Journal of Pharmaceutical Sciences 115 (2026) 104262

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Pharmaceutics, Drug Delivery and Pharmaceutical Technology

A computationally informed framework for lipid-based formulation developability: Integrating predictive models with case studies

Callum D. Ryan[a,b], Joseph P. O'Shea[a], Brendan T. Griffin[a,b,*]

[a] School of Pharmacy, University College Cork, College Road, Cork, Ireland
[b] SSPC, The Science Foundation Ireland Research Centre for Pharmaceuticals, School of Pharmacy, University College Cork, Ireland

ARTICLE INFO

Article history:
Received 22 December 2025
Revised 23 March 2026
Accepted 23 March 2026
Available online 24 March 2026

Keywords:
Lipid-based formulations
Poorly soluble drugs
Decision trees
Computational predictions

ABSTRACT

The increasing prevalence of poorly water-soluble drugs (PWSDs) in development pipelines has intensified the need for bio-enabling strategies such as lipid-based formulations (LBFs). However, formulation selection for optimal bioavailability often relies on empirical, resource-intensive approaches. This study introduces a computationally informed developability framework that integrates predictive models to streamline LBF development. The framework combines *in silico* tools for assessing key determinants of LBF suitability—including predictions of food effect, lipid solubility, and biopharmaceutical dose number (Do)—with tailored strategies such as lipophilic salt synthesis for challenging molecules. A database of >200 FDA-approved oral drugs (2010–2023) was screened through the computational framework, and four case studies (dapagliflozin, mavacamten, ospemifene, flibanserin) illustrate the framework's ability to classify risk and guide formulation decisions. Predictions identified low-risk candidates for or type-I LBFs, highlighted medium-risk molecules requiring type-IIIa formulations, and demonstrated the utility of lipophilic salts for drugs which are high-risk with respect to dose-loading and dose number limitations. This integrated approach enables rapid, data-driven formulation decisions, reducing reliance on trial-and-error methods and supporting early-stage development of PWSDs. By coupling computational predictions with targeted experimental validation, the framework offers a practical roadmap for accelerating LBF adoption and improving R&D efficiency.

Introduction

The need for "bio-enabling" techniques such as lipid-based formulations (LBFs) and amorphous solid dispersions among others[1–3] is vital given that over 70% of molecules emerging from drug discovery are classified as poorly water-soluble drugs (PWSD).[4] LBFs, which combine lipids, surfactants, and co-solvents, can enhance drug solubilization, mimic the physiological effects of dietary lipids and improve intestinal permeability. Despite these advantages, the adoption of LBFs has declined in recent years, largely due to resource-intensive development processes and favours familiar, empirically driven approaches that are typically conservative.[3] Developing multiple bio-enabling formulations in parallel, although resource-intensive, is a commonly employed strategy to minimize the risk of overlooking the optimal formulation. Consequently, there is significant potential to accelerate and improve developability decisions by applying data-informed computational tools to predict optimal formulation strategies for specific drugs.[5,6]

Current selection criteria for LBF suitability typically emphasize properties predictive of lipid solubility, such as high log P (>5), substantial lipid solubility, and low melting point.[7,8] While these rules provide a useful starting point, their conservative nature often results in rigid, all-or-nothing decisions, that may not fully capture the complexity of modern drug discovery pipelines. In practice, drug properties influencing crystallinity (*e.g.*, melting point) and solvation (*e.g.*, log P) are interdependent, and many molecular properties are interconnected. Furthermore, focusing solely on lipid solubility overlooks the importance of predicting solubility in gastrointestinal media formed during dispersion and digestion of LBFs. If the biopharmaceutical dose number (Do) is favourable *in vivo*, alternative strategies—such as supersaturated lipids or lipophilic salts—can overcome dose-loading limitations. Therefore, an integrated approach is required—one that considers both lipid solubility and intestinal solubility, while leveraging formulation technologies to unlock the full potential of LBFs for emerging drug candidates.

* Corresponding author.
E-mail address: brendan.griffin@ucc.ie (B.T. Griffin).

https://doi.org/10.1016/j.xphs.2026.104262

The adoption of AI-driven tools in drug discovery and development is accelerating rapidly. Recent industry reports indicate that integrating computational tools and AI-based predictions into R&D workflows can deliver cost and time savings of 25–50%.[9] Within this evolving drug development, advancing computational pharmaceutics clearly plays a pivotal role. Opportunities exist to move beyond simple drug-likeness filters by leveraging the broader chemical design space and predictive models already developed.[5,10] Computational tools based on molecular structure can guide rational formulation strategies and reduce the high costs associated with trial-and-error approaches.[11] For early-stage formulation selection, high levels of precision are not essential; rather, predictions should serve as directional guidance on formulation risk. Such tools can identify candidates that are low risk for development as LBFs, easing the transition from drug substance to drug product. While experimental validation remains necessary, structured computational guidance that highlights low-risk candidates can significantly streamline development pipelines.[12]

This work proposes a computationally informed developability framework for LBFs that combines predictive modelling with targeted experimental validation. The aim is to integrate computational predictions that can forecast LBF suitability including propensity for food effect on bioavailability, lipid solubility and biopharmaceutical dose number following dispersion and digestion of LBFs.[13–16] Incorporation of fail-safe nodes or tailored investigations to the framework, if undesirable outcomes were predicted, would also be required. The hypothesised framework aims to emulate the biopharmaceutics risk assessment roadmap (BioRAM) with a "learn and confirm" approach[12] whereby computational tools can inform formulation decision making which will then be experimentally validated.

The developability framework proposed herein builds upon stepwise decision trees, a widely adopted and well-established strategy in pharmaceutical formulation design.[17–19] Such structured frameworks provide systematic guidance for formulation selection and excipient screening, thereby streamlining complex development workflows. In this context, Holm (2019) emphasized the importance of flow diagrams to support lipid-based formulation design, underscoring the industrial relevance of transparent, simplified decision tools that facilitate effective decision making in development.[20] By integrating computational prediction with targeted experimental validation, the present framework enables early-stage, cost-efficient formulation decisions that account for both physicochemical drug attributes and key biopharmaceutical parameters. Previous studies have demonstrated the successful application of interpretable machine learning models and reducing empirical optimisation, in developing solid oral dosage form products, illustrating how predictive modelling can inform formulation choices before experimental screening.[21,22] However, such predictive frameworks have not been systematically integrated into lipid-based formulation strategy selection. Collectively, these examples establish a clear precedent for integrating predictive modelling with structured decision logic in formulation workflows and provide context for the approach presented in this work.

To facilitate the development of a developability framework for LBFs, the integrated models were used to screen 206 drugs which had been licenced between 2010 and 2023 by the FDA. Through this process a range of drugs with varying risk regarding "developability" as an LBF were identified. Confirmatory *in vitro* experiments were performed with case studies using four drugs (Table 1) to illustrate both the initial information obtainable from the computational decisions and the applicability of the developability framework:

1) A preliminary prediction on whether a drug candidate is anticipated to display a pharmaceutical food effect.
2) Predictions of drug solubility in medium-chain triglyceride LBF.
3) An estimate of the predicted Do of a drug following dispersion and digestion of the drug in biorelevant media.
4) Utilising lipophilic salt synthesis as a strategy to enhance dose loading in the event of poor DL.

Theoretical section

This section outlines the critical attributes and computational approaches integrated into the proposed LBF developability framework. The framework combines standard investigations—such as rDCS classification and food-effect prediction—with tailored investigations triggered by specific risk factors, including dose-loading limitations and solubility challenges. These steps aim to guide formulation decisions from early candidate selection through Phase I readiness. The framework adopts a stepwise approach inspired by the Biopharmaceutics Risk Assessment Roadmap (BioRAM) and rDCS principles.[12,23] Initial predictions assess:

1. rDCS classification and need for bio-enabling strategies.
2. Food-effect likelihood using machine learning models.
3. Drug solubility in medium-chain triglycerides (MCTs).
4. Biopharmaceutical dose number (Do) after dispersion and digestion.
5. Feasibility of lipophilic salt synthesis for challenging molecules.

Standard investigations

Standard investigations utilised here mimic the rDCS and are intended to give a first impression of whether a new oral drug candidate can be easily formulated as an immediate release dosage form with minimal risk associated. The standard investigations here incorporate the addition of a prediction of a potential pharmaceutical food effect.

Do and rDCS classification

Determining the rDCS classification of the drug and whether a bio-enabling approach such as LBF is required is a critical first step in drug development. Generally, experimentally determined inputs such as drug solubility in FaSSIF, permeability and dose are required

Table 1
List of drugs utilised as case studies.

