Predictions of biorelevant solubility change during dispersion and digestion of lipid-based formulations

Lotte Ejskjær , René Holm , Martin Kuentz , Karl J. Box , Brendan T. Griffin , Patrick J. O'Dwyer

PII: S0928-0987(24)00145-3
DOI: https://doi.org/10.1016/j.ejps.2024.106833
Reference: PHASCI 106833

To appear in: *European Journal of Pharmaceutical Sciences*

Received date: 5 April 2024
Revised date: 23 May 2024
Accepted date: 11 June 2024

Please cite this article as: Lotte Ejskjær , René Holm , Martin Kuentz , Karl J. Box , Brendan T. Griffin , Patrick J. O'Dwyer , Predictions of biorelevant solubility change during dispersion and digestion of lipid-based formulations, *European Journal of Pharmaceutical Sciences* (2024), doi: https://doi.org/10.1016/j.ejps.2024.106833

Predictions of biorelevant solubility change during dispersion and digestion of lipid-based formulations

Lotte Ejskjær[1], René Holm[2], Martin Kuentz[3], Karl J. Box[4], Brendan T. Griffin[1], Patrick J. O'Dwyer[1]

University College Cork, College Road, Cork, Ireland[1], University of Southern Denmark, Campusvej 55, Odense, Denmark[2], University of Applied Sciences and Arts Northwestern Switzerland, Hofackerstr. 30, Muttenz 4132, Switzerland[3], Pion Inc (UK), Forest Row, East Sussex, UK[4]

Corresponding author:

Patrick O'Dwyer

Email: Patrick.odwyer@ucc.ie

Address: School of Pharmacy, University College Cork, Cork, Ireland

Abstract

Computational approaches are increasingly explored in development of drug products, including the development of lipid-based formulations (LBFs), to assess their feasibility for achieving adequate oral absorption at an early stage. This study investigated the use of computational pharmaceutics approaches to predict solubility changes of poorly soluble drugs during dispersion and digestion in biorelevant media. Concentrations of 30 poorly water-soluble drugs were determined pre- and post-digestion with in-line UV probes using the MicroDISS Profiler™. Generally, cationic drugs displayed higher drug concentrations post-digestion, whereas for non-ionized drugs there was no discernible trend between drug concentration in dispersed and digested phase. In the case of anionic drugs there tended to be a decrease or no change in the drug concentration post-digestion. Partial least squares modelling was used to identify the molecular descriptors and drug properties which predict changes in solubility ratio in long-chain LBF pre-digestion (R^2 of calibration = 0.80, Q^2 of validation = 0.64) and post-digestion (R^2 of calibration = 0.76, Q^2 of validation = 0.72). Furthermore, multiple linear regression equations were developed to facilitate prediction of the solubility ratio pre- and post-digestion. Applying three molecular descriptors (melting point, LogD, and number of aromatic rings) these equations showed good predictivity (pre-digestion R^2 = 0.70, and post-digestion R^2 = 0.68). The model developed will support a computationally guided lipid-based formulation strategy for emerging poorly water-soluble drugs by predicting biorelevant solubility changes during dispersion and digestion. This facilitates a more data-informed developability decision making and subsequently facilitates a more efficient use of formulation screening resources.

Key words

Computational pharmaceutics, developability screening, in silico modelling, multivariate analysis, partial least squares, physicochemical properties, lipid-based-formulation

List of abbreviations

BCS	Biopharmaceutical classification system
rDCS	Refined developability classification system
LBF	Lipid-based formulation
LogD	Partition coefficient
MLR	Multiple linear regression
NAP_MinQ	Minimum NPA partial atomic charge
N_AromR	Number of aromatic rings
PCA	Partial component analysis
Pi_FMi2	Second component of the autocorrelation vector scaled pi Fukui-indices
Pi_AQc	Sum of absolute values of pi partial atomic charge on carbon
Pi_FMi4	Fourth component of the autocorrelation vector scaled pi Fukui-indices
PLS	Partial least square
$RMSE_{CV}$	Root-mean-square error of the cross-validation
RMSEE	Root-mean-square error of the estimate
RMSEP	Root-mean-square error of prediction
SaaCH	Hydrogen E-state index for aromatic carbons
SMILES	Simplified molecular-input line-entry system
SR	Solubility ratio
T_m	Melting point
UV	Ultra-violet

1 Introduction

In the discovery phase, computational tools are well established to identify new lead candidates with optimal receptor binding affinity and physicochemical profiles with mechanistic models or data-driven approaches, like quantitative structure activity relationship and quantitative structure property relationship (C. Bergström et al., 2016). The drug candidates are often evaluated based on their "drug-ability", which is generally defined as the likelihood of the drug being able to functionally interact with the biological target. However, a consequence of discovering drugs with more potent binding often includes the selection of more lipophilic drug candidates, which are often not readily amenable for oral delivery (C. Bergström et al., 2016). Bio-enabling formulation strategies have been developed to increase the bioavailability of these drug compounds. Examples of these bio-enabling formulation strategies are solid dispersions (Singh & Van den Mooter, 2016; Van Den Mooter, 2012), nano- and micro-sizing (Merisko-Liversidge et al., 2003), and lipid-based formulation (LBF) (Feeney et al., 2016). There is no formulation that works for all candidate drugs, therefore, the development of a lead formulation for pivotal clinical trials often includes screening of a range of excipients and formulations until a formulation with acceptable bioavailability is identified. Alternatively, multiple formulations with the drug compound are developed in parallel to increase the chance of success. Each of these approaches has high development costs, long development timelines, and includes a risk of failure to develop the new medicine (Kuentz et al., 2021; Lennernäs et al., 2014). This is particularly concerning because contemporary drug product development continues to depend on trial-and-error and/or in-house knowledge within individual pharmaceutical companies (Ditzinger et al., 2019). However, computational approaches to guide drug product development have substantial potential to make drug product development faster, better, and cheaper (C. Bergström & Larsson, 2018).

LBF is a bio-enabling formulation strategy where the drug compound is presented pre-dissolved in a lipid matrix before administration. LBFs hold for a very promising technology for solubility-rate-limited drugs and for drugs that exhibit significant positive food effects (Feeney et al., 2016). However, the development of LBFs has declined over the last decade (Bennett-Lenane et al., 2020). This can be an indication of challenges among pharmaceutical companies to adopt LBF strategies and a need for guidance on development (Holm, 2019). Drug absorption from LBFs is a dynamic process where the solubility of the drug compound dramatically changes upon both dispersion and digestion also depending on the lipid excipients in the formulation (Boyd & Clulow, 2021; Gautschi et al., 2016; Khan et al., 2016). The dispersion of LBFs in biorelevant media can lead to increased drug solubilisation, transient

supersaturation, and delayed precipitation, meaning that the drug can be present in the biorelevant media in a concentration higher than the equilibrium concentration. The pH-stat method is the most widely used *in vitro* lipolysis method for testing LBFs. With this approach, the pH of the digestion media is maintained throughout the experiment by adding sodium hydroxide to counteract the formation of free fatty acids because of the digestion of the LBFs. This method requires relatively high quantities of media, formulation, and drug compound. Moreover, it is relatively time consuming as it is a low throughput method that requires withdrawing of samples for off-line analysis (Williams et al., 2012). Optimized *in vitro* lipolysis methods have been suggested to comply with the need for higher throughput, smaller quantities, and real-time analytics (Devraj et al., 2014; Khan et al., 2022; Mosgaard et al., 2015, 2017; Tanaka et al., 2022). Recently, our research group has shown proof of concept of a small-scale lipolysis method using the MicroDISS Profiler™ which facilitates in-line higher throughput data generation of drug concentrations during dispersion and digestion (Ejskjær et al., 2023).

