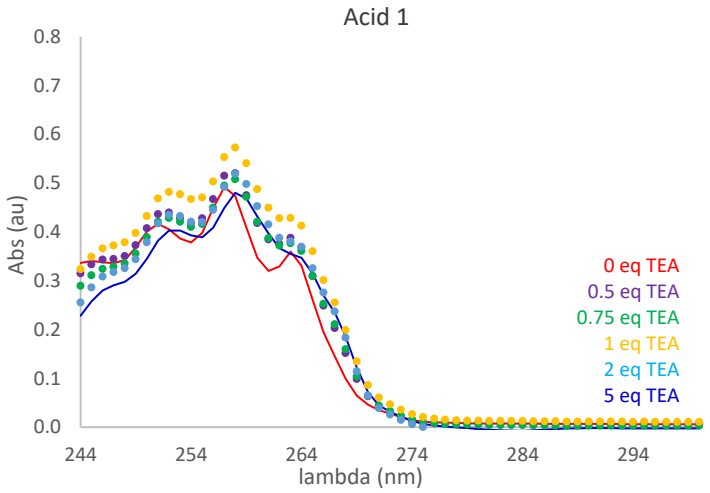
$$HInd + TEA \rightleftharpoons Ind^{-}\,TEAH^{+} \qquad Ki.\,p.\,(HInd) = \frac{[Ind^{-}TEAH^{+}]}{[HInd][TEA]}$$

Equation 83

where HInd referred to the benzoic acid (indicator).
Absorbance changes at 279 nm resulted from the ratio between the two ion-pairing constants:

$$\frac{Ki.\,p.\,(HA)}{Ki.\,p.\,(HInd)} = \frac{[A^{-}TEAH^{+}]}{[HA][TEA]} \cdot \frac{[HInd][TEA]}{[Ind^{-}TEAH^{+}]}$$

Equation 84



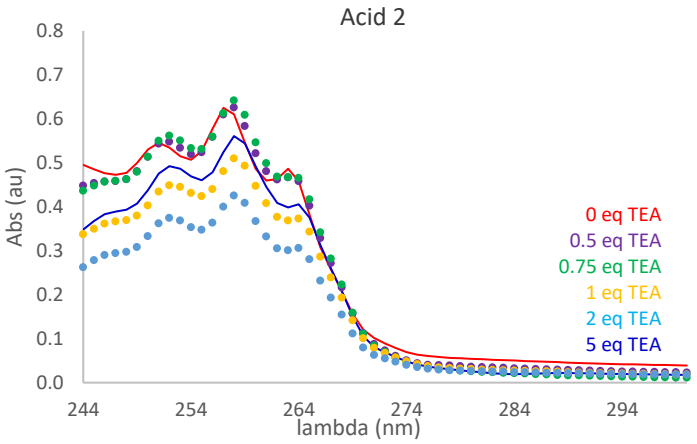


Figure 38: Absorbance profile of standard acids in the presence of different equivalents of TEA.

Assuming Ki.p. proportional to Ka of the acid, the higher the ratio, the stronger the acid that meant the lower the concentration of ion pair involving indicator in comparison to the one involving the acidic analyte.

Before testing mixtures, each acid of interest was analysed individually returning the spectra reported in Figure 37. Neither acid 1 nor acid 2 showed a remarkable difference between the absorbance profiles of the pure chemical components (*i.e.* the neutral and ionized form).

It should be noticed, moreover, that none of the analyte was responsive above 270 nm, that evidence supporting the use of benzoic acid as indicator. In fact, absorbance profile of benzoic acid showed a consistent and repeatable absorbance difference (ΔAbs) between unionized (Hind) and ionized form (Ind^-) which reached a relative maximum at 279 nm (see Figure 39*a*). At that wavelength, ΔAbs of both acids was negligible (see Figure 39*b*), so that ΔAbs could be exclusively ascribed to the ion-pairing equilibrium between the indicator and the titrant.

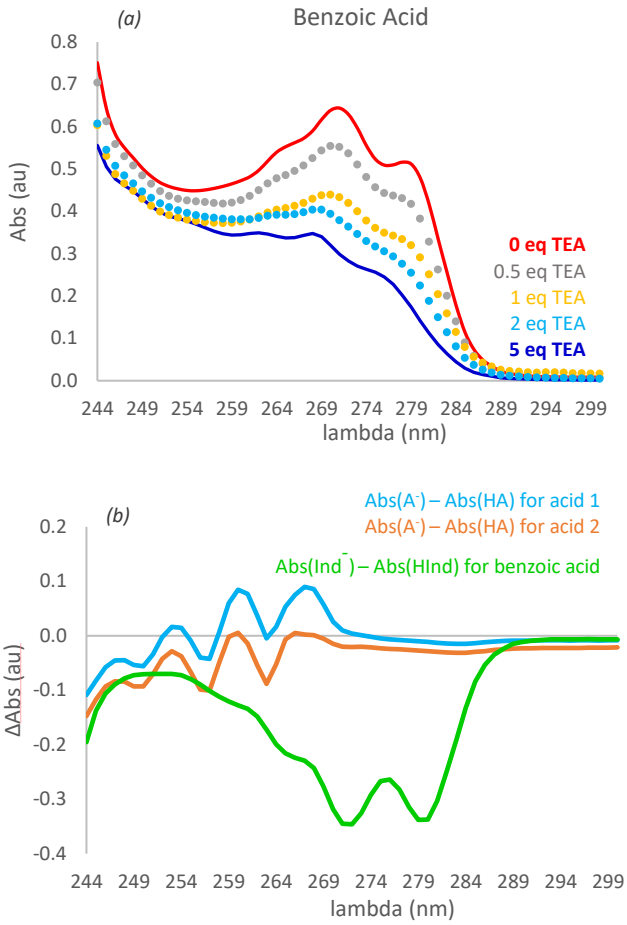


Figure 39: (a) Absorbance profile of benzoic acid in the presence of different equivalent concentration of TEA; (b) absorbance difference between neutral and ionized species of each acid plotted as a function of wavelength.

After collecting the spectra of the pure components, solutions of each acid in addition to the indicator were prepared, containing increasing amounts of basic titrant, and absorbance variation around 279 nm was monitored to

qualitatively evaluate ion-pairing propensity of each acid in non-aqueous fluid mixture.

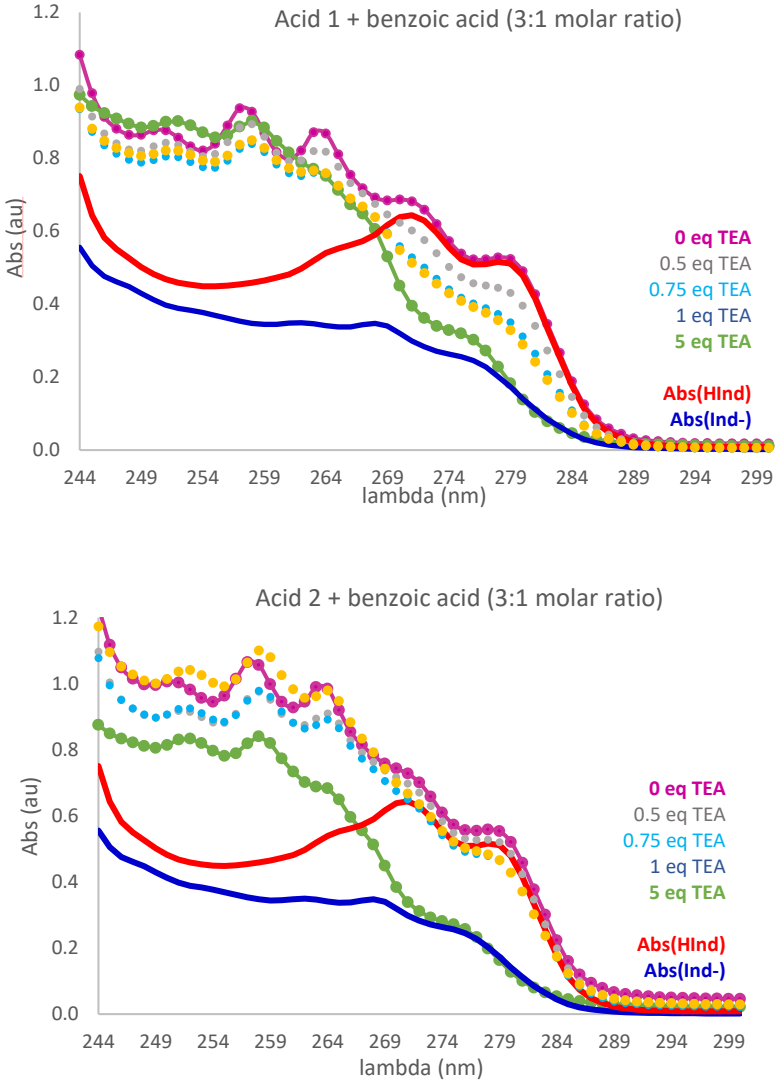
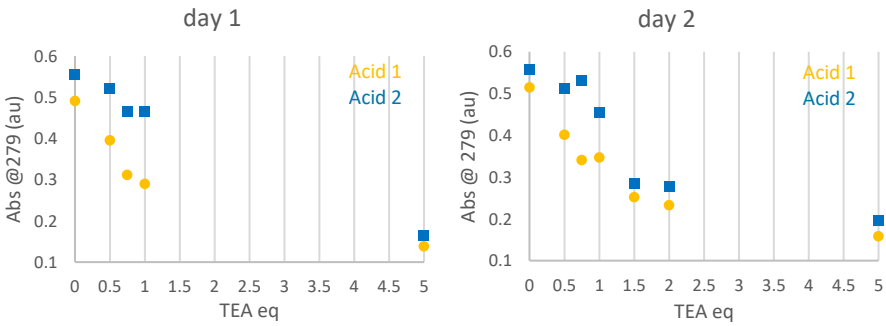


Figure 40: Absorbance profile of each acid and indicator mixture with different content of TEA.

Referring to Figure 40, acid 1 and benzoic acid mixture showed a gradual transition between Hind (higher Abs) and Ind^- (lower Abs) UV spectra, as TEA concentration increased. On the other hand, acid 2 and benzoic acid solution exhibited a sudden drop in absorbance signal only when molar excess of base was added. That suggested the indicator to interact with the base only when acid 2 had reached ion pairing equilibrium with the TEA.
For a more straightforward interpretation, further analysis focused on the absorbance at 279 nm (*i.e.* the lambda at which ΔAbs for benzoic acid reached a maximum) that can be regarded as proportional to the concentration of HInd. The absorbance of each acid mixture with the indicator was reported as function of the equivalent concentration of the titrant. Experimental points from acid 1 (reported as yellow dots) were always below data from acid 2 (reported as blue squares) confirming the latter to compete for TEA at a larger extent than the former.

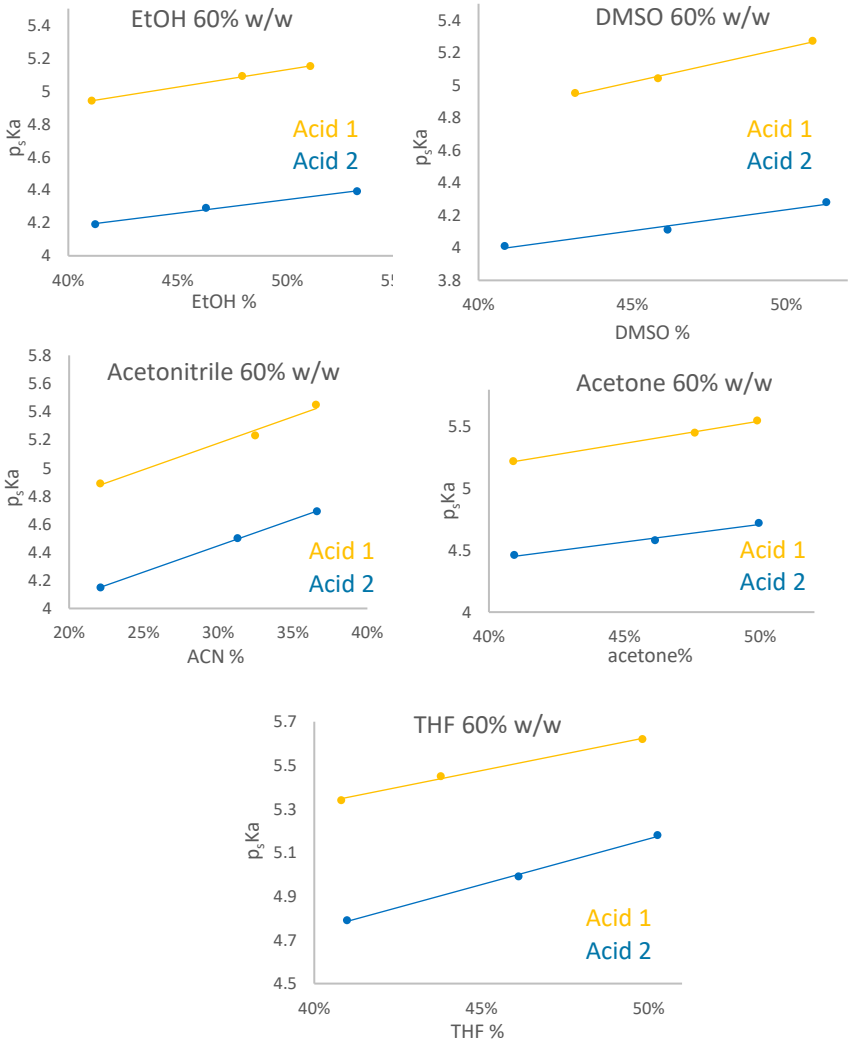


Graph 3: Absorbance as function of TEA concentration for acid 1 and acid 2, each in mixture with the indicator.

Potentiometric pKa determination in co-solvent/aqueous mixtures

Potentiometric pKas of the two commercial standards were measured in aqueous conditions resulting in acid 2 to be stronger than acid 1 (pKa = 3.20 ± 0.01 and 3.88 ± 0.01 respectively).

Then, the acids were assayed in different water/organic solvent mixtures at different percentages and the following linear trends were found by interpolating experimental points from each medium.



Graph 4: Linear interpolation of p$_s$Kas values in different organic solvent/water mixtures.

As the polarity of the medium decreased (*e.g.* higher percentages of organic solvent; solvent with lower dielectric constant), the apparent acidic strength of both standards decreased (*i.e.* higher p_sKa values), since ionization was less favoured in comparison to water. However, relative acidity of the compounds was confirmed at any percentage in any mixture tested, being p_sKa of acid 2 always lower than p_sKa of acid 1. Added to that, the trends defined by linear interpolation did not allow inferring an inversion of acidic strength relationship between the two acids.

4.4 Conclusions

In this section an investigation of acid-base equilibria of two commercial acids in non-aqueous solvents was attempted to mimic conditions of a non-water based formulation.

A dedicated UV-metric approach was employed to investigate relative ion pairing propensity of the analytes in a non-aqueous fluid mixture. For this purpose, mixtures of each acid and an UV-responding indicator (benzoic acid) were prepared employing increasing concentrations of the transparent titrant triethylamine (TEA). Absorbance profiles of each mixture were collected and compared, considering the absorbance at 279 nm related to the extent at which ion pairing between the analyte and the base occurred.

Approximating acidity strength with ion-pairing constant (Ki.p.) and evaluating spectrophotometric data from a qualitative perspective, it was found acid 2 to be stronger than acid 1.

Unfortunately, quantitative analysis was hampered by the characteristics of the absorbance data: the mixture spectra did not simply result from the additive contributions of the limiting forms (*i.e.* Abs of ionized form added to Abs of the unionized form equals Abs of the mixture at each concentration of TEA) hence the least-squares fitting of the linear combination of the two limiting forms described by Kutt [152] was not applicable.

According to literature, lack of linear combination could be ascribed to side reactions (e.g. homo- and hetero-conjugation) that would change the shape of the spectra. However, unexpected shapes were not observed and low operative concentration made parasite reactions unlikable to occur.
Thus, the lack of linear combination between spectra of the mixture and the ones of individual forms was ascribed to both intra- and inter-day oscillation of lamp source that involved erratic noise in the spectra.
A value of ion-pairing constant, however, would have been hard to relate to the acidic strengths of the analyte so the qualitative estimation of relative acidity reported in this work fit for purpose.
Furthermore, pKa of each acid was determined by using Sirius T3 potentiometric approach in ISA water (polar medium) and p_sKa were then assessed in several organic solvent/aqueous mixtures (low polar/apolar media).
Linear trends found in all systems confirmed the relative acidity observed in aqueous phase and, moreover, provided quantitative data to support results from spectrophotometric approach.

5. Lipophilicity assessment by potentiometric method

5.1. Assessment of log *P* by potentiometric approach

5.1.1 Aim of the work

The second part of the research project focused on lipophilicity assessment by using Sirius Analytical hardware-software packages (GLpKa^{TM} and T3^{TM}). The same set of literature compounds used for pKa assays was employed for the validation of potentiometric log *P* experiments in *n*-octanol/ISA water biphasic system. Then the method was transferred to ionisable NCEs at discovery and pre-development stage and advantages and limits of the technique were evaluated.
As reported for pKa experimental data, correlation of log *P* experimental results to computed values by ACD Lab Percepta 2012 v. 8 (Advanced Chemistry Development, Inc., Toronto, Ontario, Canada) was performed for both standards and drug candidates to evaluate reliability of the prediction outputs.

5.1.2 Experimental part

5.1.2.1 Materials

Commercial standards reported in paragraph **3.1.2.1 Materials** of chapter **3**. were also employed for potentiometric log *P* assays; proprietary molecules were kindly donated by Chiesi Farmaceutici S.p.A. (Parma, Italy). In addition to the reagents described in **3.1.2.1 Materials**, *n*-octanol (*n*-oct) was purchased from Sigma Aldrich (Missouri, USA). Partition solvent was prepared adding 25 mL of ISA water in a 1 L bottle and pouring *n*-oct on it. The system was stirred and equilibrated overnight.
As for the other reagents, *n*-oct saturated with ISA water was saved and employed up to three months from preparation.

5.1.2.2 Apparatus

Log *P* assays were performed by using GLpKa™ and the upgraded version of that system, the T3™. Both instruments were purchased by Sirius Analytical Ltd (Forest Row, UK). A full description of the instrumental setup has been reported in previous **1a Apparatus** paragraph of chapter **3**.

5.1.2.3 Methods

The pKa values acquired in the first part of the work were employed in subsequent log *P* assays.

As regards GLpKA™ instrument, 10 mL of 1-10 mM solutions of samples were titrated under the same conditions as in pKa determinations, but in presence of various volumes of the partitioning solvent.

The *n*-oct/water phase ratio varied according to the expected lipophilicity of the compounds, ranging from 0.01 (*i.e.* 0.1 mL *n*-oct over 10 mL ISA water) to 0.33 (*i.e.* 3.3 mL *n*-oct over 10 mL ISA water); titration of each compound was performed at three or more different phase volume ratios.

Using T3™ instrument, total partition volumes were lowered to 1-1.5 mL, but the same operative concentrations and *n*-oct/water ratios were adopted.

Theoretical background of the method was described in section **2c Potentiometric method** of paragraph **1.3.4.2**. Briefly, from titrations in the presence of the partitioning solvent, the p_oKa (*i.e.* the apparent ionization constant in the presence of *n*-oct) was obtained.

The partitioning of the sample in the organic and aqueous mixtures resulted in a shift of the flexes of the titration curves (pKas) towards higher (for acids) or lower (for bases) pHs and log *P* values were estimated from such shift by using Bjerrum plots (see Equation 42-Equation 46).

Applying the same iterative procedure reported for pKa values in section **2c pKa determination** of paragraph **1.2.2.2**, log *P* data were refined by a weighed non-linear least-squares procedure, where the aqueous pKa values were used as unrefined contributions. Refinement was iterated until

minimization of sum of weighed squares of residuals (S in Equation 8) was obtained. Such procedure was performed by using RefinementPro v. 2.2.6.6 and SiriusT3Refine v. 1.1.3.0 softwares implemented in GLpKA™ and T3™, respectively.

5.1.2.4 Data analysis

Linear regression analysis calculations and related plots were performed in Microsoft Office Excel 2013.