Drug compound	Commercial dose	Molecular weight (g/mol)	rDCS class	LogP	Melting point
Dapagliflozin	10mg	408.87	III	2.52[A]	86°C[B]
Mavacamten	15mg	273.33	IIa	2.3[A]	224.5°C[B]
Ospemifene	60mg	378.9	IIa	5.7[A]	135°C[B]
Flibanserin	100mg	390.4	IIa	3.82[A]	162°C[B]

[A] Calculated by ChemAxon.
[B] Determined via DSC.

to determine the rDCS classification of the drug in question. It is recommended that, in the absence of reliable dosing information, 3 default doses, 5 mg, 50 mg, and 500 mg, are applied for use in both standard and tailored investigations as per the rDCS.[23] Further to keep in line with the predictive nature of the proposed framework both the permeability of the drug and solubility were predicted using ADMET predictor 9.5. The solubility was criteria was determined by calculating a biopharmaceutical dose number (Do) using a dose/solubility ratio (Eq. 1)

$$Do = \frac{dose}{S_{si} \times V_{si}} \quad (1)$$

where, the dose is the highest dose, S_{si} is the apparent solubility in biorelevant media; predicted drug solubility in FaSSIF was used in this study, and V_{si} is the volume in the small intestine available for dissolution.[24,25] Drugs which fall into the category of rDCS class I can be formulated as conventional immediate release drugs due to their high solubility and permeability. Drugs which fall into the other 3 classifications can be progressed through the framework based on their need to improve their bioavailability for them to be commercially and clinically successful.

Prediction of food effect

At the drug discovery phase after identification of an oral drug candidate and its rDCS classification, the principal consideration of the BioRAM is a quality target product profile (QTTP) which includes key product attributes deemed clinically relevant for achieving product success. The presence of a pharmaceutical food effect is one such attribute that a QTTP tends to explore. A food effect can be defined as a significant change in the rate or extent of drug absorption between a drug being administered in the fed or fasted state.[26]

Early assessment of food effect can influence formulation strategy, including whether to proceed with development or explore alternative dosage forms to minimise food-dependent bioavailability. LBFs were originally proposed to mimic the positive food effect observed for PWSDs by co-administering lipids instead of food. Thus, the presence or absence of a food effect serves as an early indicator of LBF suitability.[26] Briefly, the employed model classifies whether a drug will display a food effect based on determining a relationship between significant drug properties including; Octanol Water Partition Coefficient (LogP), Number of Hydrogen Bond Donors (HBD), Topological Polar Surface Area (T_PSA) and Dose (mg).[13] ANN detects complex non-linear patterns between datasets. ANN avails of a general structure composed of input, output and hidden layers while using activation functions, connection weighs, nodes in the form of artificial neurons from the input layer to relay data to the hidden and output layers via weighted connections (synapses) to predict Y, in this case, food effect.[27,28] Drugs which displayed a positive or neutral food effect and were in need of a bio-enabling formulation due to their rDCS classification were progressed to the next stage in the framework. Drugs which display negative food effects would be excluded and require an alternative bio-enabling strategy.

Tailored investigations

Tailored investigations are triggered if the standard investigations indicate that such approaches are required. Tailored investigations address specific formulation challenges for LBFs. The criteria which trigger such criteria can be found in Table 2, with a risk level for a new drug compound associated with the investigation outlined in Table 3.

If no LBF specific tailored investigation is triggered initially, an alternative formulation strategy would be recommended. Likewise, if all tailored investigations fail and further assessment indicates the need for a different approach, the candidate is deemed unsuitable for LBF development, and an alternative bio-enabling strategy should be pursued. The fundamental theory behind each of these investigations will be explained in greater detail in the following sections.

Table 2
Triggers for the tailored investigations and why such investigations are required.

Trigger	Tailored investigation	Reason
Bio-enabling approach required due to rDCS classification & predicted food effect	Predict dose loading ($DL_{Predicted}$) in MCT	Predicted lipid solubility of the drug needs to be sufficient to allow the dose of the drug to be dissolved in MCT.
DL in Type-I formulation >5	Predict Do	Assessment of Do increases through formulation as a LBF
Do > 1, high SR & drug ionisable	Lipophilic salt strategy	Lipophilic salt could allow for higher dose loadings in type-IIIA formulations with subsequent improvement to Do.
DL >5/Do > 1, unioniable	Alternative strategy required	Drug candidate may not merit LBF technology/lipid solution unachievable.

Table 3
Different risk levels (low to high) for the assessment of effort required to develop a drug candidate as an LBF.

Tailored investigations triggered	Risk level
0	Alternative strategy required
1	Low
2	Medium
3	High
4	Alternative strategy required

Drug solubility in type I LBF

Pre-formulation profiling is critical for LBF development, particularly in assessing drug solubility in lipid excipients. However, experimental solubility measurements are resource-intensive. Leveraging predictions of solubility in lipid excipients can streamline decision-making by identifying appropriate excipients for a given candidate. Since successful LBFs require the drug to dissolve in lipid vehicles, solubility in medium-chain triglycerides (MCTs), a commonly utilized lipid excipient, forms the first stage of tailored investigations.

The Yalkowsky equation (Eq. 2) highlights a relationship between solubility (S), melting point (MP in K), and log P, where P is the partition-coefficient (P).

$$LogS = 0.8 - LogP - 0.01(MP - 0.25) \quad (2)$$

Poorly soluble drugs often fall into two categories: “brick dust” compounds, with strong crystal lattices and limited aqueous solubility, and “grease-ball” compounds, which exhibit high lipophilicity and low melting points. This relationship is clearly outlined by the Yalkowsky equation. While lipid-based excipients and surfactants can enhance solubility for grease-ball compounds, brick-dust drugs may require alternative strategies to improve dose loading, as triglyceride solubility is often insufficient.

Commercially, most LBFs are Type I formulations, favoring simplicity with single excipients.[29,30] Accordingly, the framework prioritizes Type I LBFs. Predictive models based on molecular descriptors typically achieve an RMSE of ~1 log unit (≈10-fold accuracy) compared to experimental values.[31] These predictions should guide LBF development, rather than replace, experimental validation. Drugs which are predicted to display a lipid dose number of <5 (calculated using Eq. 3) are considered low developability risk using

conventional LBFS, DL > 5 but < 50 triggers further investigation using surfactants or co-solvents. DL > 50 indicates high developability risks as a LBF warranting alternative strategies such as amorphous solid dispersions. The cut-offs for predicted dose loading of 5 and 50 are arbitrarily, informed by both the variance in prediction accuracy and the reported potential gains achievable through appropriate excipient selection or dose loading strategies for formulation technologies.[32,33]

$$DL_{Predicted} = \frac{dose}{Lipid\ Solubility} \quad (3)$$

Several machine learning models predict solubility in various lipid excipients.[16,34,35] In this framework, a model developed by Lange et al. was employed, using quantitative structure−property relationships derived from 182 structurally diverse drugs, however other models could also be employed to demonstrate predicted solubility in MCTs.[34,35] A list of the predictive outputs with respective descriptors and software utilised to determine lipid solubility in MCTs can be found in Table 2. Two case studies were conducted for predicting drug solubility in MCTs, one using 2D/3D descriptors and one using SOAP descriptors to demonstrate the applicability of this investigation using different models.

Assessment of biorelevant solubility change during dispersion and digestion

Drugs which display DL of greater than 5 but less than 50 proceed to the next stage of the development framework. Given that DL challenges have been identified at this stage, a type-IIIa LBF was selected. In contrast to poorly dispersible Type I systems, Type IIIA LBFs are mixtures of oils and surfactants that form self-emulsifying drug delivery systems (SEDDS) on dispersion in intestinal fluids. Type IIIA LBFs contain relatively high proportion of triglyceride (between 40−80%), relative to Type IIIB/IV systems, yet maintain an ability to readily self-emulsify to form a microemulsion that improves drug solubilization.

To predict biorelevant solubility during dispersion and digestion within a Type-IIIa LBF, the model developed by Bennett-Lenane and Ejskjær is applied.[14,15] Both models utilize Partial Least Squares (PLS) regression to estimate biorelevant solubility changes before and after digestion. The PLS approach extracts latent factors from **x** (drug molecular descriptors) and **y** (biorelevant solubility) to maximize covariance, establishing a linear relationship between these matrices while identifying key molecular descriptors. Details of these descriptors and the software used are provided in Table 2.