A number of studies have demonstrated the use of data-driven prediction models in the development of LBFs to estimate lipid solubility to act as a guide for maximal dose loading (L. Alskär et al., 2016; Persson et al., 2013; Sacchetti & Nejati, 2012). Even though it is useful to guide the initial understanding of the maximal dose loading in the LBFs this approach does not represent the sole criterion for LBF suitability. As previously mentioned the drug absorption from LBF is a dynamic process where the solubility can change dramatically upon both dispersion and digestion. Therefore, these aspects are also important to consider when developing a LBF. Furthermore, the solution stability, content uniformity, capsule filling and precipitation are also important aspect to consider later in the development phase (Feeney et al., 2016). A study by Bennett-Lenane et al., (2021) showed the utility of using statistical modelling to predict the biorelevant solubility change of drugs in two types of LBFs. This study provided important and valuable information about solubility upon dispersion in biorelevant media and the use of predictive tools in drug product development. However, a key limitation of the study was that the effect of digestion of the LBFs and the effect on solubility thereof was not considered. The digestion is essential to assess as most excipients used in LBFs are naturally digestible in the intestine which can significantly affect solubility (Koehl et al., 2020; Zupančič et al., 2023).

Accordingly, the objective of this study was two-fold; firstly, to demonstrate a broader utilization of the higher throughput *in vitro* lipolysis setup by assessing the influence of both dispersion and digestion for a range of drugs within an LBF. Secondly, to enhance the use of computational approaches in

LBF strategies by determining whether the solubilised drug concentrations change pre- and post-digestion can be accurately captured in a statistical prediction model, and development of two new interpretable equations which will allow rapid assessment of the viability of the LBF approach.

2 Method

2.1 Dataset selection

The dataset chosen for this study was similar to the data used to investigate the solubility gain upon dispersion in LBF using the MicroDISS Profiler™ (Bennett-Lenane et al., 2021). The dataset is composed of 30 poorly water-soluble drugs that cover a broad physicochemical spectrum (see Table 1). It consists of anionic drugs (8), cationic drugs (9), and non-ionized drugs (13) at pH 6.5. All drugs were purchased from Kemprotec Ltd (Cumbria, UK). The drug compounds have ultra-violet (UV) chromophores to be able to detect the concentration with the in-line UV probes in the MicroDISS Profiler™.

2.2 Formulations

Two formulations that previously have been investigated and characterized as a type III LBF were chosen for this study (Bennett-Lenane et al., 2021, Griffin et al., 2014). A medium-chain LBF, and a long-chain LBF. The formulations were prepared by weighing in exact amounts of excipients (Table 2) into a glass vial. The formulations were stirred at 37 °C at 300 rpm overnight.

2.3 Media preparation

Phosphate buffer was prepared according to the protocol from biorelevant.com (London, UK) and adjusted with sodium hydroxide to either pH 6.5 or pH 7.5. A pH of 6.5 was used for all drugs except for compounds with pKa values within the pH shift observed in a previous study (Ejskjær et al., 2023). These drugs (dipyridamole and ketoconazole) were run at a pH of 7.5 to avoid artificial solubilization change caused by the ionization of the drug compound. The buffer was supplemented with FaSSIF powder a day prior to the experiments.

2.4 MicroDISS Profiler™

The experimental protocol was adopted from (Ejskjær et al., 2023). In brief, the MicroDISS Profiler™ (Pion Inc., USA) was used to determine the drug concentrations (n = 3) during one hour of dispersion and three hours of digestion. The instrument settings were a temperature of 37 °C and a stirring rate of 250 rpm. In the 24-hour equilibrium solubilities upon dispersion acquired from Bennett-Lenane et al., (2021), a stirring rate of 300 rpm was used. The difference in stirring rates reflected limits in instrument settings, however, the impact of change in stirring rates was assessed to be minor, as all experiments reached a plateau in drug concentrations. The path lengths used were either 1 mm, 2 mm, or 5 mm depending on the expected concentrations and the drug's chromophore quality. Standard spectra were collected for each UV probe and each drug at the pH of the experimental run, and a linear relationship was established between concentration and UV absorbance. The experimental run was conducted in four vials. Three vials were run with drug and the fourth vial was run as a blank to consolidate for potential UV changes in the background media during time and digestion. Each vial contained 15 mL FaSSIF buffer, blank LBF in the ratio of 1:200, and a cross-stirring bar. The ratio of 1:200 between LBF and FaSSIF was used to correspond to a more biorelevant ratio compared to the standard 1:40 which is considered to have an excess of LBF compared to real-life dosing conditions. The blank LBF was added to the vessel, and after 15-20 minutes of dispersion, the UV-detection was started, and excess drug was added to three of the vials. At least double the amount of drug as previously found soluble in the LBF in FaSSIF was added to ensure an excess of drug was present (Bennett-Lenane et al., 2021). The drug dispersion ran for one hour. Following that, Palatase® 20000L was added to all vials at a concentration of 125 PLU/mL (PLU/mL is the propyl laurate unit per gram enzyme, which reflects the amount of enzyme that generates 1 µmol of propyl laurate per minute), and the digestion process continued for three hours. The digestion with Palatase® 20000L has previously been shown to correspond to digestion with the porcine pancreatin which is the standard lipase to use (Ejskjær et al., 2023). In situ scans were collected every 30 seconds. The concentrations were determined by considering area-under-the-curve in second derivative spectra. The AuPRO software (version 7, Pion Inc, MA, USA) was used to interpret the data. The pH of the media was measured at the end of the experiment.

2.5 Drug physicochemical and molecular properties

Isomeric simplified molecular-input line-entry system (SMILES) was acquired from PubChem for each drug compound. The SMILES were used as input for calculating 387 descriptors from ADMET Predictor 9.5 (Simulation Plus, USA). Physiochemical and molecular descriptors were calculated both at the pre-digestion and post-digestion pH. Only the LogD value differed at the two pH values and were both included in the dataset. Additionally, the melting point (T_m) derived from Bennett-Lenane et al., (2021) was added to the descriptor data set.

2.6 Biopharmaceutical data analysis

The solubility change between the pre-digested and post-digested solubility of the drug were calculated via Equation 1.

$$\text{Solubility change}_{(\text{post-digestion/pre-digestion})} = \frac{\text{Solubility}_{\text{post-digestion}}}{\text{Solubility}_{\text{pre-digestion}}} \quad \text{(Equation 1)}$$

Where the $\text{solubility}_{\text{post-digestion}}$ refers to the drug concentration measured after three hours of digestion, and the $\text{solubility}_{\text{pre-digestion}}$ refers to the drug concentration determined after one hour of dispersion.

The standard error (SE) was calculated from Equation 2 as previously reported (Bennett-Lenane et al., 2021).

$$\text{SE} = \sqrt{\frac{\text{SA}^2}{\text{A}^2} + \frac{\text{SB}^2}{\text{B}^2}} \quad \text{(Equation 2)}$$

Where A and B are the mean measured solubility values pre- and post-digestion, and SA and SB are the standard error for A and B.

A two-sided t-test was used to investigate if a significant ($p > 0.05$) solubility change, or loss was found post-digestion.