5.1.3 Results and Discussion

Commercial standards

Potentiometric log *P* was assessed for each of n = 14 commercial standards and compared to in-house experimental reference value provided by Sirius Analytical (reported in Table 15).

Compound	Log $P_{T3\ ref.}$	Log P_{GLpKA} ± SE	Δ $P_{T3\ ref.-GLpKa}$
procaine	2.14	1.99 ± 0.03	0.15
quinine	3.50	3.07 ± 0.03	0.43
desipramine	4.21	3.96 ± 0.07	0.25
atenolol	0.22	0.01 ± 0.01	0.21
benzocaine	1.89	n/d	n/d
phtalic acid	0.85	0.95 ± 0.06	-0.1
1-napthol	2.85	2.92 ± 0.07	-0.07
paracetamol	0.46	0.36 ± 0.06	0.1
ciprofloxacin	-1.08	-1.20 ± 0.34	0.12
sulfathiazole	0.07	-0.05 ± 0.07	0.12

papaverine	2.95	3.43 ± 0.04	-0.48
benzthiazide	1.73	n/d	n/d
ibuprofen	3.97	3.86 ± 0.02	0.11
diclofenac	4.51	4.36 ± 0.06	0.15

Table 15: Potentiometric log P determined with GLpKa reported with experimental reference (ref.) values by Sirius T3.

Issues were mainly encountered for compounds showing pKa values at the extremes of the operative range of the glass electrode (e.g. benzocaine) and/or low solubility (e.g. bezthiazide). In the former case, since the partitioning of the sample weakened acidic and basic sites, the flexes of Bjerrum curves were shifted so that determination became less accurate than the one of mid-range pKas. In the latter one, precipitation occurred when pH was approaching pKa, hampering data point collection (see Bjerrum plot for benzthiazide in Figure 41) and subsequent data refinement.

Moreover, data processing turned out to be challenging as electrode performance got worse in presence of *n*-oct. In addition, the electrode needed to be routinely cleaned with 80% MeOH to avoid tight junction occlusion by compound precipitation that was particularly critical in the case of the miniaturized combined glass electrode in Sirius T3 instrument.

Nonetheless, log *P* values showed a good agreement with reference values provided by Sirius for both soluble (Log P_{GLpKA} = 0.87(± 0.03) Log $P_{T3\ ref}$ – 0.09(± 0.07); n = 10, s = 0.16, F = 925 and R^2 = 0.992) and slightly soluble compounds (Log P_{GLpKA} = 0.57(± 0.13) Log $P_{T3\ ref}$ + 1.70(± 0.49); n = 3, s = 0.14, F = 20 and R^2 = 0.953). The accuracy in lipophilicity determination strongly depended on the accuracy and precision of previous pKa measurements, thus explaining the decrement of correlation coefficients in comparison to the ionization constants.

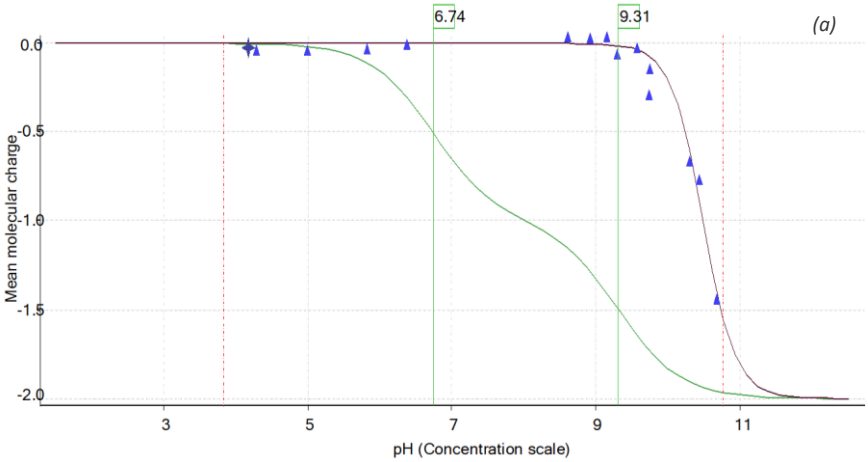
Moreover, a linear regression of experimental log *P* versus computed values (cLogP by ACD Lab Percepta v. 8) was calculated for commercial drugs:

Log P_{GLpKA} = 1.07(± 0.14)cLogP – 0.42(± 0.39)

n = 12, s = 0.77, F = 56, R^2 = 0.848

Equation 85

Correlation proved appreciable, even if lower than the one observed for pKa values. As found for experimental versus reference correlation, it could be speculated that a relatively larger error in prediction was reasonable for lipophilicity since that property accounts for ionization propensity of the compounds.
That correlation model was used to compare the prediction accuracy for commercially available drugs to the one for NCEs at discovery stage.



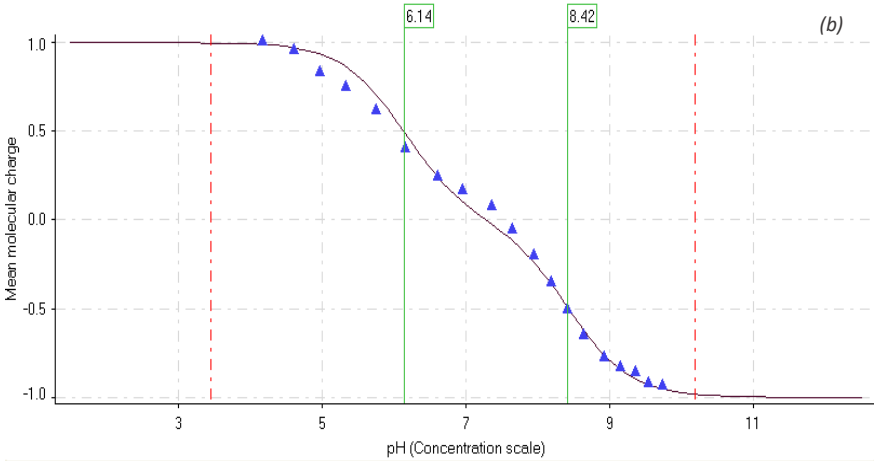


Figure 41: Bjerrum plot for (a) benzthiazide and (b) ciprofloxacin. In case (a) collection of data points was hindered by precipitation, leading to a worse fitting (RMS = 0.485) in comparison to case (b), in which the ampholyte was fully dissolved all through the assay (RMSD = 0.139).

Proprietary NCEs

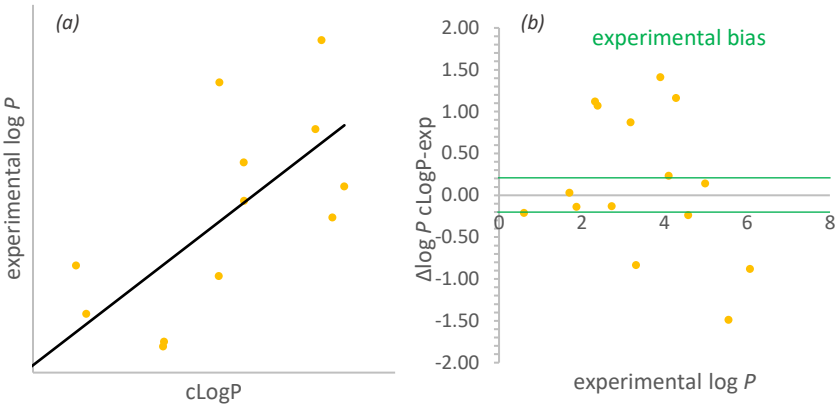
After testing commercial standards, potentiometric approach was applied to proprietary NCEs and results were correlated to cLogP predicted values by ACD Lab Percepta 2012 v. 8:

$$\text{Log } P_{GLpKA} = 0.85(\pm 0.15)\text{cLogP} - 0.38(\pm 0.58)$$

$$n = 15,\ s = 0.85,\ F = 32,\ R^2 = 0.714$$

Equation 86

Tested drug candidates were extremely insoluble and prone to precipitation over the pH range covered during the experiment or showed stability issues at strongly alkaline or acidic pHs. As discussed previously in this thesis, low solubility forced to determine pKa in co-solvent mixtures, resulting in a higher uncertainty of the final result.



Graph 5: (a) Correlation between experimental and predicted log P values for NCEs and (b) related residuals graph.

Correlation between experimental and predicted log *P* proved poor (see Equation 86), as reported in Graph 5*a*, unveiling a very low accuracy in prediction from the software. As did for pKas, the bias between computed and experimental values was calculated (see Equation 78) and plotted in a residuals graph as a function of potentiometric log *P*. The mean bias was 0.7 and error in prediction was considered acceptable (*i.e.* comparable to the experimental bias of 0.2 log unit) for only 26% of the tested compounds.

A further analysis on bias values proved mean error to increase passing from monoprotic (n = 3 NCEs, mean bias = 0.16 log unit) to multiprotic drug-like molecules (n = 12 NCEs, mean bias = 0.79 log unit).

On the basis of this evidence low predictivity could be mainly ascribed to structural and ionization complexity of NCEs and to the fact that the software did not account for ion-pair partitioning, which was indeed found for several specimens. As discussed above in this thesis, prediction of partitioning for ionized species is often hindered by the low accuracy of predicted pKas, which was an issue especially for compounds at the discovery stage that were not represented in the training set.

Thus, once again, relevance of experimental physico-chemical profiling for improving predictions about new drug candidates was highlighted.

5.1.4 Conclusions

Potentiometric approach was tested on the same set of standards previously employed in pKa assays, comparing data with reference values provided by Sirius Analytical. The technique turned out accurate for soluble (R^2 = 0.992, n = 10) as well as for poorly soluble (R^2 = 0.953, n = 4) commercial drugs that showed moderate chemical complexity. Furthermore, correlation between experimental and computed (cLogP by ACD Lab Percepta v. 8) results was lower than the one found for pKa values (R^2 = 0.848 versus 0.978) but still pleasant.

The approach was easily feasible, except for compounds displaying extreme pKas that returned more challenging data to refine. Then potentiometric log *P* was assessed for n = 15 proprietary compounds and such data were correlated to the computed ones, returning a poor correlation coefficient (R^2 = 0.714) in comparison to the previously analysed commercial set.

The study unveiled several limits of the method. Compounds showing (i) low solubility, (ii) limited stability over the pH range of titration and (ii) multiple ionisable sites (n ≥ 3) required multiple assays to return data of acceptable quality for refinement that was, however, difficult. Finally, powder material was required to reach suitable concentrations in the sample vial, a condition that was not easily fulfilled at discovery and pre-development stages when NCEs are more often available as DMSO stock solutions.

Nonetheless, potentiometric log *P* assay could be regarded as useful for NCEs with no more than n = 3 ionisable sites and, most of all, for fragments or chemical precursors of the final drug, in order to provide high quality experimental data for training computational software.

6. Set up and optimization of a micro shake-flask automated protocol to measure lipophilicity of proprietary NCEs

6.1 Aim of the work

Potentiometric approach was applied to proprietary NCEs, but some technical issues proved this approach to be unsuitable for analysing problematic compounds. In fact, the presence of multiple protonation sites in addition to a general low solubility and unknown stability of such molecules strongly affected the goodness-of-fit of experimental data to the theoretical titration curves, when the partition solvent *n*-octanol was added to the sample vessel.

To parameterize lipophilicity of neutral drugs and to overcome the technical limitations of potentiometric method, an automated miniaturized shake-flask operating procedure was set up for measuring log $D_{7.4}$ of drug candidates at early discovery phases. This approach meant to improve conventional shake-flask method (i) reducing the required amount of substance, (ii) shortening partition times and (iii) minimizing the need for manual liquid-handling by using a robotic platform.

DMSO stock solutions of the analytes were employed as the starting material and experimental variables such as organic phase/aqueous phase ratio, partition time and dilution factors were optimized. Also sample processing was performed by the robotic platform and, finally, quantitation was carried with an ultra pressure liquid chromatography-tandem mass spectrometric system (UPLC-MS/MS).

6.2 Experimental part

6.2.1 Materials

Name	Structure	MW

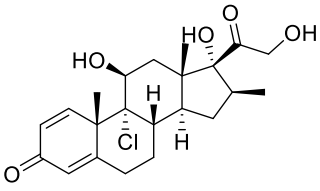
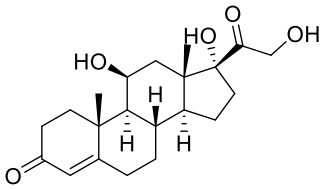
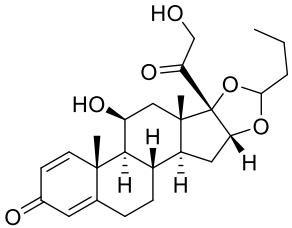
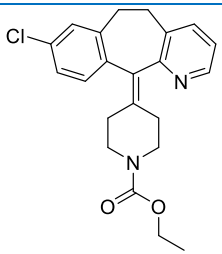
beclomethasone		521.0
hydrocortisone		362.5
budesonide		430.5
loratadine		382.9

Table 16: Commercial standards employed to set-up and validate micro-shake flask approach. Each compound is reported with respective structure and molecular weight (MW).

Beclomethasone, hydrocortisone, budesonide (reported in Table 16) and proprietary NCEs were kindly donated by Chiesi Farmaceutici S.p.A (Parma, Italy). Loratadine (see Table 16), 3-(N-morpholino)propanesulfonic acid (MOPS), potassium hydrochloride (KOH) pellets EMPLURA®, *n*-octanol (*n*-

oct), methanol (MeOH), acetonitrile (ACN) LC/MS grade dimethylsulfoxide (DMSO) and formic acid were purchased from Sigma Aldrich (Milan, Italy). Potassium chloride (KCl) was provided by Acef SpA (Piacenza, Italy). Aqueous solutions were prepared employing ultra-pure Millipore water (Darmstadt, Germany).

Aqueous phase of the biphasic system was prepared by adding a suitable amount of MOPS powder to ultra-pure water to a final concentration of 50 mM. Then pH was corrected at 7.4 by adding KOH 5 N- made from pellets- in order to avoid dilution of the buffer and ionic strength was made up to 0.15 M with KCl.

n-octanol, chosen as organic phase, was poured onto the buffer and the so made biphasic system was equilibrated by stirring overnight at room temperature. If no precipitation was detected, that was saved up to two months at room temperature.

DMSO solutions of the analytes and the internal standards (IS) at 10 mM concentration were split in aliquots of 100 µL and stock in plastic tubes (Eppendorf, Hamburg, Germany) at -20°C: doing so, only a small volume of the analyte solution was defrosted for each working day, avoiding multiple freeze-thaw cycles to the rest of the stock.

Samples were prepared in 1.0 mL deep 96-well plates with V buttom (Nunc™ by Thermo Fisher Scientific, Massachusetts, USA) that were compatible with the autosampler of the LC-MS system.

6.2.3 Apparatus

6.2.3.1 Robotic platform

Hamilton Robotics Microlab STAR Let platform (Hamilton Company, Nevada, USA) was employed to prepare and process samples into 96-well plates. It consisted of a workstation (reported in Figure 42) with a mechanical arm carrying pipetting tips, which were automatically air-displaced after each aspiration/ejection step.

Plastic vessels were filled with MOPS-saturated *n*-oct (6), *n*-oct-saturated MOPS phase (7) and MeOH (8) and served as reservoirs for sample preparation and phase dilution.

Dedicated positions of the workstation allocated 1.5 mL plastic tubes (Eppendorf, Hamburg, Germany) containing DMSO stock solutions of both analytes (3) and internal standards (1) as well as resulting biphasic systems for partitioning.

The platform was configurated so that n = 4 replicates for each of n = 4 compounds could be prepared and processed in a single cycle. All the functions of the robotic system were controlled by Venus software (Hamilton Company).

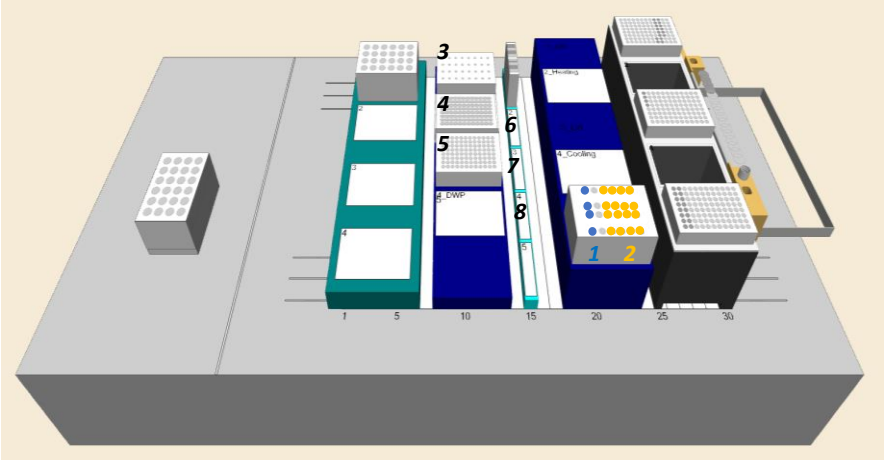


Figure 42: Layout of Hamilton Robotics Microlab STAR Let workstation. (1) IS stock solutions; (2) final samples for partitioning; (3) analyte stock solutions; (4) 96-well plate for centrifuge; (5) 96-well plate for UPLC-MS/MS analysis; (6) n-octanol; (7) MOPS buffer; (8) methanol for phase dilution.

Partition was ensured by a tube revolver (Thermo Fisher Scientific, Massachusetts, USA) equipped with 1.5 mL tube holder. Sample centrifugation was performed with a Sorvall™ ST 40 Centrifuge (Thermo Scientific, Massachusetts, USA) with 96-well plate rotor package.

6.2.3.4 LC-MS/MS system

Chromatographic methods were set up on an Acquity UPLC system (Waters, Massachusetts, USA) interfaced with a QTRAP 5500 instrument (AB Sciex, Milan, Italy) equipped with an Electrospray Ionisation (ESI) ion

source as detector. Data acquisition and processing was performed by using Analyst version 1.5.1 software by AB Sciex.
The UPLC system was equipped with an Acquity reverse-phase (RP) UPLC Kinetex (Phenomenex, California, USA)) C8 100 × 2.1 mm, 1.7 µm (Waters, Massachusetts, USA) column.

Figure 43: Experimental apparatus used to set up micro shake-flask method: (a) Hamilton Robotics Microlab STAR Let system; (b) Waters Aquity UPLC and (c) AB Sciex QTRAP 5500.

6.2.4 Methods

6.2.4.1 Sample preparation and processing

Experimenal parameter	Optimized value
n-oct/MOPS buffer ratios	0.33 (200/600 µL)
	0.42 (250/600 µL)
partition time (h)	2.5
dilution factors (DF)	5x for MOPS buffer
	500-1000-2000x for *n*-oct

Table 17: Optimized values for experimental variables.

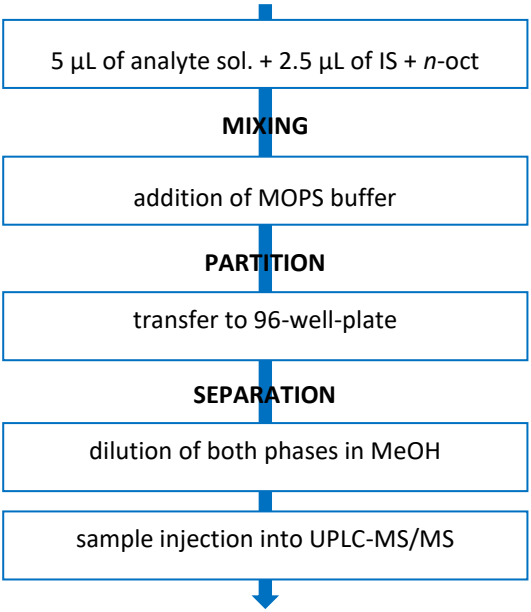


Figure 44: Workflow of sample preparation for micro shake-flask assay.

Optimization of experimental variables will be fully presented in **6.3 Results and Discussion** paragraph.