While predicting lipid solubility to guide dose loading is important, it is not the sole criterion for LBF suitability. Understanding the biorelevant solubility change *in vitro* can help inform future *in vivo* performance of LBFs through utility of a computationally informed biopharmaceutical dose number relative to the rDCS classification system. For rDCS classification using the predicted solubility ratio gains (SR) $C_{S(Predicted)}$ is used which is calculated using Eq. 4. The equations detailed here can be used for determining the Do after both dispersion and digestion.

$$C_{S(Predicted)} = SR \times S_{Si} \quad (4)$$

where SR is the predicted solubility ratio upon dispersion and digestion of the type-IIIA LBF from the predictive models and Ssi is apparent solubility of the compound in biorelevant media (FaSSIF). Incorporating Eq. 3, solubility criteria for rDCS classifications upon formulation dispersion and digestion was predicted using Eq. 5:

$$Do_{Predicted} = \frac{dose}{C_{S(Predicted)} \times V_{si}} \quad (5)$$

Solubility criteria for rDCS classification using experimentally determined biorelevant solubility after dispersion and digestion can be calculated using Eq. 6:

$$Do_{Experimental} = \frac{dose}{C_S \times V_{si}} \quad (6)$$

where, $Do_{Experimental}$ is the biopharmaceutical dose number for type-IIIA LBF, the dose is the highest dose, C_S is the apparent solubility in biorelevant media after dispersion or digestion in biorelevant media, and V_{si} is the volume in the small intestine available for dissolution.

The calculation of the $Do_{Experimental}$ produced in intestinal fluid, allowing for an assessment of the drugs rDCS class based on Eqs. 4–6 and whether a transition in rDCS class is likely.

Following dispersion and digestion of the formulation, coupled with the fact the drug is introduced into the biorelevant media in a pre-dissolved state at an appropriate dose loading, drugs which ordinarily might display poor biorelevant solubility could be effectively retained in a dissolved state in biorelevant media, which could *in vivo* translate to improved absorption. This data-driven approach supports informed developability decisions and optimizes formulation screening resources. Drugs with a $Do_{Predicted}$ below 1 should proceed to *in vitro* testing; otherwise, a third investigation into lipophilic salt synthesis may be warranted.

Lipophilic salt strategy

The final criteria considered for use in developing a framework to assist in the development of LBFs is to consider ways of increasing dose loading in lipid excipients. While approaches such as inducing supersaturation or formulating lipid suspensions show promise for overcoming dose loading limitations, particularly in preclinical toxicity and biopharmaceutical assessment, lipid solutions remain the preferred format for commercial LBF development. Therefore, the lipophilic salt strategy was adopted within this framework as a strategy to increase dose loading. Lipophilic salts are defined as salt forms which display increased drug solubility in lipids relative to the free form of the drug.[32] This is brought about through pairing an ionisable drug with a bulky, asymmetric, lipophilic counterion, such as sodium docusate or alkyl sulfates.[32,36–38] Lipophilic salts of the drug substance display higher solubilities in lipid excipients compared to the free form of the drug allowing drugs which have a DL limitations to be formulated as LBFs.[32] Furthermore, previous work has demonstrated that lipophilic salts tend to display enhanced affinity for colloidal species formed during dispersion and digestion of a formulation relative to a free form of the drug, thus further improving the biopharmaceutical performance of the drug with potential enhancements to Do.[36,37,39]

While there is currently no, machine learning algorithm that can predict drug suitability for lipophilic salt synthesis, a classical drug likeness filter approach using the ΔpKa rule can be used to assist in salt formation screening. The rule states that if the difference in pKa or ΔpKa between a drug and a counterion is greater than 3, salt formation is likely to occur. Knowledge of this rule can help to inform suitable counterions that can be used to form lipophilic salts of an ionisable compound. Recently significant work has been endeavoured to convert numerous challenging drug molecules to lipophilic salt/ionic liquid type complexes. The work has proven to be successful with many robust methods available for synthesising salt complexes of challenging molecules and both *in vivo* and *in vitro* benefits have also been reported.[32,36–38,40] As such any developability framework relating to LBF development should include the lipophilic salt strategy as a fail-safe node. Thus, the final recommendation for the developability framework would be that if an ionisable drug displays

Table 4
List of predictive outputs with respective descriptors and software utilised to assess potential for development of LBFs.

Predicted output	Technique	Descriptors required	Software utilised
Melting point	Regression based gradient model	SMILES	AAT bioquest Melting point Predictor
Food effect on bioavailability[13]	Artificial neural networks (ANN)	HBD, T_PSA, Dose, LogP	SPSS
Solubility in medium chain triglyceride[16]	Machine learning (ML)	2D/3D profiling, smooth overlap atomic positions (SOAP), Tm	SciKit learn
Solubility ratios (dispersion)[14]	Partial least square (PLS)	Tm, LogD, N_AromR, NPA_MinQ, Pi_Aqc, Pi_FMi4	SIMCA
Solubility ratios (digestion)[15]	PLS	Tm, LogD, N_AromR, SaaCH, Pi_FMi2	SIMCA

suitability for an LBF but has dose loading limitations, experimental trials should be conducted to explore the lipophilic salt strategy.

Methods and materials

Ospemifene, dapagliflozin, mavacamten and flibanserin were purchased from Kemprotec Ltd. (United Kingdom). RH40, Tween 85 and Capmul were all purchased from Merck (Ireland). Peceol and Capmul were gifted from Gattefosse (France). All solvents were of analytical grade and were purchased from Merck (Ireland) and used as received.

Database collection & compilation of physicochemical descriptors

The drug database used in this study was obtained from drug products licenced by FDA from January 2010 to December 2023. (supplementary information). The database of original new drug applications on the FDA website was searched by month for each year within the specified timeframe. Exclusion criteria were any non-oral product and any biological product. SMILES strings were obtained for all compounds using PubChem and were subsequently used to determine over 250 molecular descriptors for each drug using ADMET Predictor 9.5 (Simulations Plus, USA). A melting point predictor developed by AAT Bioquest was used to obtain melting points for drugs where the melting point was not known.

Computational predictions

Four predictive models from previously published work were used for the purpose of this study. Table 4 present details of training sets, test sets, descriptors and software utilised. The entire file containing all the physicochemical properties obtained for all drugs within the database as well as subsequent predictive outputs can be found in the supplementary information. Drugs which were used in the training set of the models were excluded from being used as case study molecules.

Confirmatory solubility studies

Solubility studies in lipid excipients were conducted for each drug in type I and type III-A formulations. The composition of these formulations can be found in Table 5. The formulations were prepared by weighing in exact amounts of excipients into a glass vial. The formulations were stirred at 37°C at 300RPM overnight (Mixdrive 15 2MAG, Germany).

Table 5
List of formulations utilised in the study.

Formulation	Constituents
Type-I long chain	Olive oil (100%)
Type-I medium chain	Miglyol-812 (100%)
Type-IIIA long chain	Olive oil (40%), Tween-85 (40%), RH40 (20%)
Type-IIIA medium chain	Miglyol-812 (40%), Tween-85 (40%), RH40 (20%)

An excess amount of each drug was added to 2 g of the formulation in a screw top-glass vial. Samples were stirred at 200RPM (Mixdrive 15 2MAG, Germany) at 25°C. Samples were withdrawn after 24, 48 and 72 hours, followed by centrifugation at 21,000 g (Mikro 200R, Andreas Hettich GmbH & Co. KG, Germany) for 30 min. The supernatant was removed and centrifuged again under identical conditions. The resulting supernatant was solubilised 1:10 v/v in a mixture of acetonitrile and ethyl acetate (1:4 v/v), followed by a further 1:10 v/v dilution in acetonitrile and ethyl acetate (4:1 v/v). Samples were analyzed by HPLC as per the method available in the supplementary information. All samples were analysed in triplicate.

Media preparation

The media used in this experiment comprised 2 mM Tris-maleate, 150 mM NaCl, and 1.06 mM $CaCl_2$ according to instructions from biorelevant.com (UK). The buffer was adjusted to pH 6.5 and supplemented with FaSSIF powder 24 hours in advance of an experiment resulting in a final composition of 3 mM taurocholate and 0.75 mM phospholipids. The final solution will be referred to as “biorelevant media” throughout this paper.