The solubility ratio (SR) was determined as the ratio between the pre- or post-digestion solubility versus FaSSIF solubility of the drug were calculated via Equation 3, as previously described (Bennett-Lenane et al., 2021). The pre-digestion solubility used for these calculations were the 24 h equilibrium solubility upon dispersion acquired from (Bennett-Lenane et al., 2021).

$$\mathrm{SR} = \frac{\mathrm{Solubility}_{\mathrm{pre-\ or\ post-digestion}}}{\mathrm{Solubility}_{\mathrm{FaSSIF}}} \quad \text{(Equation 3)}$$

2.7 Multivariate data analysis and modelling parameters

Multivariate data analysis was conducted in SIMCA 17 (Umetrics, Sweden). The raw data set used for the analysis together with the model info sheet is published with the paper in supplementary materials S1 and S2, respectively. The analysis was guided by the approaches used previously (Bennett-Lenane et al., 2021; Persson et al., 2013) and aimed to comply, where feasible, with recommendations for machine learning approaches in drug formulation published by Murray et al., (2023). A schematic of the modelling approach is shown in Figure 1. The calculated physicochemical and molecular descriptors and T_m were used as variable descriptors in the model, and the logarithm of SR (equation 3) was used as the dependent variable. Skewed descriptors were manually excluded from the data by assessment of the individual histograms. Afterward, the descriptors were de-identified, centred, and scaled to unity variance.

Principal component analysis (PCA) was used to divide the dataset into a training set and a test set. A split of 20 drugs in the training set and 8 drugs in the test set was used. Outliers were identified and placed in the test set to avoid that these distort the model development. First, the DModX plot was used to identify outliers (dipyridamole and isotretinoin), afterward the score plot was used to identify drugs that lie outside the Hotelling's T2 (95% confidence interval) ellipse (venetoclax) and the remaining drugs were randomly identified to best cover the chemical space, ionisation of the compounds, and SR range (see supplementary materials Figure S3).

Partial least squares (PLS) was used to establish important descriptors for predicting the SR. The variable reduction was performed on the training set to decrease complexity, increase interpretability, and decrease noise. The variable reduction was performed by identifying and excluding descriptors with low importance for the response and/or having information duplicated by other descriptors. This was done by consulting the coefficients plot, loading plot, and variable importance for the projection plot while monitoring the R^2 for calibration, and Q^2 for the leave-one-out cross-validation using 7 cross-validation groups. If the exclusion of a descriptor did not affect or increased the Q^2 for validation the descriptor was permanently removed from the model. The number of principal components was chosen according to the R^2 for calibration, and Q^2 for the leave-one-out cross-validation using 7 cross-validation

groups, however a limit of two principal components was chosen to limit overfitting potential. The accuracy of the model was validated by the root-mean-square error of the estimate (RMSEE) and root-mean-square error of the cross-validation ($RMSE_{CV}$). Finally, the test set was used for validation by the root-mean-square error of prediction (RMSEP).

A pre-digestion PLS model was produced with the same approach as described above. For this pre-digestion model the 24 h dispersion data acquired from Bennett-Lenane et al., (2021) were used.

2.8 Solubility equation for predicting SR

Multiple linear regression (MLR) was applied to make an easily interpretable equation to predict the SR. The MLR was performed in Excel (Microsoft Office, version 2302). The same test/training split as used in the PLS was used, and the descriptors identified in the PLS model. Equation development was performed on the training set monitored by the p-value, the f-value, and R^2-value. The test set was used to validate the equation by the root-mean squared error of the test set ($RMSE_{Test}$). The collinearity was assessed by the variance inflation factor and tolerance by IBM SPSS (version 28.0, Armonk, NY, USA).

2.9 Calculating dose number and rDCS classification

The rDCS classification of each drug was obtained using solubility and permeability as outlined previously (Butler & Dressman, 2010; Rosenberger et al., 2018). The permeability of the drugs was predicted from the ADMET Predictor 9.5 (Simulation Plus, USA), whereas the solubility criteria was obtained using a dose/solubility ratio (see Equation 4).

$$Do = \frac{dose}{(S_{si} \cdot V_{si})} \quad \text{(Equation 4)}$$

Where, the dose is the highest dose, S_{si} is the apparent solubility in biorelevant media; the determined concentrations of the drugs were utilized in this study, and V_{si} is the volume in the small intestine available for dissolution (500 mL) (Butler & Dressman, 2010; Rosenberger et al., 2018).

3 Results

3.1 Impact of lipolysis on drug solubility

The lipolysis model was tested with a diverse dataset of 30 poorly water-soluble drugs in a medium-chain and long-chain LBF. The lipolysis model was run for one hour of LBF dispersion followed by three hours of digestion. The results showed that the method was compatible with 28 out of the 30 drugs in the long-chain LBF and with 15 out of the 30 drugs in the medium-chain LBF. However, during experiments with the remaining drugs, the digestion phase resulted in a cloudy and turbid medium, where the UV-probes were unable to detect drug concentrations. As a result, no data are available for these drugs. The trend in solubility change was generally that the cationic drugs displayed higher solubilised drug concentrations post-digestion compared to pre-digestion with a median solubility ratio increase of 2.80 for the long-chain LBF and 3.36 for the medium-chain LBF. The trend for the anionic drug compounds was generally that they decreased or had no change in drug concentrations post-digestion with median solubility ratio loss of 0.91 for the long-chain LBF and 0.51 for the medium-chain LBF. There was no discernible trend between pre- and post-digestion concentrations among non-ionized drug compounds. The solubility change ranged from 0.7-5.2-fold for the long-chain LBF and 0.2-1.3-fold for the medium-chain LBF (see Figure 2).

3.2 rDCS class transition pre-digestion versus post-digestion

The concentrations of the drugs post-digestion and the solubilities of the drug compounds in the same formulation upon 24 h of dispersion (referred to as pre-digestion) acquired from Bennett-Lenane et al., (2021) were used to classify the drugs according to the rDCS. With the intention of linking the dose number to a developability framework this demonstrates greater application in a development setting. The rDCS is a classification system to assess drug candidates with respect to their developability for oral delivery based on solubility and permeability, whereas the biopharmaceutical classification system (BCS) was developed with a regulatory focus, and therefore has more conservative cut-off values (Butler & Dressman, 2010; Rosenberger et al., 2018). Overall, the majority of the drugs in the dataset maintained the same rDCS classification under post-digestive conditions as pre-digestion conditions, in 23 out of 28 drugs in the long chain LBF the digestion did not impact the rDCS classification. While the dataset for drugs was smaller for medium-chain LBFs, overall, 9 out of 15 drugs maintained the same classification between pre- and post-digestion (see Table 3). Four of the drugs (isotretinoin, mefenamic acid, nifedipine, and progesterone) made a rDCS transition in the long-chain LBF to a lower solubility classification, of which three were anionic compounds and

one was a non-ionized compound, whereas only one drug (ketoconazole) moved from rDCS Class IIb → IIa. For the medium-chain LBF, five of the drugs (celecoxib, danazol, fenofibrate, naproxen, and venetoclax) transitioned to a lower solubility classification of which three were non-ionized compounds and one was an anionic compound, and one of the drugs (ketoconazole) moved to a higher solubility classification. Ketoconazole, a cationic drug at pH 6.5, was the only compound that made a transition in classification for both long-chain LBF and medium-chain LBF and in both cases (see Table 3).

3.3 Computational prediction

Lipophilicity and T_m are two of the common parameters used to guide LBF design, however, a poor linear relationship was found between the post-digestion solubility ratio of the long-chain LBF and LogP (R^2 = 0.09), LogD (R^2 = 0.27), and T_m (R^2 = 0.33), and similar poor relationships were found for the medium-chain LBF (R^2 = 0.34, 0.48, and 0.35, respectively) (see supplementary materials S4). A machine-learning approach looking at a combination of descriptors was utilized for the long-chain LBF data set.

PCA was used to assess the data set while splitting the dataset into a test and training set. Afterward, a PLS analysis was used to identify the important predictors. The PLS model used two principal components and five descriptors. The model had a R^2 of calibration at 0.76 and a Q^2 of validation of 0.64. The five descriptors identified were the T_m, the partition coefficient at the initial pH ($LogD_{pre\text{-}pH}$), the number of aromatic rings (N_AromR), the hydrogen E-state index for aromatic carbons (SaaCH), and the second component of the autocorrelation vector scaled pi Fukui-indices (Pi_FMi2). The model is summarised in Table 4.