First, pre-made DMSO stock solutions 10 mM of analytes and IS as well as empty tubes for the resulting biphasic systems were placed in the workstation (position 3, 1 and 2 in Figure 42 respectively). To simplify sample preparation, both internal standard and analyte were added to the biphasic system at the beginning of the partition assay. Two different organic/aqueous phase ratios were tested as duplicates for each analyte: 200 or 250 µL of *n*-oct were pipetted in the empty tubes and 5 µL of analyte together with 2.5 µL of IS stock solution were added to the organic phase. Mixing was ensured by alternative aspiration and ejection of the liquid by the pipetting tips of the robot. Then 600 µL of the MOPS phase were added to the sample vials and the resulting biphasic systems were placed in the tube revolver for 2.5 h.

After partitioning, the samples were placed back in the workstation and transferred by the robot into a 96-well plate (position 4 in Figure 42) to be centrifuged (1000 g at 20 °C for 5 min). The plate was transferred to the robotic platform again for phase separation and dilution.
A suitable volume of MeOH was placed in a clean plate (position 5 in Figure 42) and spiked with *n*-oct aspirated from the specimen so that the organic phase was diluted 500 (dilution factor DF = 500) or 1000 folds (DF = 1000) according to the expected lipophilicity of the compound; similarly, fresh MeOH was pipetted in clean wells and added with MOPS to dilute aqueous phase at 1:5 ratio (DF = 5).
Mixing was performed by repeated aspiration and ejection of the solutions. Finally, the sample plate was transferred to LC-MS/MS system for analysis.

6.2.4.2 LC-MS/MS method

Time (min)	A%	B%
0.00	99	1
1.00	99	1
4.50	5	95
5.50	5	95
6.00	99	1
7.50	99	1

Compound	$[M+H]^+$	Product ion	IS
budesonide	431.3	323.3	becl.
		173.1	
beclometasone	409.4	279.3	bud.
		239.4	
hydrocortisone	363.3	326.8	becl.
		267.4	
loratadine	383.7	337.2	becl.
		267.2	

Table 18: UPLC gradient used for analysis of log D7.4 samples.
Table 19: Mass transitions selected for commercial standards and IS used for mass analysis.

The generic gradient is shown in Table 18: mobile phases consisted of water (A) and ACN (B), both added with formic acid at 0.1%, and the run time was 7.5 min. The flow rate was 0.35 mL $\cdot min^{-1}$ and the column temperature was

set at 55 °C. The injection order for each analyte was dilution solvent, MeOH (n = 2 times), buffered aqueous phase (n = 3 times), MeOH, organic phase (n = 3 times with MeOH in between). The injection volume was set at 2 µL.

MS was preferred over UV-detection because it provided higher sensitivity and wider applicability. Mass spectrometer conditions were chosen as follows: curtain gas (CUR) 20 bar; capillary voltage (IS) 5000 V; source temperature 350 °C; GS1 10 bar; GS2 30 bar. MS/MS detection method was optimized for each analyte to select two mass transitions and such optimized fragmentations are reported in Table 19 for the test compounds. Cumulative areas resulting from the two fragments were employed in the subsequent calculations. The peak of the analyte in each phase ($A_{analyte}$) was divided by the area of the IS (A_{IS}) to account for signal fluctuation:

$$A_{norm} = \frac{A_{analyte}}{A_{IS}}$$

Equation 87

Normalized area values (A_{norm}) in organic ($A_{norm,oct}$) and aqueous phase ($A_{norm,MOPS}$) were used for Log $D_{7.4}$ calculation in the following equation:

$$\text{Log } D_{7.4} = \log_{10}\left(\frac{DF_{oct} \cdot A_{norm,oct}}{DF_{MOPS} \cdot A_{norm,MOPS}}\right)$$

Equation 88

6.2.4.3 Data analysis

Calculations on MS data were performed in Microsoft Office Excel 2013 to return log $D_{7.4}$, accuracy and precision values.

6.3 Results and Discussion

Four commercial standards (*i.e.* hydrocortisone, beclomethasone, budesonide and loratadine) were chosen to cover a suitable range of lipophilicity values (1.67 < log $D_{7.4}$ < 4.56) and employed to set up the automated miniaturized shake-flask approach.
Experimental log $D_{7.4}$ of such compounds was determined by using a non-scaled down shake-flask method in which samples were made of 400 or 200 µL of *n*-oct, 10 µL of analyte stock, 5 µL of IS stock and 1 mL of MOPS buffer. The systems were let partitioning for 4 h, phases were diluted (DF = 1000 or 2000 for organic phase and DF = 5 for aqueous buffer) and quantified by using the above described LC-MS/MS method. Such values were taken as reference in the following protocol optimization phase.

Compound	Log $D_{7.4}$ ± SD
beclomethasone	1.60 ± 0.27
hydrocortisone	1.56 ± 0.05
budesonide	3.13 ± 0.16
loratadine	4.56 ± 0.01

Table 20: Reference values from manual shake-flask; SD standard deviation (n=4).

Precision and accuracy of the assays were parameterized by relative standard deviation (%RSD) and bias (%bias) respectively:

$$\%\mathrm{RSD} = \frac{\text{standard deviation}}{\text{average}} \cdot 100$$

Equation 89

$$\%\mathrm{bias} = \left(100 - \frac{\text{experimental value}}{\text{reference value}}\right) \cdot 100$$

Equation 90

and were optimized by modulating experimental variables such as (i) partition phase volume ratio and (ii) partitioning time.

6.3.1 Optimization of partitioning phase ratio

The main issue encountered across the miniaturization was to find the suitable ratio between organic and aqueous phase: in fact, since each partition well allowed an ideal operative volume of 1 mL, a mutual contamination of buffer and organic solvent during separation could easily occur.

In this context, different *n*-oct/MOPS buffer ratios for the four standards were tested (*i.e.* 100, 120, 300 and 200 µL respectively), using n = 4 replicates for each standard and partition ratio a-day for a maximum of 4 days (n = 16 samples). Partition time was set at 4.5 hours, DF were 500 and 5 for organic and aqueous phase respectively.

%RSD was calculated as a precision index to further select that volume ratio that minimized variability within samples. Results are displayed in Table 21: except for the case of budesonide, it turned out that the observed variability was higher when *n*-oct volumes were lower than 130 µL, supporting the hypothesis of mutual contamination of the phases during partition solvents separation by the robotic system.

Hence optimal solvent ratios were set at 0.33 and 0.41 (*i.e.* 200 µL of *n*-oct over 600 µL of MOPS buffer and 250 µL of *n*-oct over 600 µL of MOPS buffer). RSD% of loratadine was slightly above the mean experimental error (RSD < 15%), probably due to its extreme lipophilicity.

Compound	RSD% at low octanol volumes (100-120 µL)	RSD% at high octanol volumes (130-200 µL)
hydrocortisone	20.02	8.14
beclomethasone	13.80	10.20
budesonide	7.97	9.23
loratadine	22.04	15.91

Table 21: Inter-day precision (%RSD, n = 16) for the selected standards tested to optimize partition solvent ratios.

6.3.2 Application to commercial standards

Then the same n = 4 commercial standards were processed to test the compliance of the scaled-down assay: precision and accuracy of the automated shake-flask protocol were evaluated by measuring log $D_{7.4}$ of n = 4 reference standards on n = 4-5 consecutive days and comparing the results with non-automated shake-flask lipophilicity data, taken as reference (see Table 22-Table 25).

Partitioning time was set at two hours; to reduce differences among matrices and accounting for instrument sensitivity and compound lipophilicity, dilution factors of the two phases were chosen as DF = 5 for MOPS buffer and DF = 500-1000-2000 for the organic phase.

Hydrochortisone						
day#	**1**	**2**	**3**	**4**	**5**	**Ref. value**
Av. log $D_{7.4}$	0.92	1.12	1.71	1.64	1.67	1.56
%RSD	22.27	2.26	4.90	3.20	8.10	3.21
%bias	41.19	28.53	-9.29	-5.29	-7.21	

Table 22

Beclomethasone						
day#	**1**	**2**	**3**	**4**	**5**	**Ref. value**
Av. log $D_{7.4}$	1.57	2.76	1.87	1.82	2.29	1.60
%RSD	22.73	1.03	13.19	12.42	1.61	16.9
%bias	1.87	-73.17	-16.90	-13.97	-43.71	

Table 23

Budesonide						
day#	**1**	**2**	**3**	**4**	**5**	**Ref. value**
Av. log $D_{7.4}$	2.03	2.61	2.76	2.88	2.79	3.13
%RSD	9.24	4.66	9.56	9.43	13.25	5.11
%bias	35.03	16.52	11.96	8.04	10.87	

Table 24

Loratadine

day#	1	2	3	4	Ref. value
Av. log $D_{7.4}$	2.87	3.36	3.27	3.41	4.56
%RSD	7.41	3.05	33.71	19.46	0.22
%bias	37.11	26.30	28.25	25.12	

Table 25

Table 22-Table 25: Data of the four standards used to test the miniaturized automated shake-flask approach. Outliers and data out of the acceptable limits (%RSD > 0.2, %bias > 17) are marked with red. Each reported value is the average of four replicates (Av. log $D_{7.4}$, n = 4).

In general, precision and accuracy registered during days 1 and 2 of the automated micro shake-flask procedure revealed that some severe experimental issues still needed to be solved, resulting in an out-of-range precision and accuracy. In fact, whereas %RSD in reference manual shake-flask procedure ranged between a 0.2% and a 17%, both precision (%RSD) and accuracy (%bias) showed percentual shifts higher than 20-30%.

For at least three out of four test compounds (*i.e.* hydrocortisone, beclomethasone and budesonide), the observed variability could be reduced taking into account the setting of the robotic system regarding the pipetting tip position during phase separation. At first, the system was programmed to discriminate between the electric conductivity of the two solvents, but this approach proved unsuitable for the extremely low conductivity of *n*-oct. Hence, mutual contamination of the phases occurred leading to erratic results.

This main technical issue was solved taking into account the different superficial tension of the two liquids and the consequent counter pressure encountered by the pipetting tip entering into the two partition phases. This required adequate programming of Venus software.

After this intervention, precision for three out of four standards improved significantly (%RSD < 15 on days from 3 to 5). However, loratadine and, to a smaller extent, beclomethasone still showed the highest deviations both in precision and accuracy that was speculated to be related to partition time.

6.3.3 Optimization of partitioning time

The role of partition times was thus investigated, following the hypothesis that compounds at the extremes of the lipophilicity range might take longer times to equilibrate in comparison to the others.
To this aim, different partition times were assayed, *i.e.* 1.5 h, 2.5 h and 4 h, on hydrocortisone and loratadine, respectively (see Table 26 and Table 27).

Hydrochortisone				
Partition time (h)	**1.5**	**2.5**	**4**	**Ref. Value**
Av. log $D_{7.4}$	1.67	1.55	1.56	1.56
%RSD	8.38	1.94	3.21	3.21
%bias	-7.05	0.64	0.00	/

Table 26: Hydrocortisone Log $D_{7.4}$ assayed at different partition times. n = 4 replicates for each time assayed.

Loratadine				
Partition time (h)	**1.5**	**2.5**	**4**	**Ref. Value**
Av. log $D_{7.4}$	3.41	4.16	4.07	4.56
%RSD	19.35	0.96	2.21	0.22
%bias	25.22	8.77	10.75	/

Table 27: Loratadine Log $D_{7.4}$ assayed at different partition times. n = 4 replicates for each time assayed.

It was observed that both precision and accuracy significantly dropped if samples were revolved for 1.5 h: hence partitioning time was definitively set at the optimal value of 2.5 h.

6.3.4 Application to proprietary NCEs

The optimized micro shake-flask approach was finally applied to n = 4 NCEs compounds tested as n = 4 replicates; once again the log $D_{7.4}$ obtained

under non automated shake-flask conditions were taken as reference values (*vide supra*).
The average precision and accuracy (%RSD < 6.5 and %bias < 3) were better than the ones found for the commercial standards and, in addition, they were comparable to variability of the golden-standard potentiometric method, for which %RSD = 10% is generally accepted.

6.4 Conclusions and perspectives

A miniaturized shake-flask approach was set up and developed to assay log $D_{7.4}$ of slightly soluble, neutral or multiprotic drug candidates, for which potentiometric approach was hampered.
That miniaturized method was prioritized over potentiometry for early discovery needs since requiring minimum amount of drug sample. In that sense, coupling liquid chromatography to a mass spectrometer detector ensured low affection by impurities, detectability of degradation and higher sensitivity in comparison to UV-detection.
Micro shake-flask proved a useful technique to rank chemical homologs at early stages of discovery when synthesis was not yet optimized and analytes as DMSO stock solutions were preferably available.
The main requirement for this protocol was the compound to be stable in solution for at least 4 h that was the time period required for sample preparation, partition and analysis. n = 4 analytes in n = 4 replicates were quantified in a total time of 12 h, consuming less than 40 µL of DMSO stock solution.
In perspective, the automated micro shake-flask approach could be applied to evaluate log *D* at different biorelevant pHs to experimentally draw lipophilicity profiles of drug candidates.
In addition, different biphasic systems could be tested to gain additional information on structural properties of drug candidates (*i.e.* H-bond donor and acceptor capacity).

7. Solubility assessment by turbidimetric approach

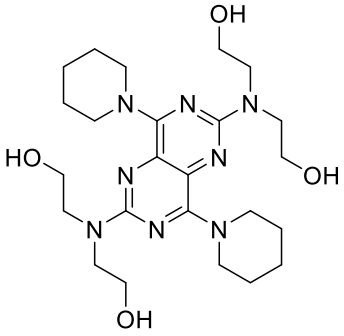
7.1 Solubility investigation by using Sirius T3™

7.1.1 Aim of the work

The third and last part of the present project focused on solubility investigation by turbidimetric approach implemented in Sirius T3™ instrument.
Intrinsic (and kinetic) solubility of some commercial standards was determined through CheqSol and Curve Fitting assays according to the chaser and non-chaser behaviour of the analyte. After practicing those methods, they were applied to proprietary NCEs at discovery and early-development stages to evaluate their suitability for industrial purpose.
Characterisation of solid state of precipitate from solubility assays was attempted to unveil pH- and time-dependent changes of the starting solid form.

7.1.2 Experimental part

7.1.2.1 Materials

Name	Structure	MW
dipyridamole		504.63

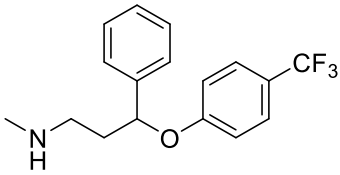
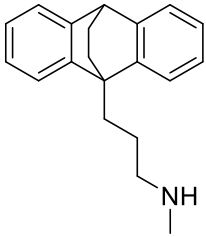
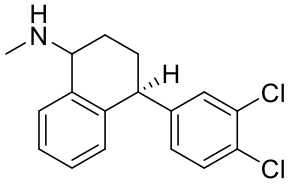
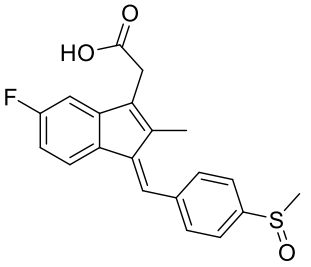
fluoxetine		309.33
maprotiline		277.40
sertraline		306.23
sulindac		356.41

Table 28: Additional standards used for solubility assay with Sirius T3, reported with respective structure and molecular weight (MW). All the other commercial drugs used in this part of the research project are reported in Table 10.

A small set of standards was chosen to include chasers (C) and non chasers-behaving (NC) drugs (see Table 10 and Table 28 for structures).
Quinine hydrochloride (NC), 1-naphtol (C), diclofenac sodium salt (C) and sulfathiazole (C) were donated by Food and Drug Department of University of Parma (Parma, Italy); ciprofloxacin (C), sulindac (C/NC) and dypiridamole (C/NC) were purchased by Sigma Aldrich (Missouri, USA); maprotiline hydrochloride (NC), sertraline hydrochloride (NC) and fluoxetine hydrochloride (NC) European Pharmacopeia standards were provided by European Directorate for the Quality of Medicine (EDQM, Strasburg, France).
Proprietary molecules were kindly donated by Chiesi Farmaceutici S.p.A. (Parma, Italy).
ISA water, co-solvent methanol 80% (MeOH 80%), acidic titrant (HCl 0.5 N) and basic titrant (KOH 0.5 N) were prepared as reported in section **3.1.2.1 Materials** of paragraph **3.1.**

7.1.2.2 Apparatus

Sirius T3^{TM} was used for all the measurements; description of the instrumental set up was reported in **3.2.2.1 Apparatus** paragraph.

7.1.2.3 Methods

Two kind of assays were performed according to the analyte, which were named CheqSol and Curve Fitting experiments respectively.

3a CheqSol assay

From 2 to 10-20 mg in form of free acid (e.g. 1-naphtol), base (e.g. quinine), ampholyte (e.g. ciprofloxacin) or a salt (e.g. hydrochlorides of maprotiline, fluoxetine and sertraline) were weighed to ensure complete solubilisation of the sample when the drug was ionized and precipitation when it was fully neutral.

1.5 mL of ISA water were used to dissolve the sample; then either HCl or KOH was added, according to the ionisable nature of the analyte, to reach a pH at which the solute was fully dissolved in its ionized form. The solution was stirred until complete dissolution was achieved.
The solution of ionized analyte was back-titrated by adding controlled volumes of 0.5 M KOH or 0.5 M HCl titrant (e.g. acids were titrated from alkaline to acidic pH and, viceversa, bases were titrated from acidic to alkaline pH) until the solution became cloudy, which indicated that precipitation of the poorly soluble neutral species occurred.
Precipitation was detected as light scattering by using the optic dip probe. The titrant volumes added during this stage were calculated to achieve a fast titration without overcoming the precipitation point by more than ~ 1 pH unit.
Precipitation was further promoted by adding aliquots of the same titrant, until the pH had changed by a predefined increment (e.g. 0.1 pH unit) or until a fixed time had elapsed (e.g. 60 sec). Sample precipitation should cause the pH to spontaneously readjust as quickly as the titrant was added. After sufficient data had been collected to calculate a solubility result, the pH was recovered to starting value to ensure that neither crystals nor solid sample remained on the apparatus. After re-dissolution, the probes were repeatedly washed with MeOH 80% before any further experiment.
For poorly soluble compounds, CheqSol assay was repeated at least three times in presence of different percentages of co-solvent.

3b Processing of data from CheqSol assay

SiriusT3Refine software v. 1.1.3.0 was employed for data processing.
At equilibrium pH gradient was invariant (*i.e.* dpH/dt = 0) since dissolution and precipitation phenomena were in balance. Under that condition, the intrinsic solubility of the neutral form was calculated as reported in the Errore. L'origine riferimento non è stata trovata. section from paragraph **1.4.6**.

Briefly, concentration of the sample at each point of the experiment was calculated and pH gradient was plotted as a function of concentration in the Crossover graph. Every time the pH gradient was inverted, the intrinsic solubility was bracketed by the neutral species concentration at the points on either side of the change and such concentrations were both included in the Cross-over graph.
Several information were needed for calculation concentration, that were (i) the formula weight of the compound, (ii) the weight of the solid sample, (iii) the concentration of the acid and base titrants and (iv) the ionization constant(s) (pKas) of the sample, both the value and type (acid or base).
All crossing points falling on y = 0 were averaged to calculate intrinsic solubility and the coefficient of variation related to that value (%CV).
Kinetic solubility (*i.e.* concentration at the pH-point at which precipitation started) was calculated knowing the sample weight, the volume of solution and the distribution of species graph.
For poorly soluble compounds apparent solubility was calculated from each set of data acquired in presence of co-solvent and aqueous solubility was finally determined by linear extrapolation using Yasuda-Shedlovsky method as described in **2c pH-metric titrations in co-solvent/aqueous media** section.

3c Curve Fitting assay

As reported for CheqSol experiment, a proper amount of drug was weighed to ensure that, at a suitable pH, the neutral form reached a concentration higher then the intrinsic solubility to be prone to precipitation.
Conversely to the previous method, complete dissolution of the starting powder was not strictly necessary in this case since the absorbance baseline could account for that initial cloudiness.
The solution of ionized analyte was back-titrated by adding controlled volumes of base or acid titrant until precipitation occurred.