In vitro *dispersion and digestion studies*

The *in vitro* dispersion and digestion studies were conducted using a pH-stat apparatus (Metrohm AG, Herisau, Switzerland) comprising an 836 Titando, an 804 Ti Stand, a pH electrode (Metrohm), and two 800 Dosino 20 mL dosing units. The system was operated by Tiamo 2.5 software (Metrohm). Pancreatin extract (8 × USP) was reconstituted immediately prior to use by adding 5 mL of biorelevant medium to 1 g of pancreatin, vortexed thoroughly, and centrifuged at 4000 g for 15 min at 4°C (Centrifuge 5702 R, Eppendorf, Germany); 4 mL of the resulting supernatant was recovered and immediately stored on ice before further usage. The LBF was introduced to biorelevant medium in a ratio of 1:40 (v/v) and stirred for 15 min to disperse the formulation. Media pH was automatically adjusted to pH 6.5, and the digestion was initiated through introduction of the lipase. A total volume of 40 mL of digestion media was reached following the addition of pancreatic lipase. The pH was kept at pH 6.5 throughout the experiment by automatic titration with 0.6 M NaOH for the medium-chain formulations and 0.2 M NaOH for the long-chain formulations. The amount of NaOH dispensed was recorded by the system and used to assess the rate and extent of digestion. After 60 min of digestion, the pH was back-titrated to pH 9 to determine the release of nonionized free fatty acids. Samples of 1 mL were withdrawn at 5 and 10 min during the dispersion phase. Pancreatin was then added, and samples of 1 mL were again withdrawn at 5, 10, 20, 40, and 60 min during the digestion

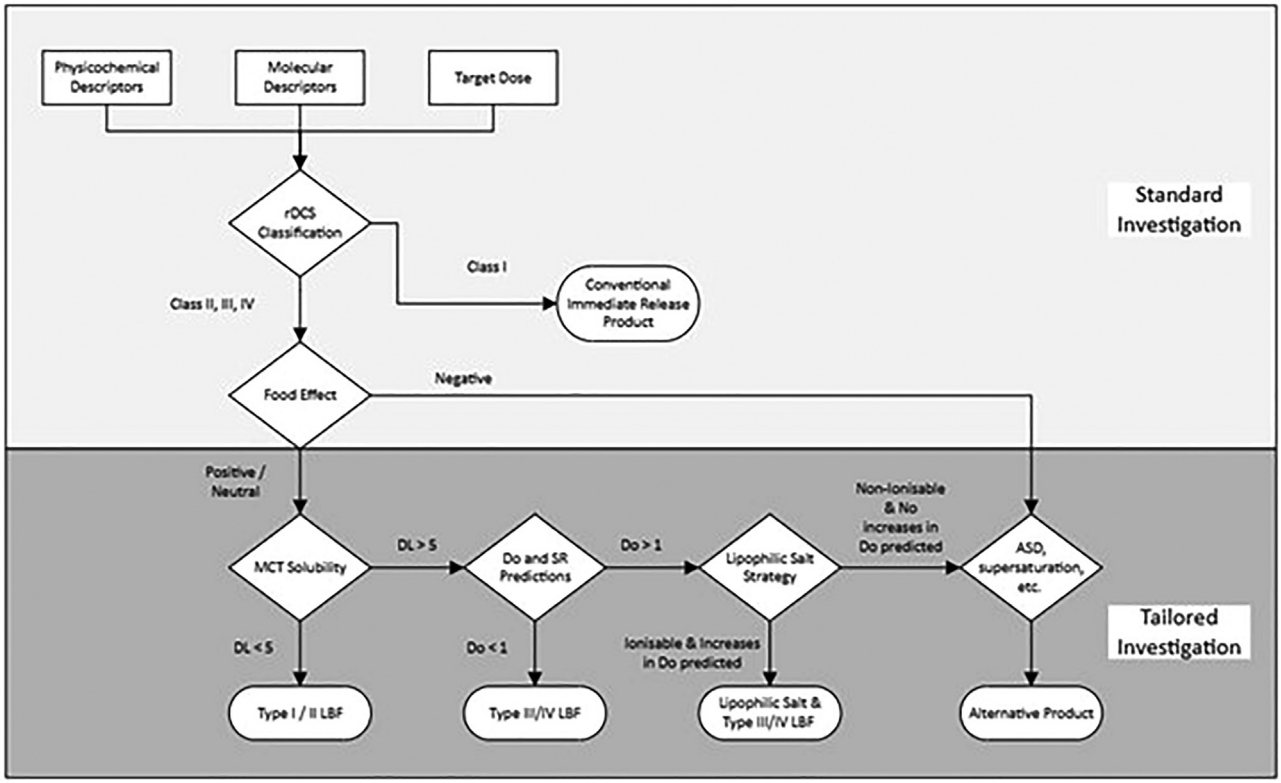

Fig. 1. Visual incorporation of the computational and *in vitro* tools utilised to form the lipid-based formulation developability framework as part of a refined drug substance to drug product development paradigm. Initial standard investigations using predictive modelling are performed relating to rDCS classification and food effect for the candidate drug. If the drug is classified as a type II, III or IV drug where low solubility and or permeability is reported, a set of tailored investigations are recommended. Computational predictions can be utilised to predict drug solubility in medium-chain triglycerides and biopharmaceutical dose number after dispersion and digestion.

phase. Each sample was immediately treated with 1 M 4-bromophenylboronic acid in methanol (5 μL/mL) to terminate the digestion process. The samples were kept at 37°C until centrifugation. The samples were centrifuged at 37°C and 21,000 g for 30 min using a benchtop centrifuge (Mikro 200 R, Hettich, Germany) to separate the drug that precipitated out of the solution and that remained solubilized. HPLC was used to quantify the concentration of the drug in the solution, the methods can be found in the supplementary information.

Lipophilic salt synthesis

Flibanserin lipophilic salts were prepared using a general procedure. Flibanserin free base and sodium docusate or sodium dodecyl suflate were dissolved in 150 mL of a biphasic solution of dichloromethane and water (1:1). 1 mmol of methanolic HCl was added to the organic phase. The mixture was stirred at room temperature overnight. The biphasic mixture was transferred to a separatory funnel and the organic phase was collected. The aqueous phase was washed using dichloromethane and the organic phase was again retained. The combined organic phases were backwashed with distilled water until a negative silver nitrate (0.02 M aq) precipitate result was obtained. The organic solution was dried with anhydrous sodium sulfate, filtered and concentrated *in vacuo*. The obtained product was placed under vacuum for 48 hours. The lipophilic salts were analysed using HPLC, differential scanning calorimetry (DSC) and nuclear magnetic resonance spectroscopy (NMR). The characterisation methods utilised can be found in the supplementary information.

Results

To explore the applicability of the proposed LBF developability framework, a set of over 200 drugs was identified as outlined previously. The database was screened using predictive algorithms assessing food effect, solubility in Type I MCT LBF, and the predicted dose number (Do) for a Type IIIA LBF formulation. Based on these modelling results, four representative drugs were selected for case studies. Each case study applied a tailored cascade of investigations aligned with the specific requirements of the drug, demonstrating the practical utility of the framework (Fig. 1).

Case study 1: tailored investigation of dose loading

Dapagliflozin, first licenced by the FDA in 2012, is licensed for the treatment of type-II diabetes, heart failure and chronic kidney disease with a clinical dose of 10 mg. Dapagliflozin is a crystalline drug with a melting point of 88°C and a cLogP of 2.82. Standard investigations classified dapagliflozin as a rDCS class III molecule which displayed a neutral food effect. A tailored investigation was triggered to determine dapagliflozin solubility in Type I MCT LBFs. Using 2D/3D molecular descriptors, dapagliflozin was predicted to have an approximate lipid solubility of 4.13mg/mL (range: 1.3 – 13.10mg/mL) in MCTs. This equated to a lipid dose number of 2.42 resulting in dapagliflozin progressing to experimental screening studies in MCTs as the requirement to proceed to confirmatory experiments from the tailored investigation was met. Experimental solubilities of dapagliflozin in miglyol-812 was determined to be 17.3±1.1mg/mL, enabling formulation as an LFCS Type I LBF using <1 mL of excipient. Dapagliflozin was thus considered to be a low-risk candidate for developability as an LBF as only one tailored investigation was triggered (Fig. 2).

Considering the chemical structure of dapagliflozin, the presence of a number of hydroxyl groups would negatively impact on drug solubility in MCT, however the presence of two benzene rings which would positively impact solubility could explain the discrepancy between predicted and experimental lipid solubility values - likely due to this chemical space being underrepresented in the model's