The five descriptors identified in the PLS model were retained to formulate a MLR equation for ease of interpretability and usability. Initially, the collinearity was assessed by the tolerance and variance inflation factor for the five descriptors identified, where no collinearity was found (see supplementary materials S5). All descriptors from the PLS model were initially included in the MLR, insignificant descriptors were removed from the equation to produce a final equation with a higher F-value. The three identified descriptors used for the MLR equation were $LogD_{pre\text{-}pH}$, T_m, and N_AromR (Table 4). The collinearity of the descriptors included in the MLR equation was assessed and no collinearity was found (see supplementary materials S5).

In the pre-digestion model the data acquired from Bennett-Lenane et al., (2021), which monitored drug concentration over a 24 hour dispersion period, was used because it was discovered that one hour of dispersion in this study was not sufficient to reach an apparent equilibrium solubility. The PLS model used one principal component and six descriptors. The R^2 of calibration was 0.80, and the Q^2 of validation was 0.72. The six descriptors identified were the T_m, LogD, N_AromR, the minimum NPA partial atomic charge (NAP_MinQ), the sum of absolute values of pi partial atomic charge on carbon (Pi_AQc), and the fourth component of the autocorrelation vector scaled pi Fukui-indices (Pi_FMi4). The model is summarised in Table 5.
Three descriptors were also identified for the MLR equation, the LogD, T_m, and N_AromR (see Table 5). No collinearity was found for the descriptors identified in the PLS model or in the MLR (see supplementary materials S6).

4 Discussion

Recently more focus has been on optimizing the identification of the most appropriate formulation strategy for a new drug compound that displays poor solubility and/or permeability (C. Bergström et al., 2016; Hossain et al., 2019; Mehta et al., 2019; Reppas et al., 2023). The formulation challenge includes both development of small-scale *in vitro* methods with higher throughput and the use of computational tools to guide the selection of the most appropriate formulation strategy. The application of predictive computational models in LBF development is increasing with more mechanistic methods and simulations to data-driven modelling that has particular merits with complex systems such as LBF that exhibit much change on dispersion and digestion. To the best of our knowledge, this study is the first attempt to make a validated predictive model for solubility ratio post-digestion of LBFs.

In general, it was found that the cationic drugs (positively charged at pH 6.5) increased drug concentration post-digestion shown by solubility change (pre-digestion/post-digestion) > 1 in (see Figure 2). This makes LBF an interesting formulation approach for these types of drugs as the digestion of the LBF seems to boost drug solubility which is also been shown in a previous study (L. Alskär et al., 2018). This is likely due to the electrostatic interactions between the positively charged drug and the negatively charged free fatty acids generated post-digestion. This was also found for the solubility of the cationic drugs upon dispersion in FaSSIF where it was postulated that the negatively charged head groups of the taurocholate bile salts improved drug concentration (Bennett-Lenane et al., 2021). Furthermore, a previous study found that not only the amount of free fatty acids released but also the type of free fatty acid was shown to be important for the solubilisation capacity of the digestion product, where the cationic drugs favoured the digestion products of the long-

chain LBF (L. Alskär et al., 2018). The anionic drugs (negatively charged at pH 6.5), generally, showed a decrease in drug concentrations or no influence in drug concentrations between the pre-digestion and post-digestion concentrations shown by solubility change (pre-digestion/post-digestion) </≈ 1. This was also in line with a previous study that found a decrease in drug concentrations post-digestion for anionic drugs (L. C. Alskär et al., 2018), and a study that showed that oleic acid was not as important for the solubility of anionic and non-ionized drugs compared to cationic drugs (Yeap et al., 2013). However, the factors influencing drug solubilisation in post-digestive conditions extend beyond electrostatic interactions. For example, some non-ionized drugs displayed higher drug concentrations post-digestion (see Figure 1). A previous study suggest that non-ionized drugs had a decrease in drug concentrations post-digestion (L. C. Alskär et al., 2018), however, the broader data set tested here shows no general trend for the solubility change of the non-ionized drugs post-digestion. Where some showed an increase in solubility upon digestion others showed a decrease in solubility which may suggest a more complex relationship and that colloidal molecular environment can have both positive and negative impact.

Using the drug concentrations to predict the rDCS classification transition upon formulating the drugs in LBFs is helpful for formulation scientists to explore the suitability of drugs in LBF and the potential solubility limited absorption for a given dose. Bennett-Lenane et al., (2021) showed that in a dispersed phase, all drugs displayed higher concentrations in a LBF dispersion relative to the solubility in biorelevant media demonstrating the advantage of solubilising lipids to improve solubility of the range of poorly soluble drugs. While that study used the 24-hour timepoint as equilibrium solubility to understand the maximum solubility, albeit not being biorelevant in terms of absorption timeframes. In this study, a dispersion phase of 1-hour was used to be more biorelevant and allow the dispersion-digestion transition. In the long-chain LBF 13 out of the 23 drugs with “low” solubility in FaSSIF (rDCS class II/IV) made a transit to “high” solubility (rDCS class I/III) or a class transit from rDCS IIb to IIa upon dispersion of which four were cationic, four were non-ionized and five were anionic. Our study advances on those observations to explore whether that these improvement drug concentrations are maintained in a post-digestive environment and showing that 9 of these drugs maintained their improved solubility classification post-digestion, whereas four reverted back to the lower solubility classification, three anionic drugs and one non-ionized drug. Furthermore, one drug (ketoconazole) shifted to a higher solubility classification upon digestion even though a shift was not seen upon dispersion (see Table 2). Similarly for the medium-chain LBF, 8 out of the 15 drugs explored moved to a higher solubility classification upon dispersion (Bennett-Lenane et al., 2021). Our study showed that three of these maintained the solubility classification (cinnarizine, clofazimine, and spironolactone) and five of these drugs (celecoxib, danazol, fenofibrate, naproxen, and venetoclax)

reverted to a lower solubility classification. Here, ketoconazole also transferred to a higher solubility classification post-digestion even though a transition was not seen pre-digestion.

Limitations of the experimental setup was the buffer capacity which was not high enough to maintain a stable pH throughout the experiment resulting in a pH drop. The pH drop was relatively minor for the long-chain LBF (from pH 6.5 to pH 6.35), however, two drugs (dipyridamole and ketoconazole) with pKa values in this region were run at pH 7.5 to reduce the ionisation effect on solubilisation capacity. Additionally, to compensate for the ionisation effect in the modelling, dipyridamole was placed in the test set and ketoconazole in the training set to minimise any biases introduced into the model. Furthermore, the post-digestion model was performed where dipyridamole and ketoconazole were excluded which only showed minor changes of the statistical results, and no differences in the descriptors identified (see supplementary materials S7).