Precipitation was maintained by adding further aliquots of the titrant and experimental points were recorded until titration was completed (*i.e.* reaching the extreme of the pH range set at the beginning of the assay).
After collecting enough data for refinement, re-dissolution of the sample was promoted by reverting the solution to a pH at which the analyte was fully ionized.

3d Processing data from Curve Fitting assay

Also in this case data processing was performed by using SiriusT3Refine software in the version 1.1.3.0.
Solubility was calculated by measuring the extent at which precipitation perturbed dissociation equilibrium, following the same theoretical principles described for potentiometric log *P* and co-solvent pKa assays (see paragraph **1.2.2.2** and **1.3.4.2**).
For this purpose, Bjerrum functions (*i.e.* average number of bound protons vs pH) were calculated to describe the behaviour of the compound in solution (solution graph) and after precipitation occurrence (precipitation graph). The latter could be calculated by using specific equations for mono- and multiprotic compounds as reported in paragraph **2c** of the **1.4.6.2 Experimental assessment of solubility** section.
As long as the sample was fully dissolved, experimental points followed the so-called solution plot: after precipitation started, all titration points deviated from such graph and fell onto the precipitation plot, which resulted from a shift of the solution graph. Distance between the curves was used to calculate intrinsic solubility of the analyte.
Such behaving compounds were not able to supersaturate herein kinetic solubility- calculated at the starting point of precipitation- was assumed equal to intrinsic solubility (S_0) that was easily determined by using the sample weight, the volume of solution and the distribution of species graph. Fitting of experimental points after detection of turbidity to the precipitation graph was optimised to refine solubility result.

Insoluble compounds were analysed in co-solvent/water mixtures (n = 3) and aqueous solubility was derived from apparent solubility results by using Yasuda-Shedlovsky extrapolation, as mentioned for the CheqSol assay.
Calculation of correlation and graphs for solubility results were made in Microsoft Office Excel 2013.

7.1.3 Results and Discussion

Solubility data of commercial standards

The 10 commercial standards were analysed to practice the methods and evaluate their technical limits. Results are summarized in Table 29 and each of them is reported as average of n = 2-5 assays. Log S_0 results were compared to reference values provided by Sirius Analytical and were found in good agreement.

Compound	assay	KS	S_0 (µg/ml) ± SE	ACD log ($1/S_0$)	T3 log ($1/S_0$)	Δ log
diclofenac	CS	65.16 µg/mL	1.7 ± 0.02	7.47	8.25	0.8
sulfathiazole	CS	741.6 µg/mL	513.3 ± 31.0	5.44	5.70	0.3
1-naphtol	CS	1.732 mg/mL	1386 ± 56.4	5.46	5.02	-0.4
ciprofloxacin	CS	443.1 µg/mL	75.7 ± 31.7	6.32	6.64	0.3
fluoxetine	CF	/	43.0 ± 2.7	7.79	6.86	-0.9
quinine	CF	/	442.8 ± 23.4	8.03	5.86	-1.2
maprotiline	CF extr.	/	5.7 ± 0.3	7.60	7.69	0.1
sertraline	CF extr.	/	5.3 ± 1.6	8.19	7.76	-0.4
sulindac	CS	96.52 µg/mL	18.0 ± 0.5	7.71	7.30	-0.4
	CF	/	96.1 ± 2.5		6.57	-1.1

dypiridamole	CS	78.58 µg/mL	8.3 ± 2.6	5.42	7.78	2.4
	CF	/	70.5 ± 11.1		6.85	1.4

Table 29: Kinetic (KS) and Intrinsic solubility (S_0) values for commercial standards determined by using Sirius T3. Kinetic solubility was calculated from experimental point at the beginning of precipitation in CheqSol assays. CS CheqSol; CF Curve Fitting; CF extr. Curve Fitting extrapolation.

Five stages could be identified in Cheqsol experiment that were (i) dissolution, (ii) initial precipitation, (iii) additional precipitation, (iv) chasing equilibrium and (v) re-dissolution. Except for (iv), the same steps could be identified in Curve Fitting assay.

(i) The powder sample was added with ISA water and either acid or base under stirring to reach full dissolution of the analyte in its ionized form. (ii) The solution was then back-titrated until precipitation of the poorly soluble neutral species occurred.

Precipitation was detected by using the optic dip probe as a sudden increase of absorbance signal (*i.e.* reduction of transmitted light to the detector) at a wavelength at which the compound did not absorb light. Monitoring light scattering allowed, in addition, to determine kinetic solubility (*i.e.* the concentration at which precipitation first occurred in solution).

However, it should be noticed that kinetic solubility value returned by CheqSol assay was not meant to be accurate or necessarily comparable to results obtained by other kinetic assays.

(iii) Precipitation was further promoted by adding aliquots of the same titrant. Sample precipitation should cause the pH to spontaneously readjust as quickly as the titrant was added. Additional precipitation stage ensured that sufficient precipitation occurred for the following step of the experiment.

(iv) In the core phase of the experiment, after precipitation of the neutral species had occurred, the solution was repeatedly changed from supersaturated to under-saturated and back again by changing the pH (*i.e.* chasing equilibrium). This stage was repeated until 20 saturation state changes had been measured. That operation meant to actively encourage

pH changes in the direction of equilibrium towards which the system was spontaneously moving.

pH changes over time (*i.e.* pH gradient) was monitored once it had settled to a sustained response and was used to calculate concentration at equilibrium (*i.e.* dpH/dt = 0). A detailed description of the theoretical bases of chasing equilibrium was provided in paragraph **2c** of section of section **1.4.6.2** in the introductive chapter. Aqueous assays were performed without any issue, while co-solvent/water mixtures required several attempts to find the optimal sample weight, balancing the need for starting dissolution and subsequent precipitation.

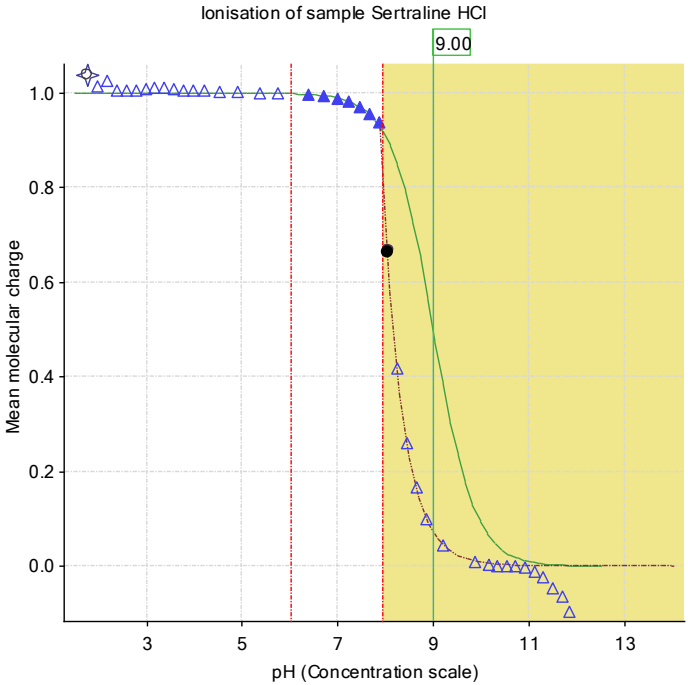


Figure 45: Bjerrum plot for curve-fitting assay of sertraline hydrochloride.

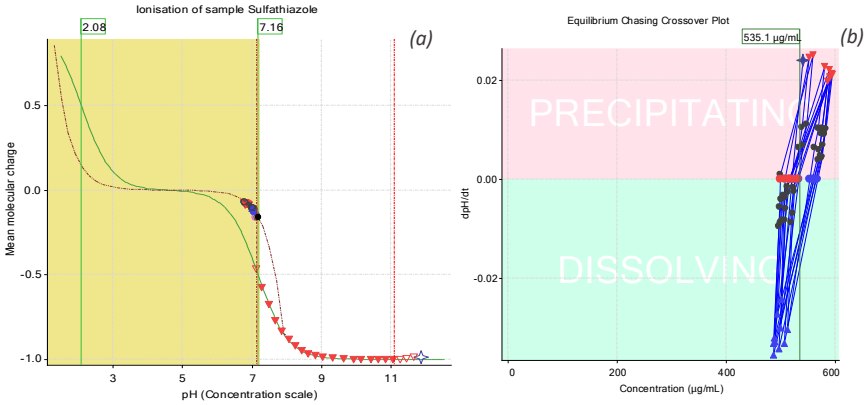


Figure 46: (a) Bjerrum plot for Cheq-sol assay and (b) Cross-over plot of sulfathiazole.

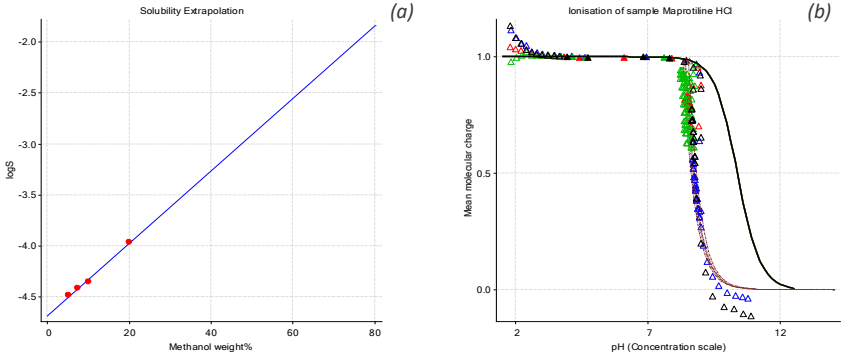


Figure 47: (a) Yasuda-Shedlovsky extrapolation and (b) Bjerrum plot for Curve Fitting assay of maprotiline hydrochloride. n = 4 Curve Fitting assays were performed at different MeOH% (5.2-20%).

Kinetic solubility was calculated for chasers even if it should be pointed out these values to be affected by relevant variability. In fact they depended upon the starting sample weight that determined, in turn, the extent of supersaturation and the time needed for precipitation to start.

Hence they could be used to get highlights about the propensity of the compound to supersaturate, but were not expected to be correlated with kinetic solubility values obtained by using other experimental approaches.

As previously reported for pKa and log *P* values, a correlation between experimental (log $(1/S_0)$) and computed values (by ACD Percepta software, ACDlog $(1/S_0)$) was attempted and a noticeably poor model was found:

$$\log (1/S_0) = 0.41(\pm 0.25)\text{ACDlog}\,(1/S_0) + 4.07(\pm 1.69)$$

$$n = 12,\ s = 0.90,\ F = 2.80,\ R^2 = 0.219$$

Equation 91

That should be explained considering solubility calculation should account for several molecular properties (e.g. ionization, lipophilicity) each of which is affected by an intrinsic error in computation. Moreover, prediction did not consider solid state of the analyte and/or propensity to solid form interconversion in solution that, indeed, strongly affected experimental data.

As a matter of fact, an average deviation of 0.81 log unit was found between ACD and experimental values (see Table 29) with higher deviation for sulindac and dipyridamole that have proved to switch between different polymorphs during solubility assay (see the following paragraph). To strengthen that hypothesis those compounds were excluded from regression model:

$$\log (1/S_0) = 0.88(\pm 0.25)\text{ACDlog}\,(1/S_0) + 0.63(\pm 1.77)$$

$$n = 8,\ s = 0.71,\ F = 12.13,\ R^2 = 0.669$$

Equation 92

Thus correlation significantly improved in Equation 92, as suggested by the increase in R^2 and F values and by the fact that slope and intercept got closer to 1 and 0 respectively. The model, however, remained far from being predictive, supporting the statement that accurate computation of solubility values could be challenging even for relatively simple structures (*i.e.* low MW and at most n = 2 ionisable sites).

Maprotiline and sertraline hydrochloride could not be assayed in purely aqueous medium and complete dissolution could be reached only in presence of MeOH 80%. Four Curve-Fitting assays at different percentages of co-solvent (*i.e.* 5-20%) were then performed for each standard and intrinsic solubility at 0% wt MeOH was obtained by applying Yasuda-Shedlovsky extrapolation (graphs for maprotiline are reported in Figure 47).

Solid state interconversion evaluated for commercial standards

Moreover, it was attempted to study the solid state of the precipitate from solubility assay and sulindac and dypiridamole were chosen as model compounds for this purpose. Box and co-authors [135] reported that sulindac displayed interconversion between two different polymorphs that was promoted by pH-change of the solution over titration (results are shown in the following table).

Compound		Exp. sol.	Ref. sol.	Exp. log $(1/S_0)$	Ref. log $(1/S_0)$	Δlog
sulindac	C	18.0 ± 0.5	11.4*	7.30	7.50	-0.20
	NC	96.1 ± 2.5	71.6*	6.57	6.70	-0.13
dypiridamole	C	8.3 ± 2.6	3.6**	7.78	8.15	-0.36
	NC	70.5 ± 11.1	/	6.85	/	/

*Table 30: Solubility data (µg/mL) for chaser (C) and non-chaser (NC) forms of sulindac and dipyridamole. Experimental data determined by Sirius T3 (exp.) are compared to reference values (ref.) and difference in log unit (Δlog $(1/S_0)$ experimental-reference) was calculated. * values from Box et al. [135]; ** value provided by Sirius Analytical.*

Bjerrum curve from one solubility assay performed for sulindac and the corresponding Neutral-Species Concentration Plot are reported in Figure 48 and Figure 49, respectively.

Titration started at alkaline pH (1) and precipitation was detected at around pH 5, at which kinetic solubility (equal to the intrinsic one) was found to be 96.1 ± 2.5 µg/mL (2): in the first part of the assay the drug behaved as a non-chaser (4). After around 30 min (4), sulindac switched to a chaser-behaving form with a lower solubility (18.0 ± 0.5 µg/mL) in comparison to the non-chaser one. Data showed good agreement with literature (see Table 30), with a difference ≤ 0.2 log unit that was assumed as acceptable experimental bias as previously reported for pKa and log *P*. By using Raman spectroscopy during the assay Box's group confirmed the two forms to be different polymorphs and that evidence was further proved by XRPD analysis of the precipitate.

In order to characterise solid forms, that experiment was repeatedly performed and precipitate was collected for analysis, but the solid quantity was insufficient to perform any assay.

Thus, investigation of solution-mediated transformation of the solid forms was limited to the analysis of Bjerrum plots.

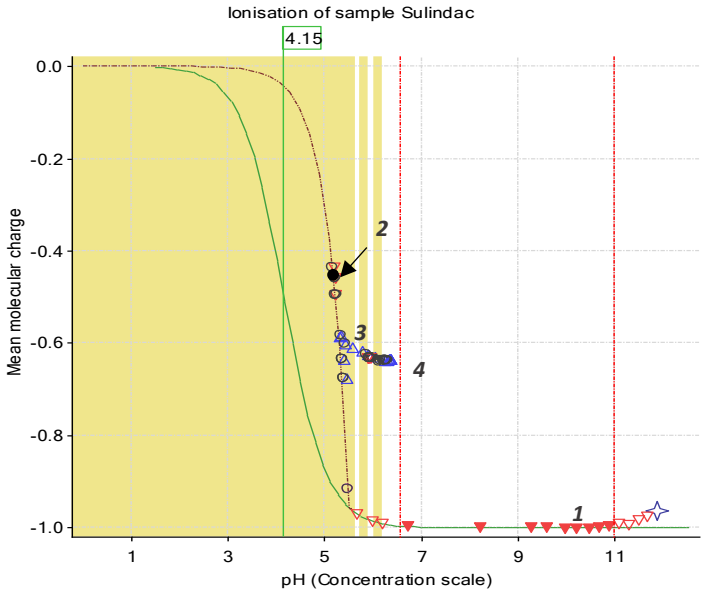


Figure 48: Bjerrum's plot for sulindac CheqSol assay.

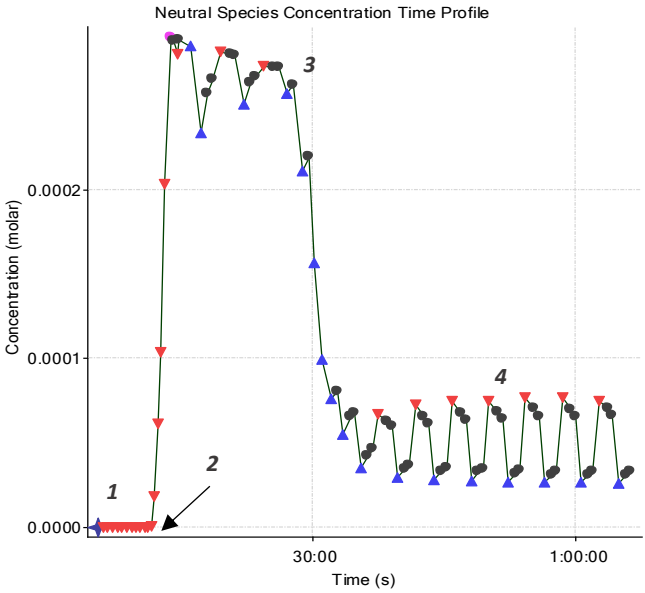


Figure 49: Neutral species concentration graph for sulindac CheqSol assay. To note the sudden transition between phase 3 and 4.

In that context, dipyridamole provided a further example of solution-mediated transformation, observed for n = 3 repeated experiments. Referring to Bjerrum curve and neutral species concentration graph (Figure 51 and Figure 50 respectively), titration started from acidic pH (1) and precipitation was detected around pH 5 (2) at which solubility of the non-chaser form was determined. Then drug slowly turned from a non-chaser to a chaser-behaviour in around 45 minutes, time after which chasing was induced and monitored to determine solubility of the new form.

As found for sulindac, the chaser showed a lower intrinsic solubility than the non-chaser, which was 70.5 ± 11.1 µg/mL and 8.33 ± 2.59 µg/mL respectively. Thus, the latter was supposed to be a more stable polymorph for both the drugs, for instance, a more crystalline solid.

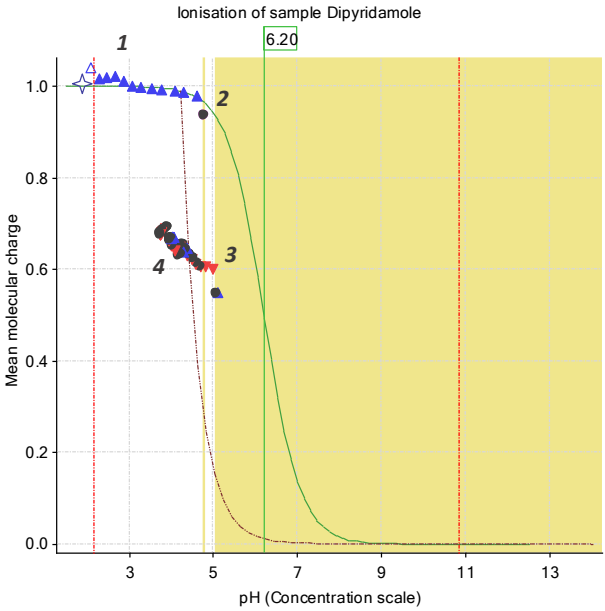


Figure 50: Neutral species concentration graph for dipyridamole. To note the slow transition to the chasing phase in the light-blue circle.

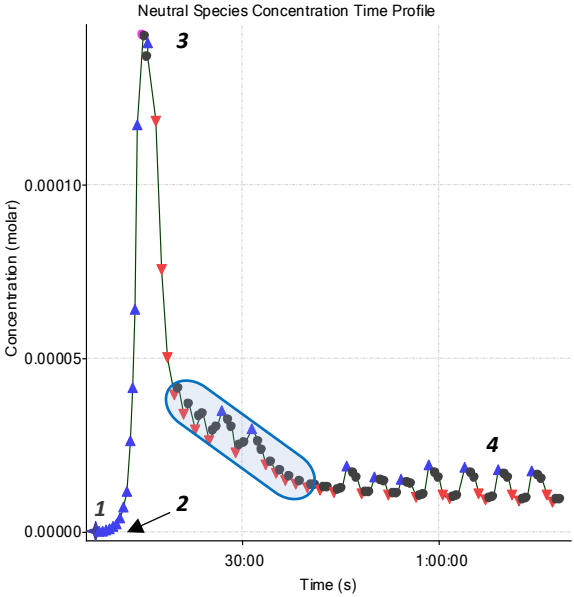


Figure 51: Bjerrum's plot of dipyridamole CheqSol assay.