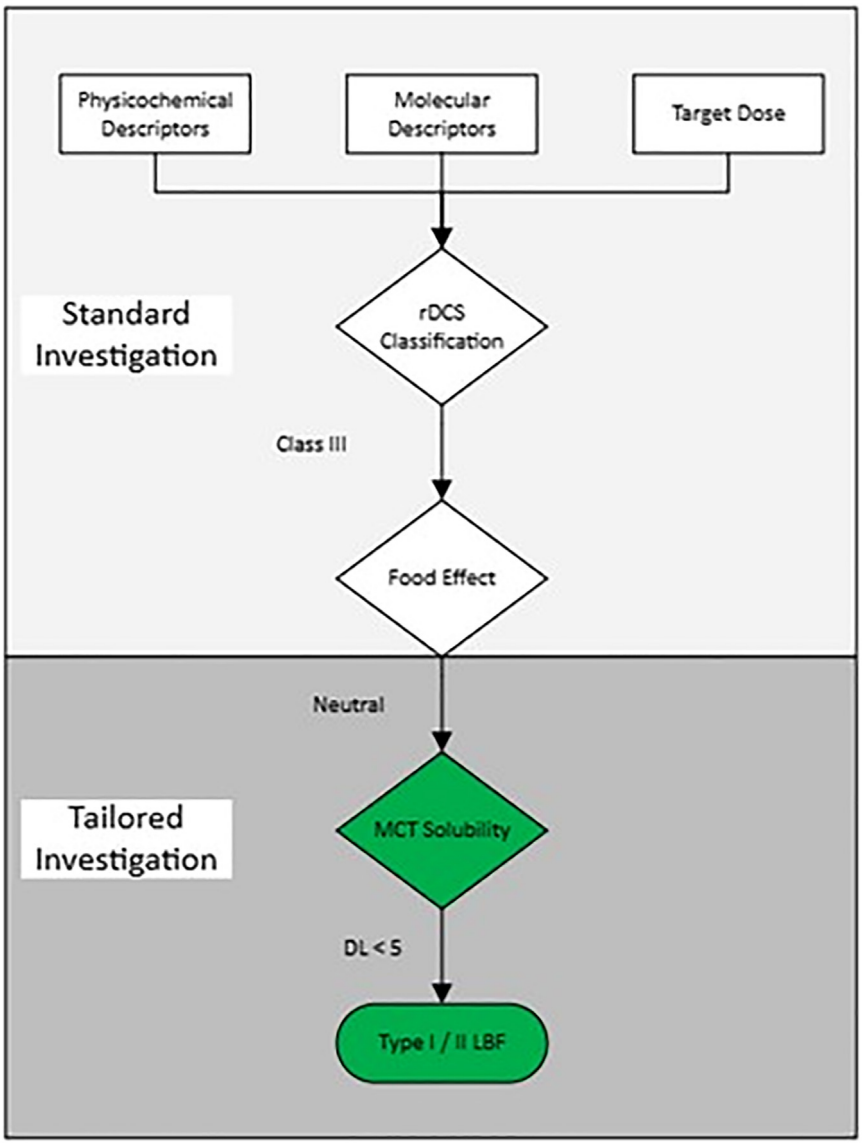

Fig. 2. Traffic light schematic illustrating that only one tailored investigation was triggered for dapagliflozin and as the dose loading in MCT was predicted to be between 0–5, dapagliflozin was proceeded to experimental validation of these results.

training set. Digitoxin, a structurally similar molecule in the training set (hydroxyl groups and benzene rings), exhibited high uncertainty due to the molecule's distinct chemical space, and this may have affected the subsequent prediction of dapagliflozin solubility.[16] This discrepancy highlights a limitation: even when a model reports high confidence, predictions can be inaccurate if the chemical space is poorly represented in the training data. Refinement could be achieved by incorporating additional data from underrepresented chemical spaces, offering an opportunity for active learning.[41] Active learning addresses out-of-distribution predictions by dynamically identifying and requesting new measurements, allowing the model to autonomously improve its understanding of the design space and suggest the next areas for evaluation.[42]

Case study 2: tailored investigation of dose loading

Mavacamten, approved by the FDA in 2022 as a "first-in-class" therapy, is indicated for the treatment of obstructive hypertrophic cardiomyopathy. With a clinical dose of 15 mg, mavacamten has a cLogP of 2.3 and a melting point of 224.5°C. Based on the initial standard investigations, mavacamten was classified as a rDCS class IIa molecule and the food effect was predicted to be neutral. Using SOAP descriptors, mavacamten was predicted to have a solubility in Type I MCT LBFs of 4.95 mg/mL (range: 1.60 – 15.32mg/mL). The predicted solubility value corresponded to a $DL_{Predicted}$ in MCTs of 3.03. Experimental validation determined the solubility in Miglyol-812 at 15± 0.63 mg/mL indicating that mavacamten is suitable for formulation as a type-I LBF in a < 1 mL of miglyol-812. As only one tailored investigation was triggered for mavacamten, it was deemed to be a low-risk drug candidate with regards to formulating it as an LBF (Fig. 3).

Case study 3: tailored investigation of biopharmaceutical dose number

Ospemifene, approved by the FDA in 2013 for the treatment of dyspareunia, has a commercial dose of 60 mg, a cLogP of 5.7, and a melting point of 135 °C. Through initial standard investigations, ospemifene was classified as a rDCS class II, and predicted to display a positive food effect, which triggered further tailored investigations to assess ospemifene's developability as an LBF. The first tailored investigation assessed solubility of ospemifene solubility in a Type I MCT LBF using SOAPS descriptors indicated a low solubility in MCTs of 10.5mg/mL, resulting in a dose number of 5.7—exceeding the threshold of 5 for Type I LBF suitability. Consequently, a second tailored investigation was initiated, to evaluate ospemifene in a Type IIIA LBF

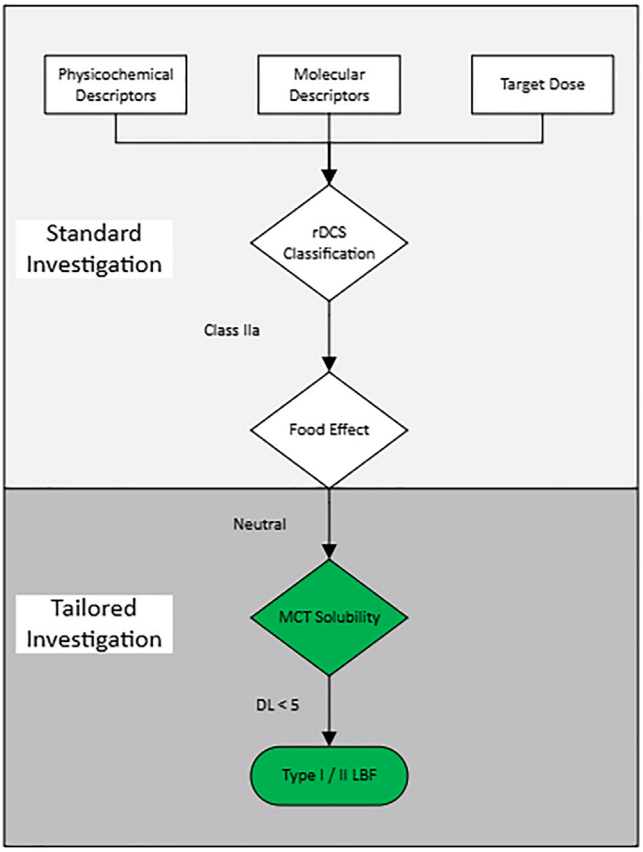

Fig. 3. Traffic light schematic indicating that only one tailored investigation was triggered for mavacamten and as the dose loading in MCT was predicted to be between 0 −5, mavacamten was proceeded to experimental validation of these results.

and predict biorelevant solubility gains during dispersion and digestion. Ospemifene was predicted to transition from rDCS class IIa drug to class I after formulation as an LBF. Based on the lipid developability framework and the risk classification table (Table 3), ospemifene is determined to have a medium risk factor when being formulated as an LBF as two tailored investigations were triggered (Fig. 4).

Experimental validation confirmed ospemifene was successfully loaded into a type-IIIA LBF at 60mg/g. Dispersion and digestion of the Type IIIA LBF significantly increased ospemifene solubility in biorelevant media compared to the unformulated drug (Table 5). This gain in solubility reflects the ability for ospemifene to be solubilised within the self-emulsified colloidal species formed during dispersion and digestion of the formulation in intestinal media. Based on the predicted solubility ratio gains, $Do_{(Predicted)}$ for ospemifene was expected to transition from a rDCS class II molecule (Do>1) to a rDCS class I (Do<1) after formulation as a type-IIIA LBF (Fig. 5 & Table 6). Experimental results confirmed that ospemifene as rDCS Class I, effectively overcoming its solubility limitations, when formulated as a Type IIIA LBF. (Table 5). This transition underscores the suitability of ospemifene for LBF development and its potential to mitigate food-effect variability.

These findings indicate that ospemifene would have been a strong candidate for LBF formulation. The formulation strategy that was adopted for the commercial product was a conventional solid oral dosage form (tablets containing mannitol, microcrystalline cellulose, povidone [E1201], pregelatinized starch, and sodium starch glycolate), which must be taken with food.[43] By applying a computationally guided development approach, this work demonstrates that ospemifene could have been steered towards development as a LBF to optimise bioavailability, offering potential advantages such as reducing food-dependent variability.

Case study 4: tailored investigation of lipophilic salt strategy

Flibanserin was licenced by the FDA in 2015, is used to treat premenopausal women with HSDD and has a commercial dose of 100 mg. Flibanserin has a cLogP of 3.82 and a melting point of 162°C. Standard investigations classified flibanserin as an rDCS class IIa and predicted the drug would display a positive food effect. The absolute oral bioavailability of flibanserin has been reported to be ∼33%, which was attributed to poor solubility and a high first pass metabolism.[44] Initial tailored investigations showed flibanserin failed to meet key thresholds (DL < 5 or $Do_{Predicted}$ < 1). Despite this, flibanserin demonstrated potential for LBF formulation due to predicted gains of up to 6-fold in $Do_{(Predicted)}$ following dispersion and digestion. However, experimental solubility remained a limiting factor: <3 mg/mL in Type I LBFs (Miglyol-812) and ∼20 mg/mL in Type IIIA formulations.