The results showed that three descriptors could be successfully employed in the MLR equations to predict the SR both pre- and post-digestion, T_m, Log D, and N_AromR (Table 3 and 4). The two models showed equal predictability in the PLS model, where the R^2 of calibration were 0.80 and 0.76, and the Q^2 of validation were 0.64 and 0.72, for the pre- and post-digestion model respectively, and in the MLR equation with R^2 of 0.70 and 0.68, respectively (see Table 4 and 5). The T_m was negatively correlated with the SR in both models. This was most likely because molecules with high T_m exhibit solid-state limited solubility ("brick dust" molecules) which will result in poor solubility in the lipid excipients and result in a more modest SR upon dispersion and digestion (L. Alskär et al., 2016). The LogD was positively correlated to the SR pre- and post-digestion. This was not unexpected as adding the LBF to the media will increase the lipophilicity of the media, and previous studies have shown that lipophilicity of drugs have an influence on the drug concentration post-digestion (Kaukonen, Boyd, Charman, et al., 2004; Kaukonen, Boyd, Porter, et al., 2004). The LogD has also in a previous study shown a strong correlation with the solubility of poorly water-soluble drugs in biorelevant media (Fagerberg et al., 2010), and in the modelling of FaSSIF/phosphate buffer (pH 6.5) ratio (Fagerberg et al., 2012). The N_AromR exhibited a positive correlation with SR in both models, contrary to the findings of a previous study on solubility modelling in human intestinal fluid (Fagerberg et al., 2015). Here aromatic compounds were found to be less hydrated compared to the drugs with more aliphatic structure, however, it might indicate that drugs with aromatic structure will have a higher affinity for lipid excipients (Ritchie et al., 2011). A number of more complex descriptors were also identified in the PLS models but were found non-significant in the MLR equations. In the pre-digestion PLS

model, NPA_MinQ, Pi_AQc, and Pi_FMi4 were identified (see Table 4), and in the post-digestion model, SaaCH and Pi_FMi2 were identified (see Table 3). These descriptors are not as straightforward to interpret as the T_m, LogD, and N_AromR. However, they describe the atomic charge within the molecule and electrophilicity; similar descriptors for topological distance and electronegativity have previously been found significant in *in silico* predictions in FaSSIF buffer and human intestinal fluid (Fagerberg et al., 2010, 2015). As the data set is relatively small, this could cause over parameterisation in the models. Nevertheless, the models show promising results for producing reliable predictions. This study demonstrated the great potential of predicting a new drug candidate performance in a LBF both pre- and post-digestion prior to starting comprehensive experimental work, which include the classical characterisation of the formulation and *in vitro* studies. However, before industrial deployment bigger datasets are needed to provide better predictions with higher accuracy (Murray et al., 2023).

5 Conclusion

In conclusion, this study successfully demonstrated the application of the higher throughput in-line digestion method as a screening and characterization tool for dispersion and digestion of LBFs. Additionally, it successfully evolved the use of data-driven models in lipid-based drug product development by development of a pre- and post-digestion solubility model. The obtained equations to predict the SR upon dispersion and digestion of LBFs included three easily available molecular descriptors. This work highlights the significant potential of computational-informed drug development, offering the opportunity to predict the likelihood of a transition in rDCS classification when formulating a new drug as a LBF. This can enhance the efficiency of formulation development by providing early indications of formulation performance during the development phase.

Acknowledgments

All authors are part of the InPharma Industrial Training Network, which was funded by the European Union's Horizon 2020 Marie Sklodowska-Curie research and innovation program under grant agreement No. 955756.

The authors would like to thank Sam Lee for training and support with the MicroDISS experiments, as well as Dr. Harriet Bennett-Lenane for help and support with the PLS modelling.

Declarations of interest

KJB is an employee of Pion Inc. The other authors disclosed no conflicts of interest related to this article.

References

Alskär, L. C., Keemink, J., Johannesson, J., Porter, C. J. H., & Bergström, C. A. S. (2018). Impact of Drug Physicochemical Properties on Lipolysis-Triggered Drug Supersaturation and Precipitation from Lipid-Based Formulations. *Molecular Pharmaceutics*, *15*(10), 4733–4744. https://doi.org/10.1021/ACS.MOLPHARMACEUT.8B00699

Alskär, L., Porter, C., & Bergström, C. (2016). Tools for Early Prediction of Drug Loading in Lipid-Based Formulations. *Molecular Pharmaceutics*, *13*(1), 251–261. https://doi.org/10.1021/ACS.MOLPHARMACEUT.5B00704

Bennett-Lenane, H., Koehl, N. J., O'Dwyer, P. J., Box, K. J., O'Shea, J. P., & Griffin, B. T. (2021). Applying Computational Predictions of Biorelevant Solubility Ratio Upon Self-Emulsifying Lipid-Based Formulations Dispersion to Predict Dose Number. *Journal of Pharmaceutical Sciences*, *110*(1), 164–175. https://doi.org/10.1016/J.XPHS.2020.10.055

Bennett-Lenane, H., O'Shea, J. P., O'Driscoll, C. M., & Griffin, B. T. (2020). A Retrospective Biopharmaceutical Analysis of >800 Approved Oral Drug Products: Are Drug Properties of Solid Dispersions and Lipid-Based Formulations Distinctive? In *Journal of Pharmaceutical Sciences* (Vol. 109, Issue 11, pp. 3248–3261). Elsevier B.V. https://doi.org/10.1016/j.xphs.2020.08.008

Bergström, C. A. S., & Larsson, P. (2018). Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting. *International Journal of Pharmaceutics*, *540*(1–2), 185–193. https://doi.org/10.1016/J.IJPHARM.2018.01.044

Bergström, C., Charman, W., & Porter, C. (2016). Computational prediction of formulation strategies for beyond-rule-of-5 compounds. *Advanced Drug Delivery Reviews*, *101*, 6–21. https://doi.org/10.1016/J.ADDR.2016.02.005

Boyd, B. J., & Clulow, A. J. (2021). The influence of lipid digestion on the fate of orally administered drug delivery vehicles. *Biochemical Society Transactions*, *49*(4), 1749–1761. https://doi.org/10.1042/BST20210168

Butler, J. M., & Dressman, J. B. (2010). The developability classification system: Application of biopharmaceutics concepts to formulation development. *Journal of Pharmaceutical Sciences*, *99*(12), 4940–4954. https://doi.org/10.1002/JPS.22217

Ditzinger, F., Price, D. J., Ilie, A. R., Köhl, N. J., Jankovic, S., Tsakiridou, G., Aleandri, S., Kalantzi, L., Holm, R., Nair, A., Saal, C., Griffin, B., & Kuentz, M. (2019). Lipophilicity and hydrophobicity considerations in bio-enabling oral formulations approaches – a PEARRL review. *Journal of Pharmacy and Pharmacology*, *71*(4), 464–482. https://doi.org/10.1111/JPHP.12984

Ejskjær, L., O'Dwyer, P. J., Ryan, C. D., Holm, R., Kuentz, M., Box, K. J., & Griffin, B. T. (2023). Developing an in vitro lipolysis model for real-time analysis of drug concentrations during digestion of lipid-based formulations. *European Journal of Pharmaceutical Sciences*, 106681. https://doi.org/10.1016/J.EJPS.2023.106681

Fagerberg, J. H., Al-Tikriti, Y., Ragnarsson, G., & Bergströ, C. A. S. (2012). Ethanol Effects on Apparent Solubility of Poorly Soluble Drugs in Simulated Intestinal Fluid. *Molecular Pharmaceutics*, *9*(7). https://doi.org/10.1021/mp2006467

Fagerberg, J. H., Karlsson, E., Ulander, J., Hanisch, G., & Bergström, C. A. S. (2015). Computational Prediction of Drug Solubility in Fasted Simulated and Aspirated Human Intestinal Fluid. *Pharmaceutical Research* , *32*, 578–589. https://doi.org/10.1007/s11095-014-1487-z

Fagerberg, J. H., Tsinman, O., Sun, N., Tsinman, K., Avdeef, A., & Bergström, C. A. S. (2010). Dissolution Rate and Apparent Solubility of Poorly Soluble Drugs in Biorelevant Dissolution Media. *Molecular Pharmaceutics*, *7*(5), 1419–1430. https://doi.org/10.1021/mp100049m

Feeney, O. M., Crum, M. F., McEvoy, C. L., Trevaskis, N. L., Williams, H. D., Pouton, C. W., Charman, W. N., Bergström, C. A. S., & Porter, C. J. H. (2016). 50 years of oral lipid-based formulations: Provenance, progress and future perspectives. In *Advanced Drug Delivery Reviews* (Vol. 101, pp. 167–194). Elsevier B.V. https://doi.org/10.1016/j.addr.2016.04.007