Focusing on region 3-4 in Figure 47 and Figure 50, a further speculation should be attempted. While phase 4 started suddenly for sulindac, with a harsh switch from the non-chasing to the chasing behaviour, it can be noticed that chasing phase arose more gradually for dipyridamole passing through a transition step (light-blue circle in Figure 51). That let suppose a slower, and thus less favoured, interconversion in solution for dipyridamole in comparison to the previous case.

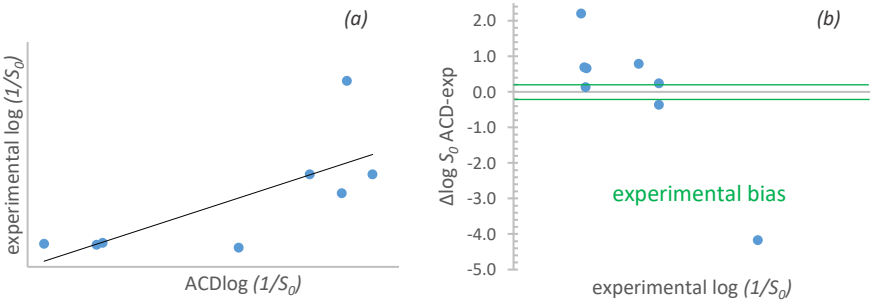
Solubility results for proprietary compounds

CheqSol and Curve Fitting assays were attempted for n = 8 NCEs to evaluate the feasibility of such methods and results were compared to computed log *(1/S_0)* values (by ACD Lab Percepta 2012 v. 8):

log *(1/S_0)* = 1.49(± 0.56)ACDlog *(1/S_0)* - 2.45(± 1.88)

n = 8, s = 1.88, F = 7.15, R^2 = 0.544

Equation 93



Graph 6: (a) Correlation between experimental and computed solubility values for NCEs and (b) related residuals graph.

Not surprisingly, the resulting correlation in Equation 93 was lower than the one observed for commercial standards (R^2 = 0.544 versus R^2 = 0.669 in Equation 92). Moreover, the mean difference between computed and

experimental values was equal to 1.16 log unit and for two over eight NCEs deviation was higher than 2 log units, strongly affecting the goodness of regression model.
From an experimental perspective, optimising the set up of the assay was both time and substance consuming, especially for insoluble compounds. Interconversion between different solid forms was not that apparent from Bjerrum curves as it was for commercial standards, although dualistic chaser- and non chaser-behaviour was observed in some cases.
Moreover, as expected, error in prediction was considerably higher than the acceptable experimental bias (*i.e.* 0.2 log units) as a consequence of molecular complexity, multiple ionisation sites and uncertainty in pKa prediction for multiprotic molecules.
Finally, some insoluble proprietary compounds required to be tested in presence of co-solvent and intrinsic solubility was then extrapolated by Yasuda-Shedlovsky method. Such protocol required several percentages of MeOH 80% to be tested in order to optimise the experimental conditions and returned results affected by higher uncertainty than the ones collected in purely aqueous medium.

7.1.3 Conclusions

Turbidimetric assays (*i.e.* CheqSol and Curve Fitting) implemented in Sirius T3^{TM} were practised to measure kinetic and intrinsic solubility of a small pool of commercial standards and were then transferred to proprietary compounds to assess feasibility of such approaches for a research and development centre needs.
As reported for the other potentiometry-based methods described in the present work, CheqSol and Curve Fitting assays required the analyte (i) to be stable over the pH range of titration and (ii) available as up to 20 mg of powder. Some drawbacks of the approach should be pointed out: (i) above all, the main issue was seeking for the optimal amount of sample to be used in each assay, which ensured complete dissolution at starting pH as well as precipitation of the neutral form during titration.

(ii) Another hinder was represented by insoluble compounds, for which co-solvent/water mixtures were used and extrapolation of aqueous intrinsic solubility required Yasuda-Shedlovsky method.
(iii) Moreover, accuracy of solubility results was limited by the quality of previously determined pKa data, involving higher uncertainty for multiprotic and slightly soluble drugs.
(iv) From a practical perspective, the set up of both aqueous and co-solvent assays for NCEs was time-demanding; each analysis requiring from 1 to 3 hours, these methods could be regarded as medium-low throughput and not suitable for screening application. Conversely, kinetic solubility of discovery drug candidates should be assessed by scale-down approaches in 96-well plates, saving intrinsic solubility determination for later stages of research.
Despite technical limits of the approach, turbidimetric assays proved informative and able to complement knowledge gained from high-throughput screening solubility assays. In that sense, CheqSol and Curve Fitting analysis could prove useful for compounds at pre-development stage (*i.e.* freely available in relatively high amounts) from several perspectives. (i) Turbidimetric assays allowed to classify drugs on the basis of their experimental behaviour and (ii) to unveil the propensity for supersaturation, which could be modulated by using different formulation tools.
In addition, (iii) Bjerrum curves gave highlights about phenomena under kinetic control. In that context, time- or pH-induced conversion within different polymorphs could be unveiled, as proved by sulindac and dipyridamole case studies. On-line techniques (*e.g.* Raman probe) should ideally allow to characterise the solid forms in suspension, while off-line characterisation of the precipitate required to mock-up the assay to collect enough material for analysis. Finally, investigating solution-mediated interconversions requires the preliminary characterisation of polymorphic forms of the analyte that is usually available when the candidate only has access to the development phase.

8. Effect of composition of simulated intestinal media on solubility, supersaturation and solid state of felodipine investigated by design of experiment

8.1 Aim of the work

Solubility in the gastro-intestinal tract may limit absorption and the subsequent therapeutic effect of orally administered drugs. Since the introduction of the first simulated intestinal fluid for the fasted state [118] (FaSSIF), the use of that kind of media for *in vitro* studies has gained increasing attention [153, 154, 155], as discussed in the Introduction of this thesis. Recent literature has been emphasized that simulated intestinal fluid (SIF) should account for intra- and inter-individually variability reported for human intestinal fluid composition [156, 157, 158, 159, 160, 161,162, 163, 164, 165, 166, 167].

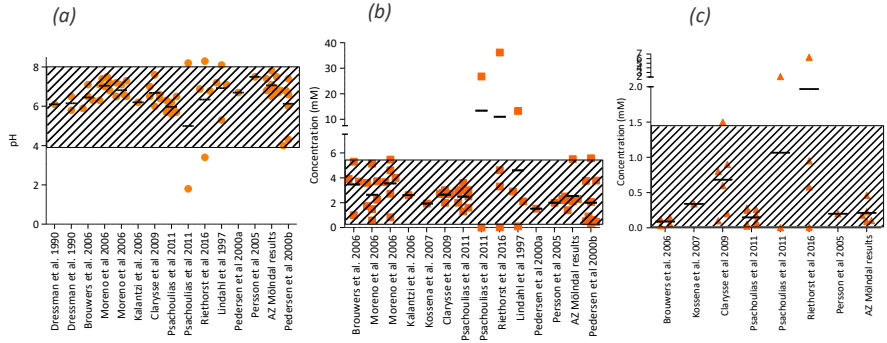


Figure 52: Variability of human intestinal fluid samples reported in recent literature. (a) pH, (b) bile salt, (b) phospholipid.

In this context, Design of Experiment (DoE) is a useful tool to provide a comprehensive view of the role of different solubilizing factors and their interactions, carrying a still manageable number of experiments in comparison to conventional solubility studies. That approach was successfully applied by Madsen and co-authors [122] whose work set the starting point for the present study.

The purpose of this part of the research project was to investigate the impact of intestinal media compositions on solubility and supersaturation propensity of felodipine, a neutral slightly soluble drug, in the set of simulated intestinal media defined in Madsen's paper, using µDiss Profiler™ (Pion Inc, California, USA).
Taking advantage from the larger scale of the experiment, in comparison to turbidimetric assays performed on Sirius T3, it was made an attempt characterise the solid state of the precipitate resulting from supersaturation studies.
That was compared to the starting material, trying to enlighten the role played by pH, bile salt and phospholipid content on solution-mediated interconversion between different solid forms.

8.2 Experimental part

8.2.1 Materials

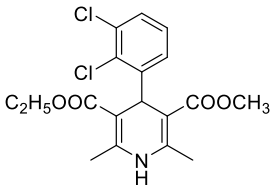


Figure 53: Felodipine. MW = 384.3. Albeit carrying a dihydropyridine moiety, the drug is regarded as neutral.

Felodipine was donated by AstraZeneca (Mölndal, Sweden). Glacial acetic acid, chloroform, 2-(N morpholino)ethanesulfonic acid (MES) hydrate, monobasic sodium phosphate and sodium taurocholate hydrate were purchased from Sigma-Aldrich (Missouri, USA). Sodium chloride (NaCl), sodium hydroxide (NaOH) and dimethylsulfoxide (DMSO) were purchased from Merck Millipore (Darmstadt, Germany). MES sodium salt was supplied by Acros Organics (Geel, Belgien). 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), was purchased from Thermo Fisher (Pennsylvania, USA). Soy derived phospholipid (Lipoid S-PC) was purchased

from Lipoid GmbH (Ludwigshafen, Germany). SIF Powder Original was purchased by Biorelevant.com (South Croydon, United Kingdom). Aqueous solution were prepared by using milli-Q water (SG Ultra Clear UV 2002, Evoqua water Technologies LLC, Barsbüttel, Germany).

8.2.2 Apparatus

DoE on simulated intestinal fluids was obtained by MODDE software (v. 12, Umetrics, Umeå, Sweden) that was also used for regression analysis.
Solubility and supersaturation studies were performed in µDiss profiler™ interfaced with Au PRO software (Pion Inc, Ca, USA). Prism version 5 (GraphPad Software Inc., California, USA) was used to plot results from both solubility and supersaturation studies.
X'Pert PRO X-ray diffractometer interfaced to a PIXcel detector (Almelo, The Netherlands) was employed for X ray powder analysis. X'Pert Data Collector and X'Pert HighScore (PANalytical B.V., Almelo, The Netherlands) were used to acquire and view the data.
A Kaiser RXN1 Microprobe (Kaiser Optical Systems, Ann Arbor, MI, USA) allowed collecting Raman spectra while data processing and analysis were performed in Microsoft Office Excel 2013 and in R Core Team (2015) (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Wien, Austria).

8.2.3 Methods

8.2.3.1 Design of Experiment: the model for solubility and solid state data

Based on Madsen's work [122], DoE on simulated intestinal fluids (SIF) was obtained by MODDE software (Umetrics, Umeå, Sweden). The following model was defined:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1.2} x_1 x_2 + \ldots + \beta_{1s} x_1^2 + \beta_{2s} x_2^2 + \ldots + \varepsilon$$

Equation 94

where y is the response (solubility or crystal likelihood); β_0 is a model constant; β_n are the coefficients of the x_n independent factors or the interaction between them; β_{ns} are the coefficients for x_n^2 (squared factors) and ε is the estimated error of prediction.

Squared parameters were considered in the fitting of Raman data, while for solubility results only linear and interaction factors were taken into account.

x_n was properly scaled to allow a straightforward comparison of β coefficients, which parameterized the impact of each factor within the model.

pH (PH), bile salt concentration (BS, sodium taurocholate hydrate), phospholipid concentration (PL, Lipoid S-PC) and interactions among them were taken into account. Considering composition of human intestinal fluids from healthy volunteers reported in literature, levels of the three factors where set as in Table 31, including a constraint: the bile salt over phospholipid concentration ratio (BS/PL) was forced to be lower than or equal to 35, to be physiologically relevant.

Factor	Abbreviation	Range	FaSSIF*	FaSSIF-V2*	FaSSIF-V2plus*
pH	PH	4.0 – 8.1	6.5	6.5	/
bile salt (mM)	BS	0.36 – 5.57	3	3	3
phospholipid (mM)	PL	0.025 – 1.5	0.75	0.2	0.2
bile salt/ phospholipid	BS/PL	≤ 35	4	15	15

*Table 31: Ranges used in DoE. Constraint: bile salt (BS) over phospholipid (PL) concentration ratio was kept ≤ 35, to be physiologically relevant. * ranges reported in Fuch's paper [119].*

Using D-optimal design and 2-factors interaction model, the software generated 10+3 SIF media (N1-N13). Lipoid S-PC has been widely employed as phospholipid in simulated intestinal fluids [112]. Taurocholate sodium salt hydrate was used since already available; commercial mixture of bile salts was avoided because the yellow colour of the resulting solution would

have interfered with the UV-spectrum of felodipine. It should be underlined that in the context of the DoE approach the choice of bile salt was not crucial.
Different buffer systems were employed with respect to the target pH of each medium: glacial acetic acid (pKa 4.75 at 37 °C (Sigma-Aldrich, 2017)) was used for media with pH 4.00; MES (pKa 5.97 at 37 °C (Sigma-Aldrich, 2017)) was used in media with pH 6.05 and HEPES (pKa 7.31 at 37 °C (Sigma-Aldrich, 2017)) was used in media with pH 8.1. Thus, the good experimental practice of being within ± 1 units of the pKa of the buffer was fulfilled.
Media N11-N13, having the same composition, were considered to test the variability of the experimental method and the general error of the model.

8.2.3.2 Preparation of SIFs

1 L of each SIF was prepared in blue-cap flask by combining different components as follows.
First, a stock solution of phospholipid 100 mM was made in chloroform and relevant buffer system (acetic acid, MES or HEPES) was prepared in milli-Q water.
A proper volume of the stock solution of phospholipid was transferred in the flask and the solvent was evaporated under a stream of nitrogen until a lipid film was formed on the bottom of the flask.
Bile salt was added together with the previously made buffer solution; finally, sodium chloride was added to adjust osmolality at 218 ± 10 mOsm. The pH was checked after equilibrating the media overnight or at least for 4 h and adjusted as needed with 1 N sodium hydroxide or 1 N hydrochloric acid to minimize dilution of the SIF.
Commercial freeze-dried powder (SIF Powder Original by Biorelevant.com) was restored as specified by the manufacturer.
Phosphate buffer was prepared with 3.954 mg/mL of monobasic sodium phosphate, 0.420 mg/mL of NaOH, and 6.186 mg/mL of NaCl dissolved in milli-Q water and adjusted to pH 6.5. SIF Powder was dissolved (2.24 mg/mL) in the phosphate buffer and stirred for 2 hours at room temperature before use. The resulting FaSSIF was stored at room

temperature, and all studies were completed within 48 h after preparation according to the manufacturer guidelines.

8.2.3.3 Solubility studies

Medium	DMSO stock (mM)	Wavelength (nm)	Probetip length (mm)
N1	5	342-435	5
N2	5	342-435	5
N3	10	342-435	2
N4	20	342-435	2
N5	20	342-435	2
N6	20	342-395	2
N7	20	342-435	5
N8	20	342-435	2
N9	20	342-435	2
N10	20	392-435	2
N11-N13	20	392-435	5

Table 32: Starting DMSO stock solution, light pathlength and wavelength used for solubility and supersaturation studies.

Solubility and supersaturation studies in the 13 media were performed in the µDiss profiler™ (Pion Inc, Ca, USA), optimizing experimental conditions for each SIF composition. 5, 10 or 20 mM stock solution of felodipine in DMSO were made in advance for this purpose.
Calibration curve was acquired for each channel adding aliquots of the suitable DMSO stock to 10 mL of medium maintained at 37 °C; the total amount of DMSO was kept below 1% to not affect drug concentration.

UV-spectrum was recorded and absorbance in the range 342–435 nm was related to the corresponding concentration. Agitation was ensured by cross-section magnets stirring at a speed of 100 rpm. 2 mm or 5 mm optic probe tips were chosen according to the turbidity of the medium as well as to the concentration reached by felodipine in solution.
After calibration, solubility was determined in each medium starting from powder, running experiments overnight in triplicates. After 16 h water concentration profile had reached a plateau that was regarded as thermodynamic solubility.
The second derivative transformation of UV-spectrum was performed through Au PRO software and used to account for particle scattering.

8.2.3.4 Supersaturation studies

Superstauration was assayed by applying solvent-shift method as described by Palmelund at al. [139], with one relevant change. Increasing concentration of felodipine were tested by adding from 10 to 100 µL of the DMSO stock to fresh medium each time, instead of iteratively adding a certain stock volume to the same SIF volume. Since supersaturation is regarded as a kinetic phenomenon, replacement of the medium after each addition of the stock ensured time to not affect the result.
Precipitation was detected as a scatter in baseline of the UV-spectrum and a subsequent deviation from linearity of the calibration curve. The highest concentration reached without observing instantaneous precipitation was chosen as maximum supersaturation concentration (C_{ss}).
Four stock solutions were then made in order to give a final concentration equal to 100% ($C_{ss100\%}$), 87.5% ($C_{ss87.5\%}$), 75% ($C_{ss75\%}$), and 50% ($C_{ss50\%}$) of C_{ss} when 200 µL of the stock solution were spiked in 10 mL of SIF. Each concentration was tested as triplicates; stirring mode and temperature were set as for solubility studies (cross section magnets, 100 rpm).
After adding 200 µL of stock into the medium, experiments were run for 60 min. Also in this case, second derivative of UV-spectrum was used to minimize signal interference induced by precipitation.

Precipitation rate (PR) was calculated for 100% and 87.5% C_{ss} in Microsoft Office Excel 2013 as follows. First, C_{mean} was obtained as the average of maximum concentration reached at supersaturation (C_{ss}) and concentration at equilibrium after precipitation of the excess solid ($C_{equil.\ sol.}$):

$$C_{mean} = \frac{(C_{ss} + C_{equil.\ sol.})}{2}$$

Equation 95

Then 3 to 5 points were considered above and below that concentration and they were interpolated by linear regression. The slope of that curve was assumed as precipitation rate (PR).

8.2.3.5 Solid state analysis by XRPD

Supersaturation assays in all media were scaled up to obtain a suitable amount of precipitate for solid state characterisation. 200 µL of 20 mM stock solution of felodipine in DMSO were added to 20 mL of SIF in a falcon tube and kept at 37 °C for 10 min, or until precipitation was detected. Tubes were then centrifuged (5000 rpm for 20 min at 37 °C) and supernatant was removed.

Resulting wet slurry was split into aliquots and analysed right after preparation, in addition to starting powder that was taken as reference. X-ray powder diffraction was performed at ambient conditions over the range of 4°-35° 2θ angles, using Cu Kα radiation (λ = 1.54187 Å) with a scan rate of 0.06734 2θ/s. Voltage and current were set at 45 kV and 40 mA respectively.

8.2.3.6 Solid state analysis by Raman

A second aliquot of each slurry was assayed by Transmission Raman spectroscopy.

A sample of the starting powder, identified as crystal form I by XRPD analysis [168] was amorphisized by quench-cooling method [169]. Spectra of both the amorphous and crystal forms were assumed as reference for data analysis. All the samples were assayed with a PhaT-probe. Dark frames was collected and subtracted after each measurement. Scattered light was recorded using 1x PhAT-probe objective, with 1 sec exposure and 5 sec accumulation. The excitation power was 200 mW at the fibre output and the wavelength was 785 nm. The Raman shift from -61 to 1841 cm^{-1} was measured, with a resolution of 0.3 cm^{-1}.

8.3 Results and discussion

8.3.1 Solubility studies

Solubility data in different SIF compositions are shown in Figure 54.

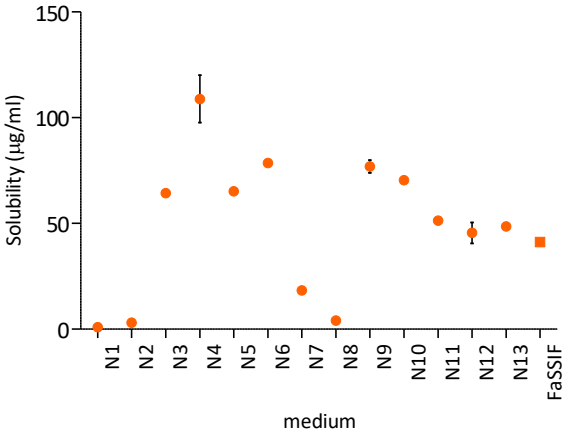


Figure 54: Scatter plot of solubility data in different SIF compositions. Experimental solubility in FaSSIF prepared using SIF powder (Biorelevant.com) is included as reference.