However, due to the ionisation potential of flibanserin it represents a class of a drug which would be a suitable candidate for lipophilic salt formation where increases to lipid solubility could bridge potential Do and DL limitations. The triggering of three tailored investigations, classified flibanserin as a high-risk drug candidate with respect to formulation as an LBF (Fig. 6).

Flibanserin lipophilic salts were successfully synthesised and isolated with docusic acid and dodecyl sulfate acid using a biphasic method, previously developed.[37] The favourable ΔpKa between the basic drug and acidic counterions ensured that ionisation was complete. Formation of a lipophilic salt was confirmed by the presence of methyl functional groups from the sulfate counterion at 13 ppm in the C_{13} spectra as well as the CH_3 terminal peaks at 0.87 ppm observed in ^1H spectra which were indicative of the presence of docusate and dodecyl sulfate which were not present in spectra for flibanserin free base. The results of the solubility studies demonstrated that flibanserin as a lipophilic salt display up to 4-fold increase in its lipid solubility in a type III-A formulation relative to the free-base form of flibanserin (Fig. 7). This increase in solubility can be attributed to both a reduction in the crystal packing of the drug and an increased lipophilicity of the salt complex.[32] Higher lipid solubilities were reported for docusate salts of flibanserin, indicating a more effective disruption to crystal packing through greater steric hinderance from the bulkier docusate molecule. Through this an improvement to lipid solubility of the drug an enhancement to Do can be realised with future *in vitro* and *in vivo* improvements due to the greater dose loading potential. This improved lipid solubility for flibanserin lipophilic salts allowed dose loadings of 80% of the drug solubility in the formulation. While this is below the targeted commercial dose of 100 mg, there is scope to improve via the addition of other excipients and co-solvents.

Discussion

Bio-enabling formulations such as LBFs play a vital role in allowing new drug candidates to enter the market as a commercially and bio-pharmaceutically viable dosage form. This study demonstrates that integrating computational tools within a structured LBF developability framework allows data-driven decisions regarding formulation suitability. The successful application of these tools highlights the importance of a unified, accessible, and user-friendly approach to deploying pharmaceutical computational algorithms. As modelling is associated with a considerable degree of uncertainty, these tools will act more so as a guide to formulation selection complementing rather than replacing *in vitro* experiments. However, informed computational screening can significantly reduce the need for complex and costly preliminary experiments, improving efficiency and reducing resource waste in R&D programmes. In this study, > 200 FDA-approved drugs were screened using a suite of computational tools to assess their suitability for LBFs. Among these, a subset showed

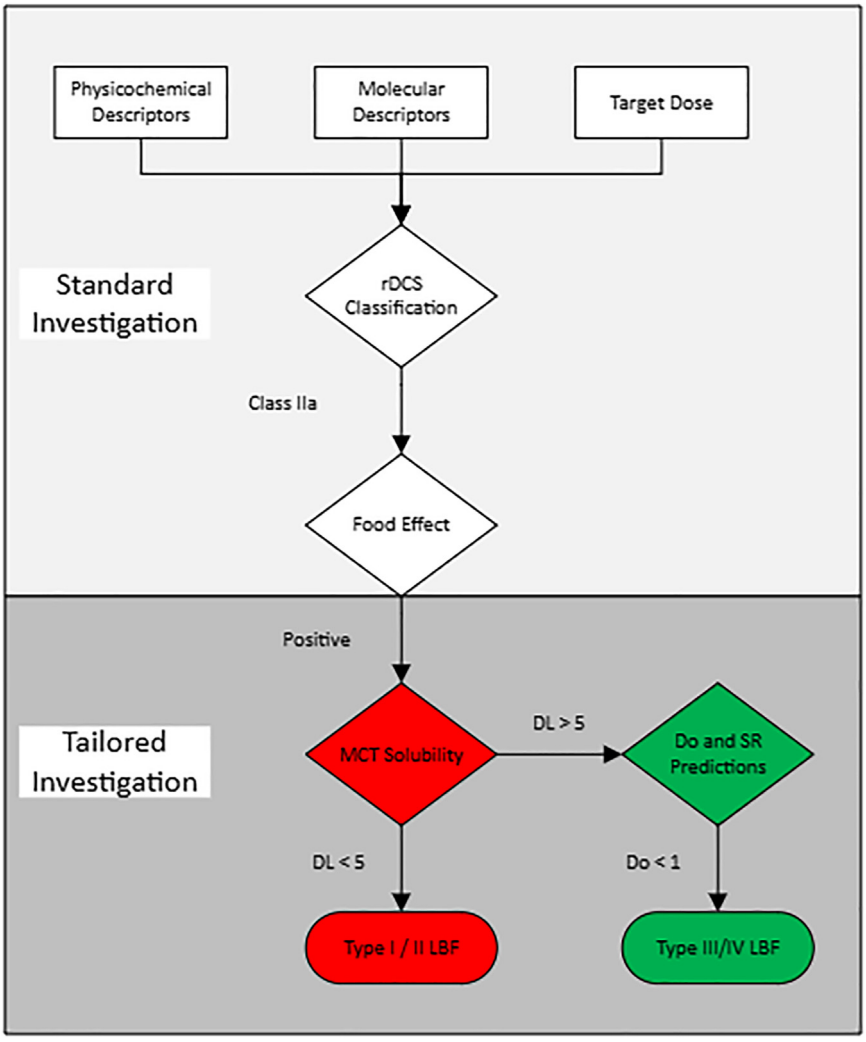

Fig. 4. Traffic light schematic indicating that two tailored investigations were triggered for ospemifene and a type-IIIA LBF was required to progress ospemifene to *in vitro* studies.

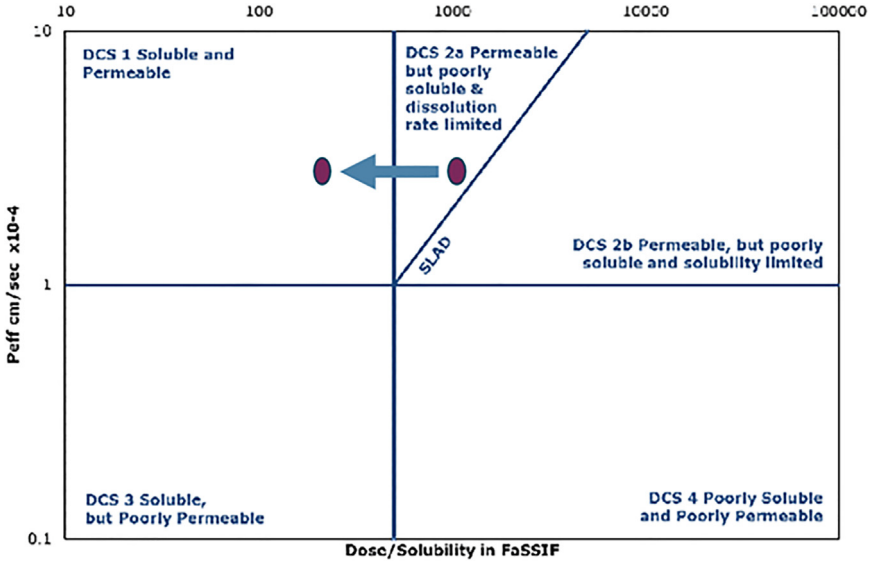

Fig. 5. rDCS infographic example depicting the transition of ospemifene from rDCS class IIa to class I after formulation as an LBF.

Table 6
Predicted biopharmaceutical dose number ($Do_{Predicted}$) and experimentally determined biopharmaceutical dose numbers ($Do_{(Experimental)}$) for formulated and unformulated ospemifene in biorelevant media.

Variable	Do	$Do_{(Experimental)}$	$Do_{(Predicted)}$	rDCS classification
FaSSIF	2.67	-	-	IIa
FaSSIF containing LC disperse	-	0.141	0.089	I
FaSSIF containing MC disperse	-	0.127	0.065	I
FaSSIF containing LC digest	-	0.32	0.083	I

Solubility values for $Do_{(Experimental)}$ were taken at the end of dispersion and digestion respectively.