Gautschi, N., Bergström, C. A. S., & Kuentz, M. (2016). Rapid determination of drug solubilization versus supersaturation in natural and digested lipids. *International Journal of Pharmaceutics*, *513*(1–2), 164–174. https://doi.org/10.1016/J.IJPHARM.2016.09.015

Griffin, B. T., Kuentz, M., Vertzoni, M., Kostewicz, E. S., Fei, Y., Faisal, W., Stillhart, C., O'Driscoll, C. M., Reppas, C., & Dressman, J. B. (2014). Comparison of in vitro tests at various levels of complexity for the prediction of in vivo performance of lipid-based formulations: Case studies with fenofibrate. *European Journal of Pharmaceutics and Biopharmaceutics*, *86*(3), 427–437. https://doi.org/10.1016/J.EJPB.2013.10.016

Holm, R. (2019). Bridging the gaps between academic research and industrial product developments of lipid-based formulations. *Advanced Drug Delivery Reviews*, *142*, 118–127. https://doi.org/10.1016/J.ADDR.2019.01.009

Hossain, S., Kabedev, A., Parrow, A., Bergström, C. A. S., & Larsson, P. (2019). Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning. *European Journal of Pharmaceutics and Biopharmaceutics*, *137*, 46–55. https://doi.org/10.1016/J.EJPB.2019.02.007

Kaukonen, A. M., Boyd, B. J., Charman, W. N., & Porter, C. J. H. (2004). *Drug Solubilization Behavior During in Vitro Digestion of Suspension Formulations of Poorly Water-Soluble Drugs in Triglyceride Lipids*.

Kaukonen, A. M., Boyd, B. J., Porter, C. J. H., & Charman, W. N. (2004). *Drug Solubilization Behavior During in Vitro Digestion of Simple Triglyceride Lipid Solution Formulations*.

Khan, J., Rades, T., & Boyd, B. (2016). The Precipitation Behavior of Poorly Water-Soluble Drugs with an Emphasis on the Digestion of Lipid Based Formulations. *Pharmaceutical Reaearch* , *33*, 548–562. https://doi.org/10.1007/s11095-015-1829-5

Koehl, N. J., Renéholm, R., Kuentz, M., Jannin, V., & Griffin, B. T. (2020). Exploring the Impact of Surfactant Type and Digestion: Highly Digestible Surfactants Improve Oral Bioavailability of Nilotinib. *Cite This: Mol. Pharmaceutics*, *17*, 3213. https://doi.org/10.1021/acs.molpharmaceut.0c00305

Kuentz, M., Holm, R., Kronseder, C., Saal, C., & Griffin, B. T. (2021). Rational Selection of Bio-Enabling Oral Drug Formulations – A PEARRL Commentary. In *Journal of Pharmaceutical Sciences* (Vol. 110, Issue 5, pp. 1921–1930). Elsevier B.V. https://doi.org/10.1016/j.xphs.2021.02.004

Lennernäs, H., Aarons, L., Augustijns, P., Beato, S., Bolger, M., Box, K., Brewster, M., Butler, J., Dressman, J., Holm, R., Julia Frank, K., Kendall, R., Langguth, P., Sydor, J., Lindahl, A., McAllister, M., Muenster, U., Müllertz, A., Ojala, K., … Abrahamsson, B. (2014). Oral biopharmaceutics tools – Time for a new initiative – An introduction to the IMI project OrBiTo. *European Journal of Pharmaceutical Sciences*, *57*(1), 292–299. https://doi.org/10.1016/J.EJPS.2013.10.012

Mehta, C. H., Narayan, R., & Nayak, U. Y. (2019). Computational modeling for formulation design. *Drug Discovery Today*, *24*(3), 781–788. https://doi.org/10.1016/J.DRUDIS.2018.11.018

Merisko-Liversidge, E., Liversidge, G. G., & Cooper, E. R. (2003). Nanosizing: a formulation approach for poorly-water-soluble compounds. *European Journal of Pharmaceutical Sciences*, *18*(2), 113–120. https://doi.org/10.1016/S0928-0987(02)00251-8

Murray, J. D., Lange, J. J., Bennett-Lenane, H., Holm, R., Kuentz, M., O'Dwyer, P. J., & Griffin, B. T. (2023). Advancing Algorithmic Drug Product Development: Recommendations for Machine Learning Approaches in Drug Formulation. *European Journal of Pharmaceutical Sciences*, 106562. https://doi.org/10.1016/J.EJPS.2023.106562

Persson, L. C., Porter, C. J. H., Charman, W. N., & Bergström, C. A. S. (2013). Computational Prediction of Drug Solubility in Lipid Based Formulation Excipients. *Pharmaceutical Research 2013 30:12*, *30*(12), 3225–3237. https://doi.org/10.1007/S11095-013-1083-7

Reppas, C., Kuentz, M., Bauer-Brandl, A., Carlert, S., Dallmann, A., Dietrich, S., Dressman, J., Ejskjaer, L., Frechen, S., Guidetti, M., Holm, R., Holzem, F. L., Karlsson, E., Kostewicz, E., Panbachi, S., Paulus, F., Senniksen, M. B., Stillhart, C., Turner, D. B., … O'dwyer, P. J. (2023). Leveraging the use of in vitro and computational methods to support the development of enabling oral drug products: An InPharma commentary. *European Journal of Pharmaceutical Sciences*, *188*, 106505. https://doi.org/10.1016/j.ejps.2023.106505

Ritchie, T. J., MacDonald, S. J. F., Young, R. J., & Pickett, S. D. (2011). The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discovery Today*, *16*(3–4), 164–171. https://doi.org/10.1016/J.DRUDIS.2010.11.014

Rosenberger, J., Butler, J., & Dressman, J. (2018). A Refined Developability Classification System. *Journal of Pharmaceutical Sciences*, *107*(8), 2020–2032. https://doi.org/10.1016/J.XPHS.2018.03.030/ATTACHMENT/7EAA972C-D702-4DFF-A9CF-CDEA77FAC0B9/MMC1.DOCX

Sacchetti, M., & Nejati, E. (2012). Prediction of drug solubility in lipid mixtures from the individual ingredients. *AAPS PharmSciTech*, *13*(4), 1103–1109. https://doi.org/10.1208/S12249-012-9830-3

Singh, A., & Van den Mooter, G. (2016). Spray drying formulation of amorphous solid dispersions. *Advanced Drug Delivery Reviews*, *100*, 27–50. https://doi.org/10.1016/J.ADDR.2015.12.010

Van Den Mooter, G. (2012). The use of amorphous solid dispersions: A formulation strategy to overcome poor solubility and dissolution rate. *Drug Discovery Today: Technologies*, *9*(2), e79–e85. https://doi.org/10.1016/J.DDTEC.2011.10.002

Williams, H. D., Sassene, P., Kleberg, K., Bakala-N Goma, J.-C., Calderone, M., Jannin, V., Igonin, A., Partheil, A., Marchaud, D., Jule, E., Vertommen, J., Maio, M., Blundell, R., Benameur, H., Carriére, F., Mullertz, A., Porter, C. J. H., & Pouton, C. W. (2012). *Toward the Establishment of Standardized In Vitro Tests for Lipid-Based Formulations, Part 1: Method Parameterization and Comparison of In Vitro Digestion Profiles Across a Range of Representative Formulations*. https://doi.org/10.1002/jps.23205

Yeap, Y. Y., Trevaskis, N. L., & Porter, C. J. H. (2013). *The Potential for Drug Supersaturation during Intestinal Processing of Lipid-Based Formulations May Be Enhanced for Basic Drugs*. https://doi.org/10.1021/mp400035z