The values spanned between 0.90 ± 0.02 and 109 ± 11 µg/mL, while a concentration of 41.2 ± 0.1 µg/mL was found in FaSSIF medium prepared from commercially available powder by Biorelevant.com. The replicates of

the centre point matched well, proving variability of the method to be acceptable.

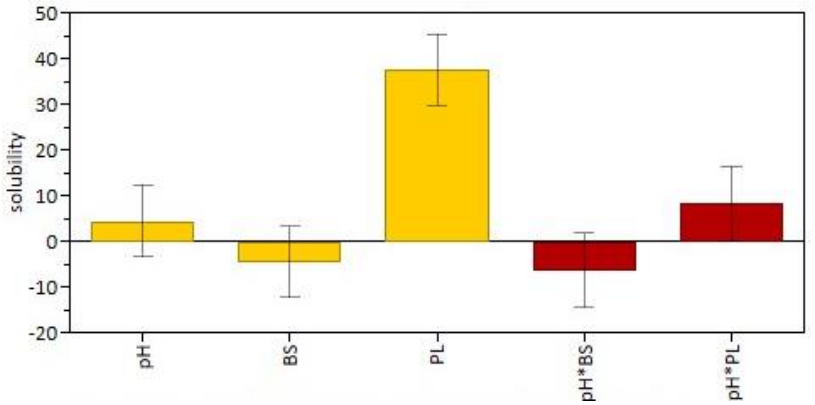


Figure 55: MODDE model for solubility data of felodipine (n = 13, df = 7, R^2 = 0.95, s = 9.76, F = 27.2, Q^2 = 0.76).

To describe the impact of each solubilizing factor, interaction model generated by MODDE software was used to fit experimental data. The heigth of the columns could be directly compared since independent viariables were centered and scaled.

Except for BS*PL interaction, which was found to be not relevant, all the other factors were included in the model. It provided well fitting, good descriptive power (R^2 = 0.95) and acceptable predictivity (Q^2 = 0.76).

Those data were compared to the ones obtained for the 24-SIFs DoE reported in Madsen's paper [122]. However, it should be taken into account that the ranges adopted in the previous work were different (*i.e.* pH 5.6-7.8; BS 1.40-5.90 mM; PL 0.093-0.6 mM) from the ones chosen for the present design.

In agreement with the previous study, this model confirmed the remarkably positive effect of phospholipid, but, on the contrary, pointed out a slightly negative impact of bile salt. It is worthy note that the influence of bile salt on solubilisation of low-soluble drugs is not always straightforward to predict. In a highly compound-dependent manner, interaction between taurocholate and an active pharmaceutical ingredient

(API) could result in a more as well as a less soluble salt [170] than the free drug.
Added to that, this model highlighted pH to be a relevant factor, even for a neutral drug as felodipine, not actually by itself, but almost in combination with bile salt (pH*BS) and phospholipid (pH*PL).
In particular, the positive effect of phospholipid was magnified at higher rather than lower pH. Characterisation of commercial FaSSIFs pointed out that phospholipid combines with bile salt to give mixed micelles [171] that keep poorly water-soluble drugs in solution.
Bile salt conjugated with taurine are sulfonic acids whose strength is decreased in micellar systems in comparison to monomeric solution [172], so that it can be speculated to be less soluble at acidic rather than neutral or alkaline pH. Since taurocholate could be regarded as freely soluble between pH 6 and 8 [173], formation of mixed micelles and consequent solubilisation of felodipine should be promoted in those conditions.

8.3.2 Supersaturation studies

Supersaturation propensity, expressed as C_{ss}, spread from 30 to 392 µg/mL (that corresponds to the DS range 1.1-33.1), as a result of different composition of the media. In 10 over 13 SIFs (*i.e.* all media except for N3, N4 and N9) precipitation curves showed an expected shape (example reported for N1 in Figure 56): generally, a supersaturated state in which concentration exceeds thermodynamic solubility was initially reached. Being intrinsically unstable, that system spontaneously reverted to equilibrium solubility (*i.e.* more energetically favourable condition) by precipitating the excess compound.
In turbid media N3 (reported as example in Figure 56), N4 and N9 concentration smoothly decreased to thermodynamic equilibrium solubility and sigmoidal supersaturation-precipitation curves were not displayed. Moreover concentration in N2, N6, N8 and N11-13 (Figure 56) showed persistent turbidity after stock addition a *(i.e.* system looked cloudy for a longer time than the one needed to generate a homogenous

solution by stirring) and an initial decrease followed by the classic supersaturation-precipitation pattern.

Trying to explain this behaviour, it can be speculated that liquid-liquid phase separation (LLPS) took place right after addition of felodipine DMSO stock into the SIFs leading to aggregation of the poorly soluble drug in droplets.

It should be recalled SIFs are complex media in which the aqueous phase, the drug-rich droplets and the solute rich-mixed micelles coexist. It was rationalized that LLPS in aqueous solutions of hydrophobic compounds displaying high melting points occurred in a metastable region of the phase diagram where concentration in solution was higher than crystalline solubility [174]. That was the condition observed in a supersaturated system in which LLPS may thus be a precursor to precipitation [175].

Looking at the curves it can be supposed an instantaneous (i.e. N2, N11-13) or rapid (i.e. N8) achievement of LLPS: solute was trapped in micelles up to saturation and then free solute concentration in the aqueous phase increased until precipitation was triggered [171]. Due to the high concentration of felodipine and the relative speed of precipitation, it might be expected the resulting solid to be amorphous or slightly crystalline (paragraphs **8.3.3** and **8.3.4** will cope with this topic).

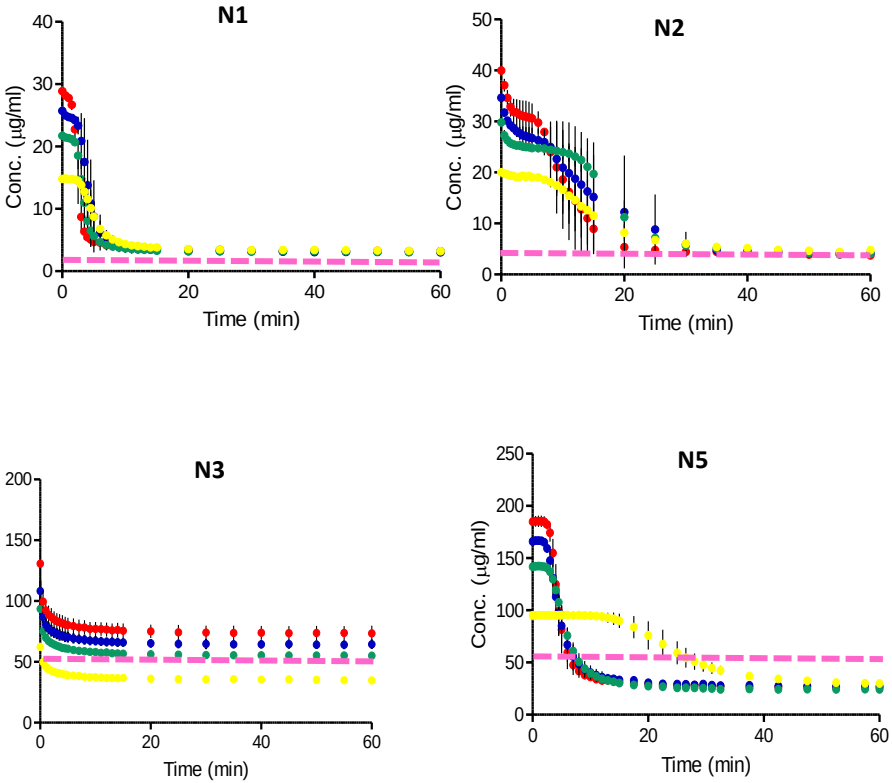
However, characterization of the micellar system of the media could help to cast light on that phenomenon; in addition to that on-line sampling and analysis by microscopy should be performed to prove LLPS took place.

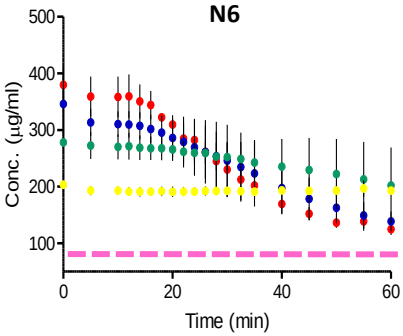
Generally speaking, supersaturation experiments hinted C_{ss} to follow a similar trend to solubility, but since turbid media were excluded by supersaturation data analysis only qualitative considerations should be made.

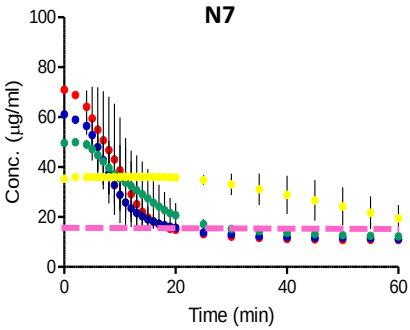
Time at which supersaturated concentration starts to decrease by meaning of precipitation is defined induction time (t_{ind}) [138]. Considering supersaturation/precipitation curves, in some media (*i.e.* N2, N6, N8, N10, N11-13) it can be noticed a weak relation between induction time and starting concentration: the higher the $\%C_{ss}$, the shorter the time supersaturation last before reverting to lower concentration. In the other

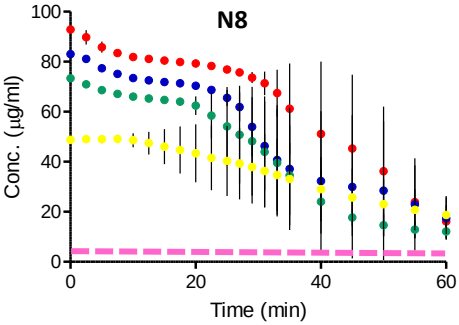
SIFs (*i.e.* N1, N5, N7), for at least 3 over 4 percentages (100%, 87.5% and 75% C_{ss}) this dependency could not be assessed.

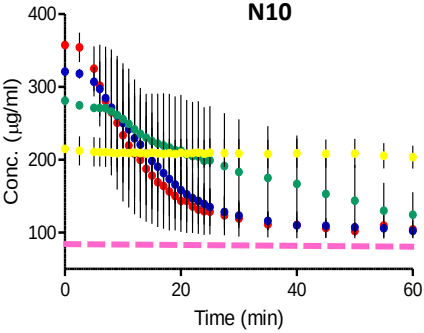
A remarkable variability was found in each medium within 3 replicates (see Precipitation rate ± SD in Table 33) and that was not surprising keeping in mind the stochastic and kinetic nature of supersaturation phenomenon. Herein considerations should be limited, again, to qualitative inspection of the curves. Considering $C_{ss100\%}$ and $C_{ss87.5\%}$, precipitation rate seemed C_{ss}-dependent at some extent in SIF N2, N6, N7, N11, N13.

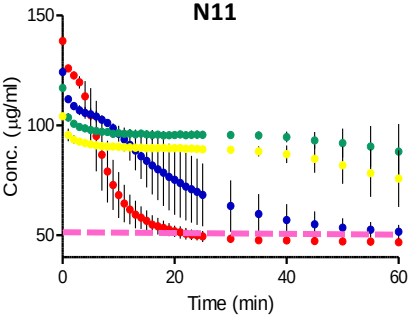


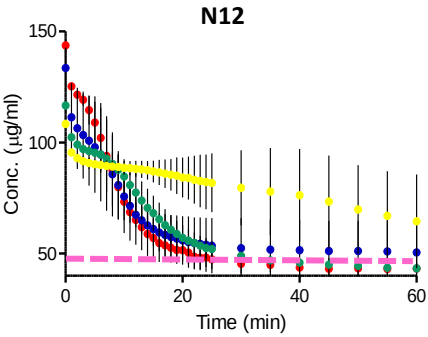

N6
Conc. (µg/ml)
500
400
300
200
100
0
20
40
60
Time (min)


N7
Conc. (µg/ml)
100
80
60
40
20
0
0
20
40
60
Time (min)


N8
Conc. (µg/ml)
100
80
60
40
20
0
0
20
40
60
Time (min)


N10
Conc. (µg/ml)
400
300
200
100
0
20
40
60
Time (min)


N11
Conc. (µg/ml)
150
100
50
0
20
40
60
Time (min)


N12
Conc. (µg/ml)
150
100
50
0
20
40
60
Time (min)

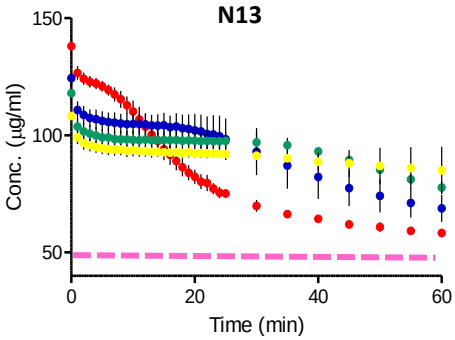


Figure 56: Supersaturation curve in DoE media. Tested concentrations: 100% Css (red dots), 87.5% Css (blue dots), 75% Css (green dots) and 50% Css (yellow dots). Each experimental point is reported as average of n = 3 replicates with SD.

8.3.3 Solid state analysis on wet slurries from supersaturation assays by XRPD

Since supersaturation did not occur as expected in media N3, N4 and N9, they were excluded from following XRPD analysis (results are reported in Table 33). Diffractogram of each sample was compared to the one of the starting powder (crystal form I) [168]: based on a qualitative analysis, the precipitate was classified as "crystalline" when any of the peaks of the reference was found in the diffractogram, even if the extent of crystallinity could not be estimated. On the contrary, when only a halo was present the precipitate was regarded as "amorphous".

Both forms were found among the 10 SIFs tested, thus proving composition to affect the solid state of the precipitate from supersaturation studies. Although correlation to solubilizing factors was hindered by the qualitative nature of data returned by such technique, XRPD could be combined to supersaturation results to further explain the shape of precipitation curves in SIFs and find a trend.

As reported in paragraph **1.4.8 Supersaturation and classical nucleation theory**, precipitation is conventionally described as a step-wise process that includes two main events: nuclei formation and crystal growth [140].

During the first stage, nuclei-precursors are continuously forming and breaking until a critical radius is reached. Above that radius, proper nuclei are generated leading to the step of crystal growth. Nucleation is rationalized by using classical nucleation theory, in which t_{ind} is assumed as inversely proportional to nucleation rate.

In the present case, felodipine uderwent a solution-mediated interconversion between crystal and amorphous form which was sensitive to pH, supersaturation extent and surfactants within the media [176]. Thanks to its di-hydropyridinic and carboxylic moieties, felodipine could establish hydrogen-bonds together with phospholipid and bile salt, which in turn can affect nucleation and precipitation.

Since felodipine precipitated as amorphous or slightly crystalline form in all SIFs, it can be speculated that classical nucleation theory did not apply to such conditions, that would explain lack of correlation between C_{ss} and t_{ind} or PR.

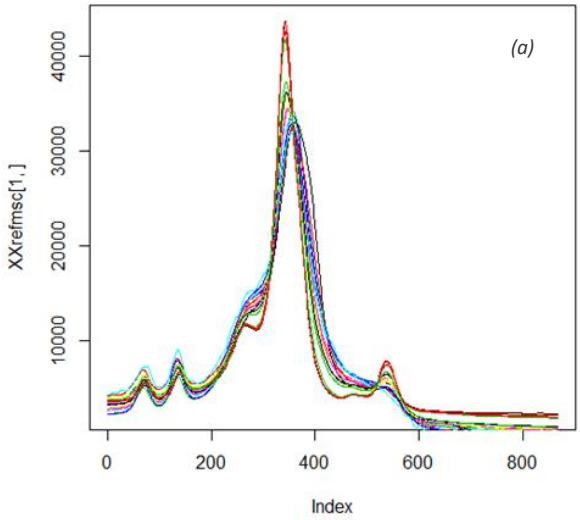
Medium	pH	BS (mM)	PL (mM)	BS/PL	Solid form by XRPD	Ppt rate (µg/min) ± SD	Crystal likelihood
N1	4	0.36	0.025	14.4	AMO	195 ± 21	0.27
N2	8.1	0.36	0.025	14.4	CRY	79 ± 19	0.08
N3	4	0.36	1.5	0.24	/	/	0.41
N4	8.1	0.36	1.5	0.24	/	/	0.96
N5	4	5.57	1.5	3.71	AMO	592 ± 290	/
N6	8.1	5.57	1.5	3.71	AMO	155 ± 18	/
N7	4	5.57	0.159	35.03	CRY	87 ± 16	0.16
N8	8.1	5.57	0.159	35.03	AMO	55 ± 18	0.33
N9	4	2.965	1.5	1.98	/	/	0.41
N10	6.05	5.57	1.5	3.71	CRY	247 ± 182	/

N11	6.05	2.547	0.6418	3.97	CRY	116 ± 18	0.67
N12	6.05	2.547	0.6418	3.97	CRY	96 ± 27	0.56
N13	6.05	2.547	0.6418	3.97	CRY	36 ± 12	0.86

Table 33: Media composition, solid form of the precipitate (AMO amorphous; CRY crystalline), precipitation rate (Ppt rate) calculated for 100% Css and crystal likelihood estimated from Raman spectra.

8.3.4 Solid state analysis on wet slurries from supersaturation assays by Raman spectroscopy

Supersaturation samples were freshly prepared as mentioned in the previous section right before collecting Raman spectra. Turbid media were included in this analysis to broaden the data set.



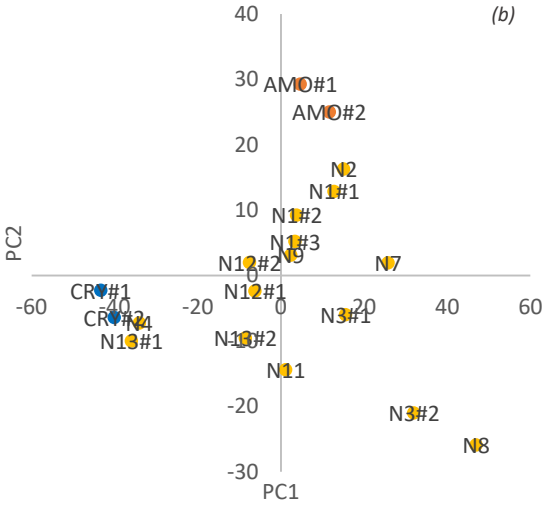


Figure 57: (a) Raman spectra within 1540 and 1800 cm^{-1} after MSC correction. (b) PCA on data refined with MSC correction. Only the significant range within 1540 and 1800 nm was considered and outliers N6 and N10 were excluded by preliminary PCA on raw data.

Samples from several media, such as N1, N3, N11 and N12, as well as reference forms (crystal form I: CRY, and amorphous: AMO), were prepared as duplicates to estimate if repeated precipitation in the same conditions returned consistent spectra.

The region with the most remarkable difference between crystal and amorphous was found within 1540 and 1800 cm^{-1}; thus spectra were truncated, pre-treated and analysed by using Principal Component Analysis (PCA) from R pls package [177] to allow the detection of outliers (*i.e.* N5, N6, and N10). For such media quality of the spectra was regarded as too low to allow further investigation.

Remaining spectra were pre-treated by using Multiplicative Scatter Correction (MSC function) [178] to correct baseline drift and a final PCA were performed on refined data (scores are plot in Figure 57*b*). Replicates from references, N1 and centre points N11-N13 clustered well except for N13#1, proving the correction effective and the repeatability acceptable among independent samples.

	PC1	PC2	PC3	PC4	PC5
SD	24.92	14.49	4.49	0.27	1.90
var%	0.72	0.24	0.02	0.01	0.00
cum. var%	0.72	0.96	0.98	0.99	0.99

Table 34: summary of PCA performed on MSC-corrected spectra after truncation; each principal component (PC) is reported with its standard deviation (SD), % of explained variance (%var) and cumulative percentage of explained variance (cum. var. %).