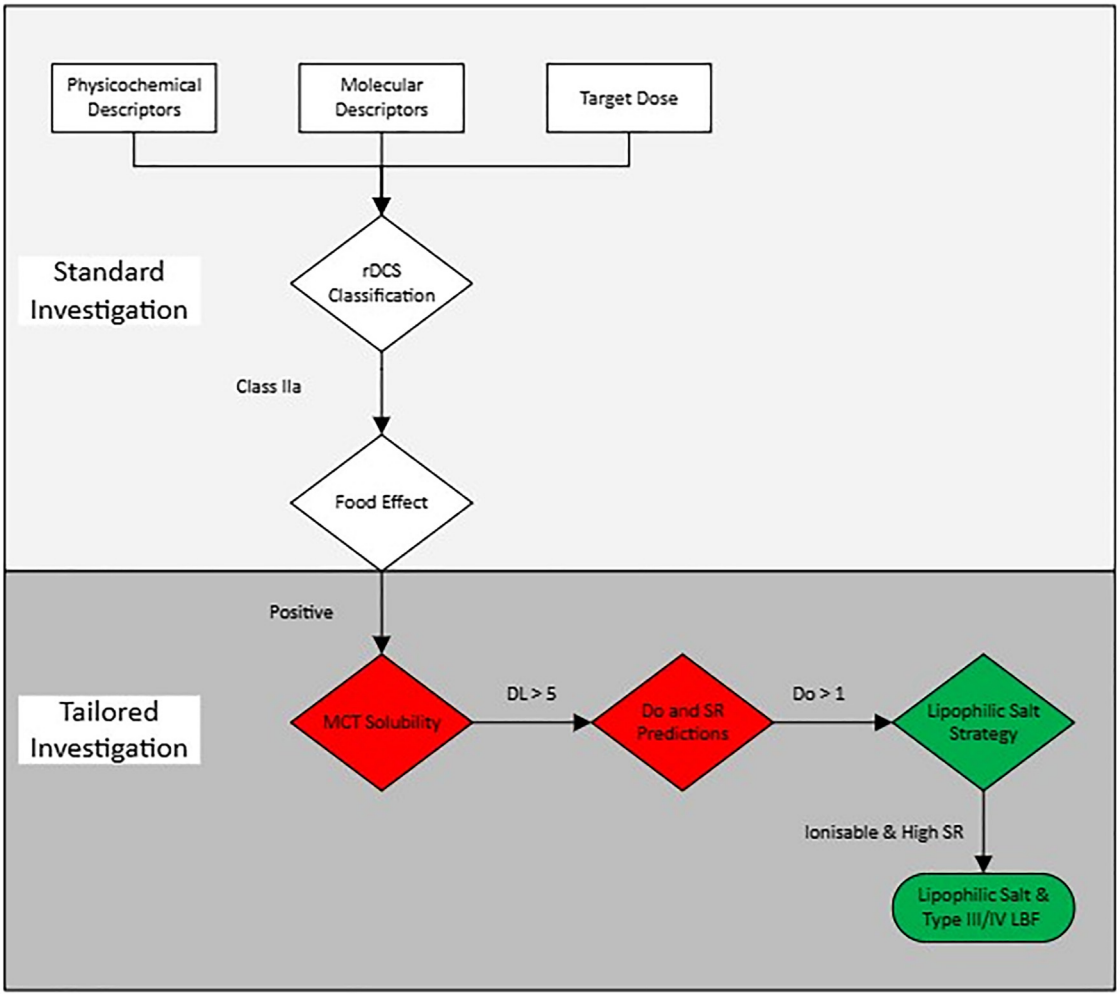

Fig. 6. Traffic light schematic indicating that three tailored investigations were required to formulate flibanserin as an LBF.

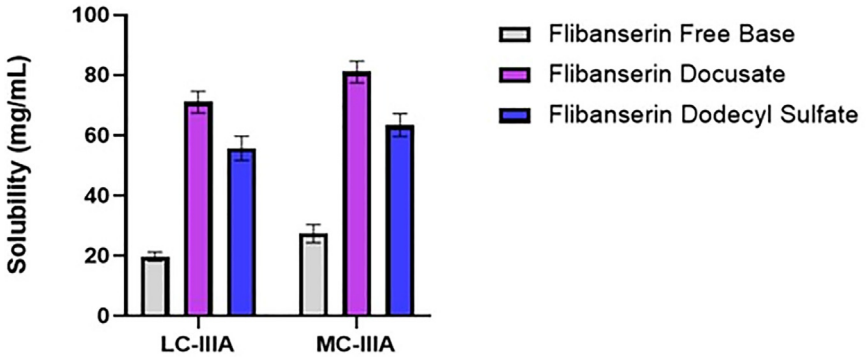

Fig. 7. Illustration of the solubility gain achieved in both medium (MC-IIIA: Miglyol-812 oil (40%), Tween-85 (40%), RH40 (20%)) and long-chain (LC-IIIA: Olive oil (40%), Tween-85 (40%), RH40 (20%)) formulations following conversion of flibanserin free base to a lipophilic salt complex.

positive or neutral food effects, a DL <5 in Type I LBFs, enhancements to Do when formulated as a Type III LBF as well as suitability for synthesis as lipophilic salts. Collectively, these parameters were incorporated into an rDCS-based framework, where targeted experimental investigations validated the predictive outputs and demonstrated drug suitability for LBFs.

The successful integration of a variety of computational tools relating to the development of LBFs and the development of a refined drug developability framework has been demonstrated. The models applied here are readily available and utilise easily accessible software and molecular descriptors. For example, the model utilised in this framework to predict food effect only requires 4 easily attainable input drug properties (Dose, HBD, S+LogP and T_PSA) to generate predictions, while with regards to the predicting of biorelevant solubility change during dispersion and digestion, the PLS models use descriptors which are readily available. Complementary multiple linear regression equations are also available, making this framework practical for widespread use and future integration with additional models.[15] The developability framework utilised seeks to move beyond basic drug-likeness filters for specific formulation types that often overlook drug: formulation potential. Instead of using single property thresholds, such as logP, or melting point, this framework leverages data-driven models that capture complex interactions among multiple drug attributes, addressing real-world challenges in drug developability. Success within this sphere relies on continued data sharing and expansion of the chemical design space, enabling broader applicability and refinement of predictive tools. However, the field of computational pharmaceutics remains a relatively new field and for the full potential to be realised various recommendations for advancing ML-informed drug formulation including reproducibility, data stewardship, interpretability and feature selection need to be considered. [6]

In cases where lipophilic salt formation is identified as a viable strategy, the stability of the resulting salt complex follows the same fundamental principles as conventional salt stability. To ensure stable salt formation through effective proton transfer, the pKa of the drug and counterion should differ by at least two units—ΔpKa. Generally, the greater the acidity or basicity of the drug or counterion, the more likely a stable lipophilic salt can be formed. As with any potential commercial product, stability studies are conducted prior to regulatory submission. Physical and chemical stability, along with capsule compatibility, should be monitored and optimized as appropriate, ideally in parallel with formulation performance testing. For lipid solutions, physical stability assessment includes visual and microscopic inspection for drug precipitation during prolonged storage under ICH-recommended conditions. Previous reports have described at least 3 months physical stability for LBFs containing lipophilic salts when stored under ICH long-term conditions (25°C/60% RH).[36,40]

While limitations of this approach include an inherent bias towards LBFs, in the absence of large-scale data bases and predictive tools for all formulation types, specific frameworks for specific formulations are sufficient for now.[3] Prior to industrial deployment larger datasets are required to ensure more accurate predictions whereas much chemical spaces as possible can be represented.[6] While the models employed here used training sets of 30+ drugs—large for pharmaceutics but small by data science standards—future refinement through active learning and expanded datasets will enhance robustness. Additionally, incorporating predictive models for other excipients and formulation types (*e.g.*, soybean oil, Captex-355) could broaden applicability.[35] Further, both approved drugs and ones under development display a wide range of physicochemical properties and may not be fully represented by molecules within the training dataset used by the models employed here. Expansion of such training sets through active learning could result in a refinement in models' accuracy across a more diverse chemical space.

A notable source of bias in regulatory data arises from the under-reporting of failed drug products. Formulations that demonstrate insufficient oral bioavailability in preclinical studies are typically excluded from early clinical evaluation, limiting opportunities to compare them with bio-enabling approaches. Similarly, drug candidates discontinued for commercial reasons are often absent from publicly accessible development datasets. Compliance with reporting requirements on ClinicalTrials.gov, the largest clinical trials registry, remains suboptimal[45] and even late-stage trial failures are frequently not published in peer-reviewed journals.[46] This lack of transparency may skew the evidence base and overestimate the success of formulation strategies. Although such bias may limit the breadth of the framework's applicability, it does not diminish its value within its intended scope, namely, providing screening support and guiding decision making for drug candidates with development relevant properties. Incorporating near-miss or failed development candidates, as well as non-approved clinical compounds, in future studies or as separate case studies could improve the generalizability of the framework and provide a more comprehensive assessment of formulation strategies. Looking ahead, there is a clear opportunity to expand and refine the predictive capabilities of the models utilised in this work by increasing the diversity and scope of training datasets through active learning. This will ensure the robustness of the framework in accommodating novel drug candidates and emerging therapeutic compounds.[6] There is further scope to incorporate additional predictive tools into the framework as advances are made in models predicting food effects, drug concentration profiles, and the extent of precipitation thereby supporting future LBF development.