Zupančič, O., Kushwah, V., & Paudel, A. (2023). Pancreatic lipase digestion: The forgotten barrier in oral administration of lipid-based delivery systems? *Journal of Controlled Release*, *362*, 381–395. https://doi.org/10.1016/J.JCONREL.2023.08.024

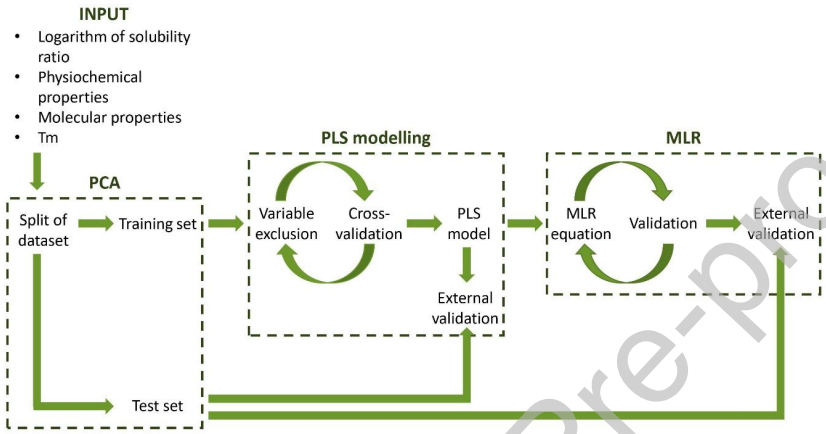

Figure 1: Schematic of the modelling approach. The physicochemical and molecular properties and the T_m of the drugs were used as input in the model together with the logarithm of the solubility ratio. The data set was split into a training and test set based on principal component analysis (PCA), whereafter the training set was used for a partial least square (PLS) modelling approach and the test set used for validation. Multiple linear regression (MLR) was performed by selecting independent descriptors on the training set and the test set was used for external validation.

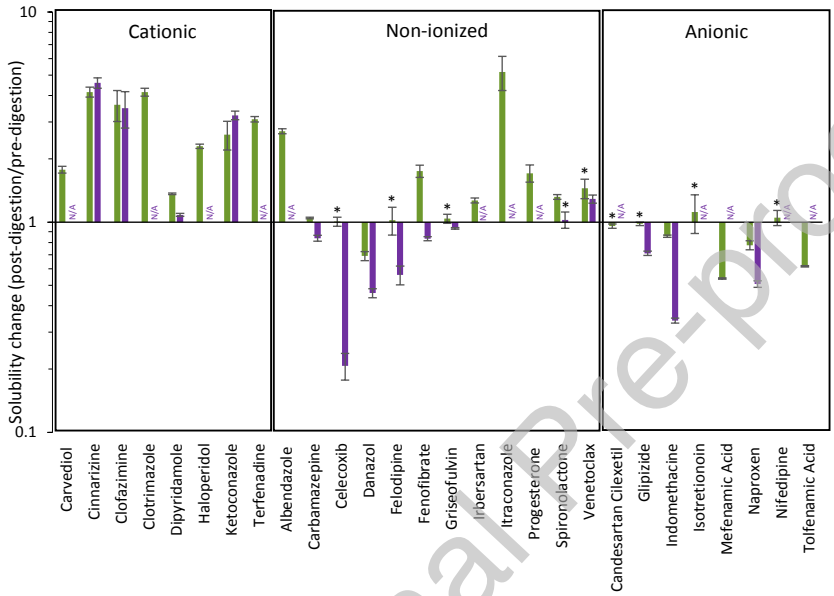

Figure 2: Solubility change (drug solubility post-digestion/pre-digestion) divided into cationic, non-ionized, and anionic drugs (pH 6.5). Generally, a solubility increase is seen for cationic drugs, and a solubility decrease or no influence is seen for anionic drugs. No trend is seen for solubility for the non-ionized drugs. The data shown are for long-chain LBF (green) and medium-chain LBF (purple). All data shown is n = 3 ± SE. The star (*) indicates that there was no significant difference between the drug solubility pre- and post-digestion. The N/A indicates experiments where data is not available because the post-digestion media exceeded the limit of the UV-probes.

Graphical Abstract

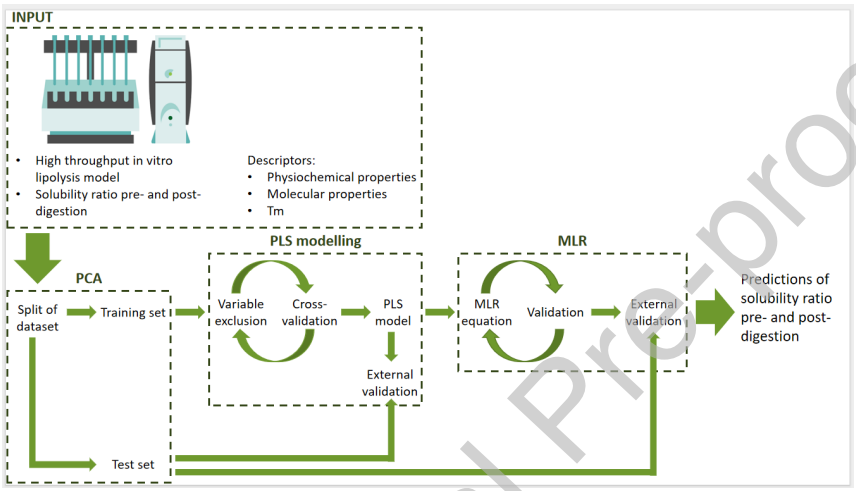

*Table 1: Selection of the physiochemical and molecular properties collected from the literature or ADMET Predictor 9.5. *The %ionised at post-pH is only for the long-chain LBF.*

Drug compound	MW (g/mol)	cLogP	LogD (pre-pH)	LogD (post-pH)	Acid/Base/Neutral	pKa	%ionised at pre-pH	%ionised at post-pH*	Tm (°C)	Max Dose (mg)
Albendazole	265.34	2.81	2.81	2.80	Ampholyte	10.26; 2.8	0; 0	0; 0	209	200
Candesartan cilexetil	610.67	5.70	2.89	2.92	Ampholyte	6	76	69	163	32
Carbamazepine	236.28	2.41	2.40	2.40	Basic	13.9	0	0	190.2	300
Carvedilol	406.49	3.94	2.43	2.28	Basic	7.8	95	97	114.5	25
Celecoxib	381.38	3.81	3.81	3.81	Acidic	11.1	0	0	158	200
Cinnarizine	368.53	5.01	4.08	3.95	Basic	8.4	99	99	119	25
Clofazimine	473.41	7.11	4.54	4.39	Basic	8.51	99	99	211	50
Clotrimazole	344.85	5.10	5.08	5.07	Basic	6.7	61	64	142	10
Danazol	337.47	4.26	4.26	4.26	Neutral	-	-	-	227	200
Dipyridamole	504.64	3.05	3.04	3.02	Basic	6.59	11	31	163	200
Felodipine	384.26	5.03	5.03	5.03	Basic	5.07	4	6	143	10
Fenofibrate	360.84	5.20	5.20	5.20	Neutral	-	-	-	79	150
Glipizide	445.54	2.14	1.50	1.60	Acidic	5.9	75	69	201.5	10
Griseofulvin	352.77	2.51	2.51	2.51	Neutral	-	-	-	220	500
Haloperidol	375.87	3.90	2.14	1.99	Basic	8.3	98	99	151	20
Indomethacin	357.80	4.03	1.45	1.58	Acidic	4.5	99	98	160	50
Irbesartan	428.54	3.73	2.90	3.02	Ampholyte	4.12; 7.4	0; 11	1; 9	180.5	300
Isotretinoin	300.44	6.07	4.00	4.15	Acidic	4	100	99	174	40
Itraconazole	705.65	4.89	4.89	4.89	Basic	3.7	0	0	166	100
Ketoconazole	531.44	3.74	3.72	3.68	Basic	6.75; 4.22	15; 0	34; 0	146	200
Mefenamic acid	241.29	4.90	2.36	2.51	Acidic	3.89	100	100	230.5	500
Naproxen	230.27	3.21	1.10	1.25	Acidic	4.15	100	99	153	500
Nifedipine	346.34	3.10	3.10	3.10	Acidic	3.93	100	100	173	300
Progesterone	314.47	3.94	3.94	3.94	Neutral	-	-	-	128	200
Spironolactone	416.58	3.28	3.28	3.28	Neutral	-	-	-	134.5	100
Terfenadine	471.69	5.62	3.64	3.49	Basic	10	100	100	147	60
Tolfenamic acid	261.71	5.13	2.44	2.57	Acidic	5.11	96	94	213	200
Venetoclax	868.46	6.84	6.70	6.71	Ampholyte	3.4; 10.3	100; 100	100; 100	138	100

Table 2: Composition of the LBF investigated.