As reported in Table 34, variation within spectra was almost described by two principal components (72% of variation by PC1 and an additional 24% by PC2).

Apart from N4 and N13#1, which plotted close to crystalline reference, all the other samples spread in the score plot hinting that a certain contribution of amorphous limit form was found in each precipitate.

Discriminant analysis was then performed by using Microsoft Office Excel 2013, in an attempt to estimate which factors affected solid state of the precipitate. y = 1 and y = 0 were assigned to crystal and amorphous respectively and considered as reference. Linear interpolation was calculated by least square regression method and coefficients of such curve were employed in the calculation of a new y value for each sample. The output was proportional to the proximity towards CRYs or AMOs and referred as “crystal likelihood”; values were then imported in MODDE software as y responses and the model reported in paragraph **8.2.3.1** was used to fit the data. Not significant factors and interactions were removed and a simplified model was obtained (see Figure 58).

Multi-linear regression on Raman spectra (n = 10) revealed phospholipid and pH to positively affect crystallinity of the sample both individually and in combination. To reach a good fitting, the squared term pH*pH was included after verifying the confounding effect to be insignificant (correlation with the interaction factor pH*PL returned r = 0.11).

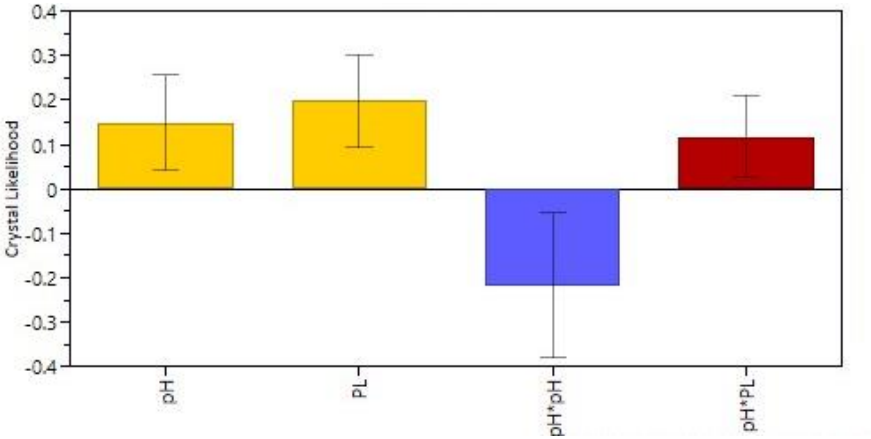


Figure 58: MODDE model for Raman data of felodipine (n = 10, df = 5, R^2 = 0.91, s = 0.12, F = 11.9, Q^2 = 0.35).

As discussed for solubility results in paragraph **8.2.3.3 Solubility studies**, the higher the pH in the range, the higher the proportion of mixed micelles that are generated into solution. Solubilisation of felodipine in micelles could promote concentration of the solute and therefore crystallization [179]; on the contrary, in a low-phospholipid SIF precipitation as amorphous rather than crystalline form occurred. *In situ* Raman analysis supported by microscopic analysis of the precipitate could lead to a morphologic characterisation of the particles.

Moreover, in order to define a predictive model of the solid form of the precipitate in supersaturated SIFs, a more rigorous calibration is required. In fact reference spectra were collected in dried condition that did not represent the complex mixtures of phospholipid and bile salt proper of simulated media.

8.4 Conclusions

Human intestinal fluids exhibit intra- and inter-individual variability that SIFs used for *in vitro* tests should represent. If solubility is the limiting step for absorption, DoE set up could be chosen as a useful tool to explore the impact of different composition of SIFs on solubility and supersaturation

propensity of orally administered drugs. Considering three factors (bile salt, phospholipid and pH) and interaction among them, D-optimal design space included a manageable number of media (10 + 3 centre points) in which both solubility and supersaturation assays were performed.

Solubility data covered a range of two order of magnitude (0.09 to 109 µg/mL) and were well fitted by a multi-linear interaction model ($R^2 = 0.95$) with acceptable predictive power ($Q^2 = 0.76$).

On the other hand, supersaturation studies returned some useful information, such as propensity of felodipine to undergo liquid-liquid phase separation, but only from a qualitative point of view.

Moreover, it was made an effort to investigate impact of different solubilizing factor on solid state of precipitate in different SIFs. After preliminary characterisation with XRPD, Raman spectroscopy was chosen as a more suitable technique for this purpose. Different composition proved to affect solid form of felodipine, and, in particular, relatively high phospholipid content appeared to promote crystallization together with medium or alkaline pH.

In conclusion present DoE turned out sub-optimal as regards supersaturation and subsequent analysis of the precipitate, hence requiring some major changes.

Recent literature suggests to update SIFs to lyso-phospholipid to better mimic intestinal condition, and to increase BS/PL ratio in order to reach critical micellar concentration [119].

Following these evidences, a new DoE has been developed and tested in the Pharmaceutical Department of KU Copenhagen in order to overcome limits of the design described in this chapter and to provide more consistent results concerning supersaturation propensity of poorly-soluble drugs in SIFs.

9. General conclusions and perspectives

It is widely accepted that the evaluation of the physico-chemical profile of a new chemical entity (NCE) allows scientists to predict and understand key aspects of the overall performance of an active molecule. Computational models have been increasingly sought to get information on structural features related to physico-chemical and biopharmaceutical properties, but the increasing molecular complexity displayed by candidates since the beginning of the new century may not be properly represented by commercial software training sets, thus leading to high errors in prediction. Conversely, experimental profiling of physico-chemical properties in discovery stage is still warranted with the double aim of implementing expert systems for prediction and of returning useful information for the following stages of research.

The overall aim of the present Ph.D. project was to set up and develop an operative workflow to experimentally profile relevant physico-chemical properties (*i.e.* ionization constants, lipophilicity and solubility) for proprietary small-molecule drug candidates at discovery and pre-development stage.

This approach was expected to be applied to new chemical entities (NCEs) that have passed early drug discovery phases (*i.e.* optimization of the synthesis) thus being freely accessible in solid form.

However, the workflow was supposed to include medium/high-throughput analytical methods suitable for those candidates that were exclusively available as DMSO stock solutions that needed a preliminary evaluation.

Before transferring the methods to actual drug candidates, those were tested on commercially available standards to evaluate their feasibility, advantages and drawbacks.

pKa assessment was performed by using Sirius GLpKA™ and Sirius T3™ instruments and strenght and weakness of potentiometric and UV-metric approaches were highlighted. The potentiometric method required higher operative concentrations than the UV-metric (10^{-4}-10^{-5} M versus 10^{-6}-10^{-7} M) and showed a lower accuracy for pKa measurements at the extreme of

the operative pH range of the glass electrode (pKa below 3 and higher than 11). On the contrary, UV-metric approach allowed to overcome such issues, but required the analyte to display chromophore(s) next to the ionisable site(s) (no more than n = 2-3 bond distance).

Micro-speciation of cetirizine was studied as a model of multiprotic compounds for which site-specific protonation constants were to be assigned. The deductive approach described by Pagliara et al. [6] was applied and it proved effective, albeit substance- and time-demanding. Thus, it was considered suitable only for selected NCEs that required deeper investigation for a specific purpose.

Moreover, acid-base equilibria of two commercial acids were investigated in a non-aqueous solvent by an *ad hoc* UV-spectrophotometric assay [152] which returned qualitative data. The obtained results were supported by parallel pH-metric assays in organic solvents/water mixtures and the overall approach allowed to estimate relative acidity strength of the two compounds in conditions that mimicked a non-aqueous formulation.

Potentiometric method implemented in Sirius GLpKa™ and T3™ equipment was applied to log *P* determination of ionisable compounds: this approach required a previous measurement of pKa whose accuracy deeply affected that of lipophilicity results. Compounds displaying several ionisation sites or, on the opposite, no ionisation site were not suitable for potentiometric log *P* determination and a semi-automated miniaturized shake-flask approach was set up and optimized for NCEs analysis in early discovery. Taking advantage from both automation and mass spectrometry detection, this approach allowed to determine log $D_{7.4}$ of those compounds that were only available as DMSO stock solutions, consuming a minimum amount of substance (< 40 µL of a 10 mM stock for n = 4 replicates). Added to that, the method could be regarded as medium-throughput (n = 4 compounds analysed as quadruplicates in less than 12 h) and hence suitable for discovery other than development phase. It is worthy noticing that the micro-shake flask workstation was found to be versatile and suitable for the set up of other different assay configurations: for instance, it could be implemented to test lipophilicity in different biphasic systems

or at different pH buffers in order to experimentally draw lipophilicity profiles of the NCEs. In that sense, the set up could also be optimized to perform kinetic solubility assays in 96-well plates.

Finally, solubility was assessed by turbidimetric methods by using Sirius T3™ and Pion µDiss profiler ™. The former allowed to discriminate between NCEs acting as "chasers" and "non-chasers", according to their ability to supersaturate under the assay conditions. Furthermore, analysis of the precipitation curves gave hints about solid form interconversions that were pH- or time-dependent as proved for sulindac and dipyridamole. Although meaningful, such approach required many attempts to find a suitable experimental set up and was substance consuming (2-20 mg/assay); moreover, it did not afford an easy off-line characterisation of the precipitate solid form.

Investigation of solid state interconversion in solution and of supersaturation was further developed during a 6-month research period at The University of Copenhagen, under the supervision of Prof. Anette Mullertz [180]. Taking advantage from the work previously done in the research group [122, 139], the impact of simulated intestinal fluid composition on solubility and supersaturation propensity of felodipine was assessed by the mean of µDiss Profiler™ which combined turbidimetric and spectrophotometric detection. To provide a comprehensive view of the role of different solubilizing factors and their interactions Design of Experiment was adopted. Precipitate from supersaturation assay was analysed by X-ray powder diffraction and Raman spectroscopy proving the composition of the media to induce partial amorphisation of the API.

Study of solubility and supersaturation processes could be potentially transferred to any biorelevant medium of interest to return useful data to supplement *in vitro-in vivo* correlation studies.

To sum up, the main goal reached by this Ph.D. project was the definition of an operative work-flow (see the decision trees in chapter **10**) that allowed to tailor from case to case the most suitable technique to each NCE in order to meet industrial practical needs and limitations. Starting from the straightforward application of known techniques, this study tried to

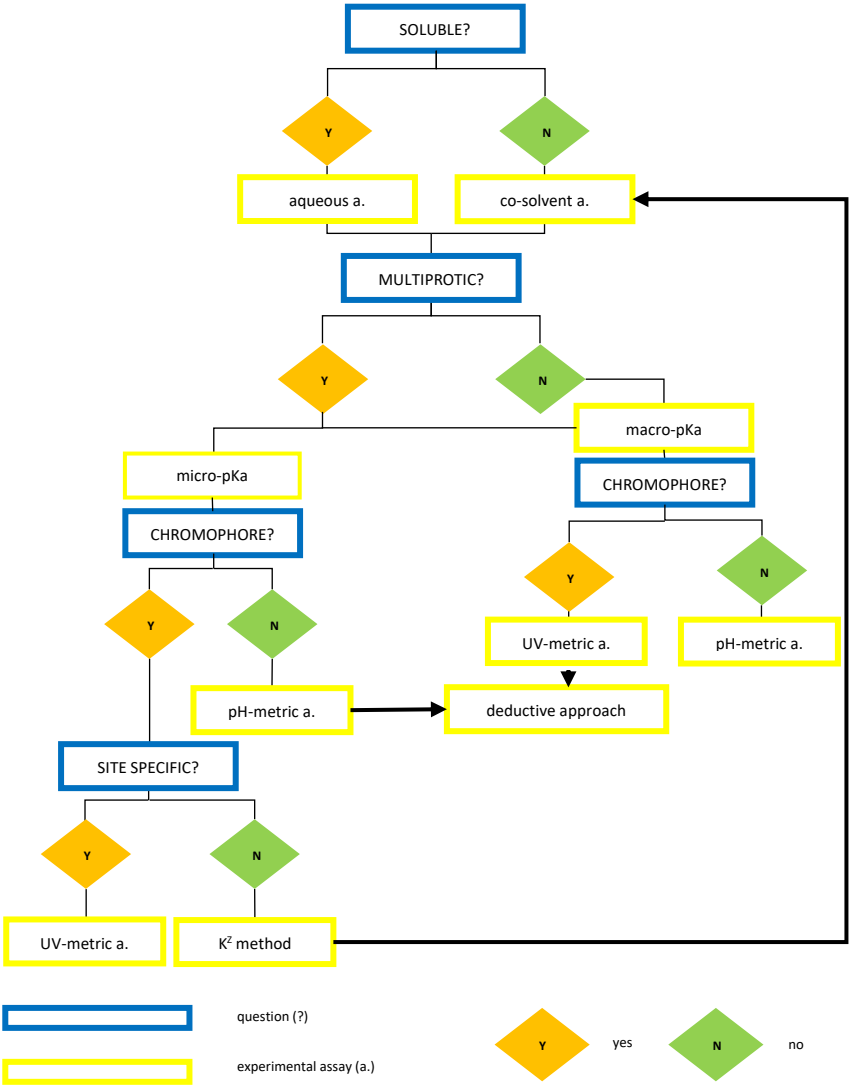
implement such approaches and investigate complex situations (e.g. from aqueous pKa to non-aqueous equilibria; from solubility in water environment to supersaturation and solid form interconversion in simulated intestinal media) resembling true cases that could occur in pre-clinical research.

Experimental data generated from these studies were used to implement expert systems for prediction of physico-chemical properties trying to fill the gap between computational and experimental approaches, which is still a shortcoming of overall pharmaceutical research.

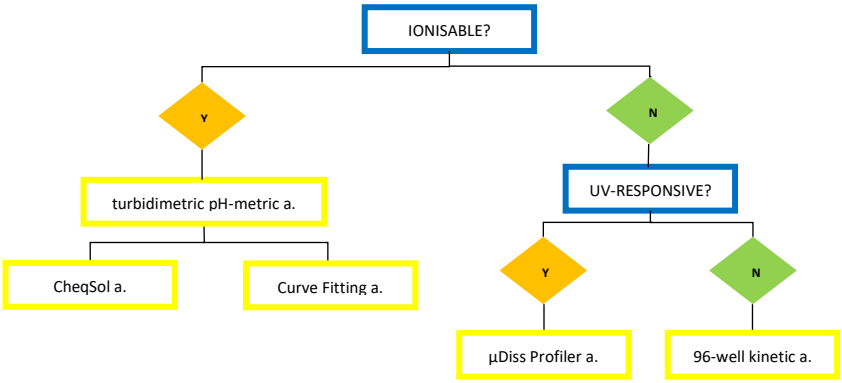
It should be highlighted that confidentiality of produced results prevented their publication in respect to Chiesi Farmaceutici S.p.A. legal policy.

10. Physico-chemical profiling of a NCE: decision trees

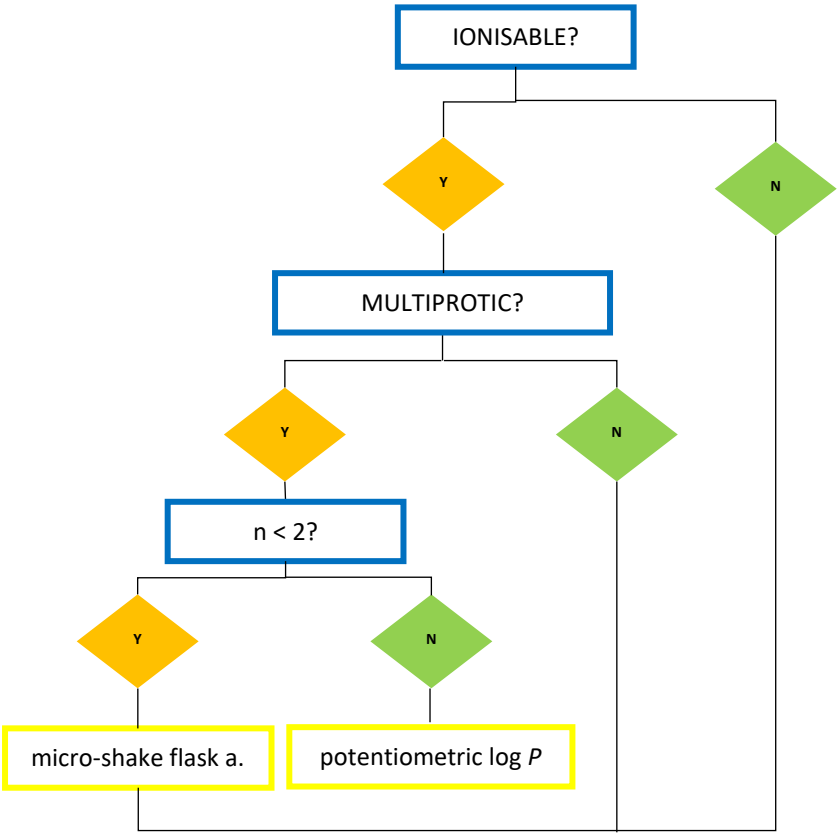
10.1 Dissociation constants determination



10.2 Solubility determination



10.3 Lipophilicity determination



Glossary

ADMET	absorption, distribution, metabolism, elimination and toxicity
AMOEBA	atomic multipole optimised energetics for biomolecular applications
ASD	amorphous solid dispersions
BCS	biopharmaceutics classification system
Bj	mean molecular charge
BMC	bio-partitioning micellar chromatography
BS	bile salts
CCC	counter current chromatography
CHI	chromatographic hydrophobicity index
Chol	cholesterol
CNT	classical nucleation theory
D-PAS	dip probe absorption spectroscopy
DS	degree of supersaturation
DSC	differential scanning calorimetry
ESI	electro-spray ionisation
EtOH	ethanol
FaSSCoF	fasted state simulated colonic fluid
FaSSGF	fasted state simulated gastric fluid
FaSSIF	fasted state simulated intestinal fluid
FaSSIF-V2	fasted state simulated intestinal fluid version 2
FeSSCoF	fed state simulated colonic fluid
FeSSGF	fed state simulated gastric fluid

FeSSIF	fed state simulated intestinal fluid
FeSSIF-V2	fed state simulated intestinal fluid version 2
GC	glycocholate
GCDC	glycochenodeoxycholate
GDC	glycodeoxycholate
GOF	goodness of fit
GPFX	grepafloxacin
GSE	general solubility equation
HIF	human intestinal fluid
HTS	high throughput screening
I	ionic strength
IAM	immobilized artificial membrane
ILC	immobilized liposome chromatography
IS	internal standard
Ksp	solubility product
K_W	autoprotolysis or self-dissociation constant of water
K^Z	tautomeric equilibrium constant
Lipoid S-PC	soy-derived phosphatidylcholine
LLE	ligand lipophilicity efficiency
Log $D_{7.4}$	(logarithm of) distribution coefficient at pH 7.4
Log *k*	logarithm of retention factor
Log *P*	(logarithm of) partition coefficient
Log *p*	(logarithm of) micro-partition coefficient
LVFX	levofloxacin
MeOH	methanol

MLC	micellar liquid chromatography
MMPA	matched molecular pair analysis
MOPS	*3-(N*-morpholino*)*propanesulfonic acid
MPS	multiple parallel synthesis
MW	molecular weight
NCE	new chemical entity
NMR	nuclear magnetic resonance
n-oct	*n*-octanol
OECD	organization for economic co-operation and development
PCA	principal component analysis
PEG	polyethyleneglycol
PK	pharmacokinetics
pKa	(negative logarithm of) dissociation constant
pka	(negative logarithm of) micro-dissociation constant
PL	phospholipids
p_oKa	apparent pKa in presence of octanol
PR	precipitation rate
p_sKa	apparent pKa in presence of co-solvent
QCTFF	quantum chemical topology force field
QSAR	quantitative structure activity relationships
QSPR	quantitative structure property relationships
RMSD	root mean square deviation
Ro5	rule of five
RP-HPLC	reversed-phase high-performance liquid chromatography
RP-TLC	reversed-phase thin layer chromatography

RT	retention time
S_0	intrinsic solubility
S_a	apparent solubility
SDDS	supersaturated drug delivery systems
SDS	sprayed dried systems
SEDDS	self emulsifying drug delivery systems
SEIF	simulated endogenous intestinal fluid
SGF	simulate gastric fluid
SIF	simulated intestinal fluid
SO	sodium oleate
S_t	thermodynamic solubility
TC	taurocholate
TCDC	taurochenodeoxycholate
TDC	taurodeoxycholate
TFA	target factor analysis
THF	tetra-hydro furan
Tm	melting point (°C)
UPLC-MS/MS	ultra-high pressure liquid chromatography tandem mass spectrometry
XRPD	X-ray powder diffraction
ΔG	Gibb's free energy
ΔH	enthalpy (contribution to the free energy of a system/process)
ΔS	entropy (contribution to the free energy of a system/process)
ε	dielectric constant

References

1 Andres A. et al., Eur. J. Pharm. Sci. 2015, 76: 181–191

2 Sugano K. et al., Drug Metab. Pharmacokinet. 2007, 22(4): 225–254

3 Lin B. et al., J. Pharm. Biomed. Anal. 2016, 122: 126–140

4 Oprea T. I., J. Computer-Aided Mol. Des. 2002, 16: 325–334

5 Hann M. M., Med. Chem. Commun. 2011, 2: 349-355

6 Pagliara A. et al., J. Med. Chem. 1998, 41: 853-863

7 Pagliara A. et al., Chem. Rev. 1997, 97: 3385-3400

8 Chen C. et al., Curr. Med. Chem. 2008, 15: 2173-2191

9 Leeson P. D. et al., Nat. Rev. 2007, 6: 881-890

10 Waring M. J. et al., Nat. Rev. Drug Discov. 2015, 14(7): 475–486

11 Lipinski C. A. et al., Adv. Drug Deliv. Rev. 1997, 23: 3-25

12 Nat. Rev. Drug Discov. 2007, 6: 853

13 Griesenauer R. H. and Kinch M. S., Adv. Drug Del. Rev. 2016, 101: 62-74

14 Lipinski C. A., Adv. Drug Del. Rev. 2016, 101: 34–41

15 Choy Y. B. and Prausnitz M. R., Pharm. Res. 2011, 28: 943–948

16 Hansch C., Bjorkroth J. P. and Leo A., J. Pharm Sci. 1987, 76: 663–687

17 Mignani S. et al., Drug Discovery Today 2016, 21(4): 573-584

18 Tian S. et al., Adv. Drug. Del. Rev. 2015, 86: 2-10

19 Settimo L. et al., Pharm. Res. 2014, 31: 1082–1095

20 Cheng T. J. et al., J. Chem. Inf. Model. 2007, 47: 2140–2148

21 Ten Brink T. and Exner T. E., J. Comput. Aided Mol. Des. 2010, 24: 935–942

22 Thomae A. V., Biophys. J. 2005, 89: 1802–1811

23 Gotink K. J., Clin. Cancer Res. 2011, 17(23): 7337-7346

24 Shelley J. C., Comput. Aided Mol. Des. 2007, 21: 681–691

25 Liao C. and Nicklaus M. C., J. Chem. Inf. 2009, 49: 2801-12

26 Caron G. and Edmondi G., Drug Discovery Today 2017, 22(6): 835-840

27 Albert A. and Serjeant E. P., The Determination of Ionization Constants, 3rd edn., Chapman and Hall, London, 1984

28 Avdeef A., Quant. Struct. Act. Relat. 1992, 11: 510–517

29 Avdeef A. et al., Anal. Chem. 1993, 65: 42–49

30 Slater B. et al., J. Pharm. Sci. 1994, 83: 1280–1283
31 Takács-Novák K. et al., J. Pharm. Biomed. Anal. 1994, 12: 1369–1377
32 Takács-Novák K. and Avdeef A., J. Pharm. Biomed. Anal. 1996, 1: 1405–1413
33 Sweeton F. H. et al., J. Solution Chem. 1974, 3: 191–214
34 Yasuda M., Bull. Chem. Soc. Jpn. 1959, 32: 429–432
35 Shedlovsky T., in: B. Pesce (Ed.), Electrolytes, Pergamon, New York, 1962
36 Avdeef A. et al., J. Pharm. Biomed. Anal. 1999, 20: 631-641
37 Takács-Novák K. et al., Int. J. Pharm. 1997, 151: 235–248
38 Allen R. I. et al., J. Pharm. Biomed. Anal. 1998, 17: 699–712
39 Malinowsky E. R., Anal. Chem. 1977, 49(4): 612-617
40 Woodruff H. B., Anal. Chem. 1981, 53: 81-84
41 Malinowski E. R., J. Chemometr. 1987, 1: 33–40
42 McCue M. and Malinowski E. R., Appl. Spectrosc. 1983, 37: 463–469
43 Szakács Z. et al., J. Math. Chem. 1999, 26: 139-155
44 Takács-Novák K. et al., J. Pharm. Biomed. Anal. 2000, 21: 1171–1182
45 Yang Z. et al., Future Med. Chem. 2016, 8(18): 2245-2262
46 Hirano T. et al., Biochim. Biophys. Acta 2006, 1758: 1743–1750
47 Judicka et al., Chem. Phys. Lett. 2018, 701: 58-64
48 Hongmao S., in: Elsevier (Ed.), A Practical Guide to Rational Drug Design, Woodhead Publishing, 2016, pp. 193–223
49 Bhal S. K. et al., Mol. Pharmaceut. 2007, 4: 556–560
50 Bannan C. C. et al., J. Chem. Theory Comput. 2016, 12: 4015−4024
51 Leo A., Hansch C. and Elkins D., Chem. Rev. 1971, 7(6): 525-616
52 Frank H. and Evans M., J. Chem. Phys. 1945, 13: 507-532
53 M. J. Kamlet et al., J. Org. Chem. 1983, 48: 2877-2887
54 M. J. Kamlet et al., J. Phys. Chem. 1988, 92: 5244-5255
55 Wassvik C. M. et al., J. Med. Chem. 2008, 51: 3035–3039
56 Gursoy R. N. and Benita S., Biomed. Pharmacother. 2004, 58(3): 173-182
57 Dressman J. B. et al., Pharm. Res. 1998, 15(1): 11-22
58 Andríc F. and Héberger K., J. Pharm. Biomed. Anal. 2015, 115: 183–191
59 Yalkowsky S. H. et al., J. Pharm. Sci. 1983, 72: 866–870
60 Pinsuwan S. et al., J. Chem. Eng. Data 1995, 40: 623–626

61 Ran Y. et al., Chemosphere 2002, 48: 487–509
62 Mortenson P. N. and Murray C. W., J. Comput. Aided Mol. Des. 2011, 25: 663–667
63 Xing L. et al., J. Chem. Inf. Comput. Sci. 2002, 42: 796–805
64 Finizio A. et al., Chemosphere 1997, 34: 131–161
65 Machatha S. G. and Yalkowsky S. H., Int. J. Pharm. 2005, 294: 185–192
66 Schroeter T. et al., Mol. Pharmaceut. 2007, 4: 524–538
67 Cheng T. J. et al., J. Chem. Inf. Model. 2007, 47: 2140–2148
68 Tetko I. V. et al., J. Chem. Inf. Comput. Sci. 2001, 41: 1407-1421
69 Tetko I. V. and Tanchuk V. Y., J. Chem. Inf. Comput. Sci. 2002, 42: 1136-1145
70 Viswanadhan V. N. et al., J. Chem. Inf. Comput. Sci. 1989, 29: 163-172
71 Ghose A. K., J. Comput. Chem. 1988, 9: 80-90
72 Moriguchi I. et al., Chem. Pharm. Bull. 1992, 40: 127-139
73 Moriguchi I. et al., Chem. Pharm. Bull. 1994, 42: 976-978
74 Csizmadia F. et al., J. Pharm. Sci. 1997, 86: 865-871
75 Hilal S. H. et al., QSAR Comb. Sci. 2004, 23: 709-720
76 Kier L. B. and Hall L. H., J. Chem. Inf. Comput. Sci. 1997, 37: 548-552
77 Wang R. et al., J. Chem. Inf. Comput. Sci. 1997, 37(3): 615-621
78 Kah M. and Brown C. D., Chemosphere 2008, 72: 1401–1408
79 Liang C. and Lian H., Trends Anal. Chem. 2015, 68: 28–36
80 Danielsson l. G. and Zhang Y. H., Trends Anal. Chem. 1996, 15(4): 188-196
81 Hartmann T. and Schmitt J., Drug Discov. Today Technol. 2004, 1: 431-439
82 Poole S. K. and Colin F. Poole C. F., J. Chromatogr. B 2003, 797(1-2): 3-19
83 Escuder-Gilabert L. et al., J. Chromatogr. B 2003, 797(1-2): 21-35
84 Yamagami C. et al., Chem. Pharm. Bull. 1994, 42: 907–912
85 Giaginis C. and Tsantili-Kakoulidou A., J. Liq. Chromatogr. Relat. Technol. 2007, 31: 79–96
86 Valko K. et al., J. Chromatogr. A 1993, 656: 501–520
87 Valko K. et al., Anal. Chem. 1997, 69: 2022–2029

88 Krass J. D. et al., Anal. Chem. 1997, 69: 2575–2581
89 Liu X. et al., Pharm. Res. 2011, 28: 962–977
90 Buchwald P. and Bodor N., Curr. Med. Chem. 1998, 5: 353-380
91 Morikawa G. et al., J. Pharm. Biomed. Anal. 2016, 117: 338–344
92 Dohta Y. et al., Anal. Chem. 2007, 79: 8312-8315
93 Alelyunas Y. W. et al., J. Chromat. A 2010, 1217: 1950–1955
94 Diamond R. M., J. Phys. Chem. 1963, 67: 2513-2517
95 Takács-Novák K. et al., Pharm. Res. 1999, 16(10): 1633-1638
96 Avdeef A., J. Pharm. Sci. 1993, 82: 183-190
97 Avdeef A. et al., J. Pharm. Sci. 1995, 84: 523-529
98 Slater B. A. et al., J. Pharm. Sci. 1994, 83: 1280-1283
99 Comer J. et al., SAR QSAR Environ. Res. 1995, 3(4): 307-313
100 Takács-Novák K. and A. Avdeef, J. Pharm. Biomed. Anal. 1996, 14: 1405-1413
101 Takács-Novák K. et al., J. Pharm. Biomed. Anal. 1994, 11: 1369-1377
102 Mazák K. et al., Eur. J. Pharm. Sci. 2011, 44: 68–73
103 Benet L. Z. et al., AAPS J. 2011, 13: 519-547
104 Bhattachar S. N., Drug Discovery Today 2006, 11(11-22): 1012-1018
105 Calvo N. L. et al., J. Pharm. Biomed. Anal. 2018, 147: 518-537
106 Jain S. et al., Drug. Dev. Ind. Pharm. 2015, 41(6): 875–887
107 Mannam Krishnamurthy M. and Naidu S. R. (2012). "8". In Lokeswara Gupta. Chemistry for ISEET - Volume 1, Part A (Ed.). Hyderabad, India: Varsity Education Management Limited, p. 298
108 Bergström C. A. S. and Larsson P., Int. J. Pharm. 2018, 540: 185–193
109 Di L. et al., Drug Discovery Today 2012, 17(9-10): 486-495
110 Di L. and Kerns E. H. 2007, Solubility Issues in Early Discovery and HTS. In: Augustijns P., Brewster M. (Eds.), Solvent Systems and Their Selection in Pharmaceutics and Biopharmaceutics. Biotechnology: Pharmaceutical Aspects, vol VI. Springer, New York, pp. 111-136
111 Faria T. N. et al., Mol. Pharm. 2004, 1: 67–76
112 Marques M. et al., Dissolut. Technol. 2011, 15-28
113 Vertzoni M. et al., Eur. J. Pharm. Biopharm. 2005, 60(3): 413–417
114 Jantratid E. et al., Pharm. Res. 2008, 25(7): 1663–1676

115 Galia E. et al., Pharm. Res. 1998, 15(5): 698–705
116 Vertzoni M. et al., Pharm. Res. 2010, 27(10): 2187–2196
117 Fuchs A. and Dressman J. B., J. Pharm. Sci. 2014, 103: 3398–3411
118 Dressman J. B. et al., Pharm. Res. 1998, 15(1): 11–22
119 Fuchs A. et al., Eur. J. Pharm. Biopharm. 2015, 94: 229–240
120 Khadra I. et al., Eur. J. Pharm. Sci. 2015, 67: 65–75
121 Perrier J. et al., Eur. J. Pharm. Sci. 2017, 111: 247–256
122 Madsen C. M. et al., Eur. J. Pharm. Biopharm. 2018, 111: 311–319
123 Murdande S. B., Pharm. Dev. Tech. 2011, 16(3): 187–200
124 Palucki M. et al., J. Med. Chem. 2010, 53: 5897–5905
125 Strickley, R. G., Pharm. Res. 2004, 21: 201–230
126 Jain N. and Yalkowsky S. H., J. Pharm. Sci. 2001, 90: 234–252
127 Skyner R. E. et al., Phys. Chem. Chem. Phys. 2015, 17: 6174--6191
128 Kühne R. et al., Chemosphere 1995, 30: 2061–2077
129 Baka E. et al., J. Pharm. Biomed. Anal. 2008, 46: 335–341
130 Nakashima S. et al., Chem. Pharm. Bull. 2013, 61(12): 1228–1238
131 Heikkilä T. et al., Int. J. Pharm. 2011, 405(1-2): 132-136
132 Alsenz J. and Kansy M., Adv. Drug Del. Rev. 2007, 59: 546–567
133 Avdeef A., Pharm. Pharmacol. Commun. 1998, 4(3): 165–178
134 Stuart M. and Box K., Anal. Chem. 2005, 77(4): 983-990
135 Box K. et al., Chem. and Biod. 2009, 6: 1767-1788
136 McNamara D. P. et al., Pharm. Res. 2006, 23(8): 1888-97
137 Edueng K. et al., Pharm. Res. 2017, 34: 1754–1772
138 Ozaki S. et al., J. Pharm. Sci. 2012, 101(1): 214-222
139 Palmelund H. et al., J. Pharm. Sci. 2016, 105: 3021-3029
140 Erdemir D. et al., Acc. Chem. Res. 2009, 42(5): 621-662
141 Gebauer D. et al., Chem. Soc. Rev. 2014, 43: 2348-2371
142 Tarn K.Y. and Takács-Novák K., Pharm. Res. 1999, 16: 374-381
143 Mitchell R. C. et al., J. Pharm. Biomed. Anal. 1999, 20(1-2): 289-295
144 Tam K. I. et al., Talanta 1999, 49(3): 539-554
145 Hanocq M. et al., Anal. Lett. 1989, 22: 117-140
146 Fisicaro E. and Braibanti A., Talanta 1988, 35(10): 769-774
147 Kaljurand I. et al., J. Org. Chem. 2000, 65: 6202-6208

148 Kaljurand I. et al., J. Org. Chem. 2002, 67: 1873-1881
149 Kaljurand I. et al., J. Org. Chem. 2005, 70: 1019-1028
150 Kutt A. et al., J. Org. Chem. 2006, 71: 2829-2838
151 Kutt A. et al., J. Org. Chem. 2011, 76: 391–395
152 Kutt A., Ph.D. thesis: Studies of acid-base equilibria in non-aqueous media, 2008
153 Klein S., AAPS J. 2010, 12(3): 397-406
154 Frank K. J. et al., Eur. J. Pharm. Sci. 2012, 47: 16–20
155 Reppas C. and Vertzoni M., J. Pharm. Pharmacol. 2012, 64: 919–930
156 Dressman J. B. et al., Pharm Res. 1990, 7(7): 756-61
157 Brouwers J. et al., J. Pharm. Sci 2006, 95(2): 372–383
158 Mariangeles Perez de la Cruz M., J. Pharm. Pharmacol. 2006, 58(8): 1079-1089
159 Kalantzi L. et al., Pharm. Res. 2006, 23(1): 165-176
160 Clarysse S. et al., J. Pharm. Sci. 2009, 98(3): 1177-1192
161 Psachoulias D. et al., Pharm. Res. 2011, 28(12): 3145-58
162 Riethorst D. et al., J Pharm. Sci. 2016, 105(2): 673-681
163 Lindahl A. et al., Pharm. Res. 1997, 14(4): 497-502
164 Pedersen B. L. et al., Pharm. Res. 2000a, 17(2): 183-189
165 Pedersen B. L., Pharm. Res. 2000b, 17: 891-894
166 Persson E. M. et al., Pharm. Res. 2005, 22: 2141–2151
167 Kossena G. A. et al., Pharm. Res. 2007, 24(11): 2084-2096
168 Rollinger J. M. et al., J. Pharm. Sci. 2001, 90(7): 951-959
169 Willart J. F. et al., Mol. Pharm. 2008, 5(6): 905-920
170 Charman W. N. et al., J. Pharm. Sci. 1997, 86(3): 269-282
171 Elvang P. A. et al., J. Pharm. Sci. 2016, 105: 2832-2839
172 Hofmann A. F. and Roda A., J. Lipid Res. 1984, 25: 1477-1499
173 Hofmann A. F. and Mysels K. J., J. Lipid Res. 1992, 33: 617-626
174 Taylor L. S. and Zhang G. G. Z., Adv. Drug Del. Rev. 2016, 101: 122–142
175 Ilevbare G. A. and Taylor L. S., Cryst. Growth Des. 2013, 13: 1497−1509
176 Lehto P. et al., J. Pharm. Sci. 2009, 98 (3): 985-996

177 Mevik B. H., Wehrens R. and Liland K. H. 2015, pls: Partial Least Squares and Principal Component Regression. R package, version 2.5-0. http://CRAN.R-project.org/package=pls

178 Huang J. et al., Am. Pharm. Rev. 2010, 13(6): 116-125

179 Rodriguez-Hornedo N. and Murphy D., J. Pharm. Sci. 2004, 93(2): 449-460

180 Effect of composition of simulated intestinal fluid on the solubility and supersaturation of felodipine investigated by design of experiment, Poster session of 11th World Meeting on Pharmaceutics, Biopharmaeutics and Pharamaceutical Technologies, Granada, 21/04/2018

Aknowledgements

At the end of this Ph.D. I would like to aknowledge all the people that allowed this thesis to be accomplished.
Firstly, Prof.ssa Federica Vacondio deeply deserves my gratitude for being my Supervisor. I am indebted to her for introducing me to the characterization of physico-chemical properties and for her scientific guidance throughout these three years. I feel honored for the kindness, patience and confidence that you have constantly showed me.
I would aknowledge my gratitude to the Analytics & Preformulation Department of Chiesi Farmaceutici and especially to Dott.ssa Silvia Catinella for giving me the chance to develop this work. I am deeply grateful to my Advisor Dott.ssa Valentina Mileo for believing in this project and supporting its development. A special thanks goes to Dott.ssa Elisa Moretti for accepting the role of external referee of my thesis.
I would thank Prof. Anette Müllertz who welcomed me in the Physiological Pharmaceutical (former RODD) Group at KU in Copenhagen and made my stay so fruitful. I would express my thanks to all the P. P. people for the special time we had together and I would especially name some of them. I am grateful to Matthias Manne Knopp for teaching me the basis of solid state studies and for kindly judging me suitable for the Doctor Europaeus title. Among my lab mates, I would especially thank Jakob Plum for sharing his project and for letting me discover the world of supersaturation and biorelevant fluids and Juliane for her helpfulness and funny company during my stay. I will always recall my exchange period as one of the best time I have ever experienced.
My gratitude goes to Prof. Stefanie-Dorothea Krämer who expressed a positive judgement for Doctor Europaeus title and to Prof. Caterina Temporini as my external referee.
I would aknowledge my appreciation to all my Ph.D. fellows and the Master's students that developed their thesis into lab 084 making the

atmosphere positive and lively. I would particularly thank Lara, Roberta, Katia and Giorgia that I am now lucky enough to call friends.
Thanks to my lifetime friends Valeria, Anna, Mattia and Ottavia for making me feel their affection despite the distance between us. A heartfelt thanks goes to Davide for being a kind and patient mate and a true model of life and work ethics.
Above all the special people that I mentioned, my parents and my brother deserve the greatest appreciation: thank you for the unconditional support and love you make me feel day by day. Without your positive impulse I would have never started this Ph.D. neither written this thesis.