Formulation	Excipients
Medium-chain LBF	40% Miglyol 812
	40% Tween 85
	20% Cremophor RH 40
Long-chain LBF	40% Olive oil
	40% Tween 85
	20% Cremophor RH 40

Table 3: rDCS Classification of the 28 drugs using pre- and post-digestion solubility values. Some data is absent for the medium-chain LBF because post-digestion media exceeded the limit of the UV-probes. The pre-digestion data is 24 hour equilibrium dispersion data acquired from (Bennett-Lenane et al., 2021). $^{\uparrow}$ *classification transition to a higher solubility classification compared to FaSSIF.*

		FaSSIF	Long-chain LBF			Medium-chain LBF		
			Pre-digestion	Post-digestion		Pre-digestion	Post-digestion	
	Drug	Do rDCS Class	Do rDCS class	Do rDCS class	rDCS class transition pre- and post-digestion	Do rDCS class	Do rDCS class	rDCS class transition pre- and post-digestion
Cationic	Carvedilol	I	I	I				
	Cinnarizine	IIa	I$^{\uparrow}$	I$^{\uparrow}$		I$^{\uparrow}$	I$^{\uparrow}$	
	Clofazimine	IIb	IIa$^{\uparrow}$	IIa$^{\uparrow}$		IIa$^{\uparrow}$	IIa$^{\uparrow}$	
	Clotrimazole	IIa	I$^{\uparrow}$	I$^{\uparrow}$				
	Dipyridamole	IV	IV	IV		IV	IV	
	Haloperidol	I	I	I				
	Ketoconazole	IIb	IIb	IIa$^{\uparrow}$	IIb → IIa	IIb	IIa$^{\uparrow}$	IIb → IIa
	Terfenadine	IIa	I$^{\uparrow}$	I$^{\uparrow}$				
Neutral	Albendazole	IIb	IIb	IIb				
	Carbamazepine	IIa	IIa	IIa		IIa	IIa	
	Celecoxib	IIb	IIa$^{\uparrow}$	IIa$^{\uparrow}$		I$^{\uparrow}$	IIa$^{\uparrow}$	I → IIa
	Danazol	IIb	IIb	IIb		IIa$^{\uparrow}$	IIb	IIa → IIb
	Felodipine	I	I	I		I	I	
	Fenofibrate	IIb	IIa$^{\uparrow}$	IIa$^{\uparrow}$		I$^{\uparrow}$	IIa$^{\uparrow}$	I → IIa
	Griseofulvin	IIb	IIb	IIb		IIb	IIb	
	Irbesartan	IV	IV	IV				
	Itraconazole	IIb	IIb	IIb				
	Progesterone	IIb	IIa$^{\uparrow}$	IIb	IIa → IIb			
	Spironolactone	IIa/b	IIa$^{\uparrow}$	IIa$^{\uparrow}$		IIa$^{\uparrow}$	IIa$^{\uparrow}$	
	Venetoclax	IV	IV	IV		III$^{\uparrow}$	IV	III → IV
Anionic	Candesartan Cilexetil	IV	III$^{\uparrow}$	III$^{\uparrow}$				
	Glipizide	III	III	III		III	III	
	Indomethacin	I	I	I		I	I	
	Isotretinoin	IIa	I$^{\uparrow}$	IIa	I → IIa			
	Mefenamic acid	IIb	IIa$^{\uparrow}$	IIb	IIa → IIb			
	Naproxen	IIa	I$^{\uparrow}$	I$^{\uparrow}$		I$^{\uparrow}$	IIa	I → IIa
	Nifedipine	IIb	IIa$^{\uparrow}$	IIb	IIa → IIb			
	Tolfenamic acid	IIa	IIa	IIa				

Table 4: Overview of the PLS and MLR models for the SR_{LC} based on drug descriptors. Tm = melting point, Log D at pH 6.5, N_AromR = the number of aromatic rings, CHaaCH =atom-type hydrogen E-state index for aromatic carbons, and Pi_FMi2 = Second component for the autocorrelation vector of scaled pi Fukui-indices (electrophilic).

PLS model (post-digestion)					
Y-variable	Log (SR_{LC})				
X-variable	Tm, $LogD_{pre-pH}$, N_AromR, SaaCH, Pi_FMi2				
Explained Y-variance	76.1 %				
No. of principal components	2				
RMSEE	0.27				
$RMSE_{CV}$	0.30				
RMSEP test set	0.35				
R^2 (calibration)	0.76				
Q^2 (validation)	0.64				
MLR Equations					
Y-variable	R^2	$RMSE_{Training}$	$RMSE_{Test}$	F-value	p-value
Log (SR_{LC})	0.68	0.28	0.43	11.37	$3.1 \cdot 10^{-4}$

$$\mathrm{Log}\,(\mathrm{SR_{LC}}) = 0.81 + 0.15 \cdot (\mathrm{Log\,D})_{\mathrm{pre-pH}} + 0.14 \cdot (\mathrm{N_{AromR}}) - 0.006 \cdot (\mathrm{T_m})$$

Table 5: Overview of the PLS and MLR models for the SR_{LC} based on drug descriptors. The pre-digestion data used is 24 hour equilibrium dispersion data *acquired from (Bennett-Lenane et al., 2021). T_m = melting point, Log D at pH 6.5, N_AromR = the number of aromatic rings, NAP_MinQ + the minimum NPA partial atomic charge, Pi_Aqc = the sum of absolute values of pi partial atomic charge on carbon, and Pi_Fmi4 = Fourth component for the autocorrelation vector of scaled pi Fukui-*

PLS model (pre-digestion)					
Y-variable		Log (SR_{LC})			
X-variable		Tm, LogD, N_AromR, NPA_MinQ, Pi_Aqc, Pi_FMi4			
Explained Y-variance		80.1 %			
No. of principal components		1			
RMSEE		0.20			
$RMSE_{CV}$		0.23			
RMSEP test set		0.39			
R^2 (calibration)		0.80			
Q^2 (validation)		0.72			
MLR Equations					
Y-variable	R^2	$RMSE_{Training}$	$RMSE_{Test}$	F-value	p-value
Log (SR_{LC})	0.70	0.24	0.39	12.99	$1.2 \cdot 10^{-4}$

$$\mathrm{Log}\,(\mathrm{SR_{LC}}) = 0.31 + 0.18 \cdot (\mathrm{LogD}) + 0.13 \cdot (\mathrm{N_{AromR}}) - 0.003 \cdot (\mathrm{T_m})$$

indices (electrophilic).

Declarations of interest

KJB is an employee of Pion Inc. The other authors disclosed no conflicts of interest related to this article.