The quote "essentially all models are wrong, but some are more useful than others" coined by the statistician George E.P. Box is particularly relevant in the case of this developability framework. While the results obtained from a computational model cannot be expected to entirely remove the need for experimental validation due to the constant presence of degrees of uncertainty, they still carry a significant degree of weighting. Ultimately, integrating *in silico* and *in vitro* approaches offers the most effective path forward for accelerating drug development and improving R&D productivity.

Conclusion

Overall, this work underscores the importance of early predictors and indicators of developability, demonstrating how understanding and applying drug properties can guide the selection of suitable formulation strategies—specifically LBFs. The computationally informed developability framework proposed here offers a practical, straightforward tool to improve formulation decision-making for LBF development. The computational tools integrated into this framework rely on readily available molecular descriptors, enabling rapid *in silico* screening and guiding experimental validation through targeted *in vitro* studies. This approach provides actionable insights for formulation scientists, particularly at the early developmental stage, and can significantly reduce costs and resource use—an advantage for start-ups and smaller companies with limited expertise. Additionally, minimizing material waste through computational screening aligns with sustainability goals and climate action initiatives. Given the high attrition rates in drug development, embracing artificial intelligence and machine learning is critical for promoting innovation and accelerating timelines. Structured frameworks such as the one presented here eliminate guesswork, optimize pre-formulation processes, and support data-driven decision-making, ultimately improving R&D efficiency and productivity.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This publication has received funding from the Science Foundation Ireland (SFI) through the SSPC, The SFI Research Centre for Pharmaceuticals under grant number (12/RC/2275_P2).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.xphs.2026.104262.

References

1. Bennett-Lenane H, O'Shea JP, O'Driscoll CM, Griffin BT. A retrospective biopharmaceutical analysis of >800 approved oral drug products: are drug properties of solid dispersions and lipid-based formulations distinctive? *J Pharm Sci*. 2020;109(11):3248–3261. https://doi.org/10.1016/j.xphs.2020.08.008.
2. Griffin BT, Dressman JB. The PEARRL reviews – innovative drug development strategies tailored to facilitate earlier access to new oral medicines. *J Pharmacy Pharmacol*. 2019;71(4):439–440. https://doi.org/10.1111/jphp.13083.
3. Kuentz M, Holm R, Kronseder C, Saal C, Griffin BT. Rational selection of bio-enabling oral drug formulations – a PEARRL commentary. *J Pharm Sci*. 2021;110(5):1921–1930. https://doi.org/10.1016/j.xphs.2021.02.004.
4. Kalepu S, Nekkanti V. Insoluble drug delivery strategies: Review of recent advances and business prospects. *Acta Pharm Sin B*. 2015;5(5):442–453. https://doi.org/10.1016/j.apsb.2015.07.003.
5. Bannigan P, Hickman RJ, Aspuru-Guzik A, Allen C. The dawn of a new pharmaceutical Epoch: can AI and robotics reshape drug formulation? *Adv Healthc Mater*. 2024. https://doi.org/10.1002/adhm.202401312. Published online.
6. Murray JD, Lange JJ, Bennett-Lenane H, et al. Advancing algorithmic drug product development: recommendations for machine learning approaches in drug formulation. *Eur J Pharm Sci*. 2023;191. https://doi.org/10.1016/j.ejps.2023.106562.
7. Ditzinger F, Price DJ, Ilie AR, et al. Lipophilicity and hydrophobicity considerations in bio-enabling oral formulations approaches – a PEARRL review. *J Pharmacy Pharmacol*. 2019;71(4):464–482. https://doi.org/10.1111/jphp.12984.
8. Pouton CW. Lipid formulations for oral administration of drugs: non-emulsifying, self-emulsifying and 'self-microemulsifying' drug delivery systems. *Eur J Pharm Sci*. 2000;11:S93–S98. https://doi.org/10.1016/S0928-0987(00)00167-6.
9. Wellcome. *Unlocking the potential of AI in Drug Discovery Current status, barriers and future opportunities*. 2023.. Accessed 16 December 2025; https://wellcome.org/insights/reports/unlocking-potential-ai-drug-discovery.
10. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. *Adv Drug Deliv Rev*. 2021;175:113806. https://doi.org/10.1016/j.addr.2021.05.016.
11. Bergström CAS, Charman WN, Porter CJH. Computational prediction of formulation strategies for beyond-rule-of-5 compounds. *Adv Drug Deliv Rev*. 2016;101:6–21. https://doi.org/10.1016/j.addr.2016.02.005.
12. Selen A, Dickinson PA, Müllertz A, et al. The biopharmaceutics risk assessment roadmap for optimizing clinical drug product performance. *J Pharm Sci*. 2014;103(11):3377–3397. https://doi.org/10.1002/jps.24162.
13. Bennett-Lenane H, Griffin BT, O'Shea JP. Machine learning methods for prediction of food effects on bioavailability: a comparison of support vector machines and artificial neural networks. *Eur J Pharm Sci*. 2022:168. https://doi.org/10.1016/j.ejps.2021.106018.
14. Bennett-Lenane H, Koehl NJ, O'Dwyer PJ, Box KJ, O'Shea JP, Griffin BT. Applying computational predictions of biorelevant solubility ratio upon self-emulsifying lipid-based formulations dispersion to predict dose number. *J Pharm Sci*. 2021;110(1):164–175. https://doi.org/10.1016/j.xphs.2020.10.055.
15. Ejskjær L, Holm R, Kuentz M, Box KJ, Griffin BT, O'Dwyer PJ. Predictions of biorelevant solubility change during dispersion and digestion of lipid-based formulations. *Eur J Pharm Sci*. 2024;200. https://doi.org/10.1016/j.ejps.2024.106833.
16. Lange JJ, Anelli A, Alsenz J, et al. Comparative analysis of chemical descriptors by machine learning reveals atomistic insights into solute-lipid interactions. *Mol Pharm*. 2024;21(7):3343–3355. https://doi.org/10.1021/acs.molpharmaceut.4c00080.
17. Zane P, Gieschen H, Kersten E, et al. In vivo models and decision trees for formulation development in early drug development: a review of current practices and recommendations for biopharmaceutical development. *Eur J Pharm Biopharm*. 2019;142:222–231. https://doi.org/10.1016/j.ejpb.2019.06.010.
18. Van Den Bergh A, Van Hemelryck S, Bevernage J, et al. Preclinical bioavailability strategy for decisions on clinical drug formulation development: an in depth analysis. *Mol Pharm*. 2018;15(7):2633–2645. https://doi.org/10.1021/acs.molpharmaceut.8b00172.
19. Kuentz M, Holm R, Elder DP. Methodology of oral formulation selection in the pharmaceutical industry. *Eur J Pharm Sci*. 2016;87:136–163. https://doi.org/10.1016/j.ejps.2015.12.008.
20. Holm R. Bridging the gaps between academic research and industrial product developments of lipid-based formulations. *Adv Drug Deliv Rev*. 2019;142:118–127. https://doi.org/10.1016/j.addr.2019.01.009.
21. Lou H, Lian B, Hageman MJ. Applications of machine learning in solid oral dosage form development. *J Pharm Sci*. 2021;110(9):3150–3165. https://doi.org/10.1016/j.xphs.2021.04.013.
22. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. *Adv Drug Deliv Rev*. 2021;175:113806. https://doi.org/10.1016/j.addr.2021.05.016.
23. Rosenberger J, Butler J, Dressman J. A refined developability classification system. *J Pharm Sci*. 2018;107(8):2020–2032. https://doi.org/10.1016/j.xphs.2018.03.030.
24. Butler JM, Dressman JB. The developability classification system: application of biopharmaceutics concepts to formulation development. *J Pharm Sci*. 2010;99(12):4940–4954. https://doi.org/10.1002/jps.22217.
25. Rosenberger J, Butler J, Muenster U, Dressman J. Application of a refined developability classification system. *J Pharm Sci*. 2019;108(3):1090–1100. https://doi.org/10.1016/j.xphs.2018.10.044.
26. O'Shea JP, Holm R, O'Driscoll CM, Griffin BT. Food for thought: formulating away the food effect – a PEARRL review. *J Pharmacy Pharmacol*. 2019;71(4):510–535. https://doi.org/10.1111/jphp.12957.
27. Tu JV. *Advantages and Disadvantages of using Artificial Neural Networks Versus Logistic Regression for Predicting Medical Outcomes*. 1996. Vol 49.
28. Bourquin J, Schmidli H, Van Hoogevest P, Leuenberger H. Basic concepts of Artificial Neural Networks (ANN) modeling in the application to pharmaceutical development. *Pharm Dev Technol*. 1997;2(2):95–109. https://doi.org/10.3109/10837459709022615.
29. Müllertz A, Ogbonna A, Ren S, Rades T. New perspectives on lipid and surfactant based drug delivery systems for oral delivery of poorly soluble drugs. *J Pharmacy Pharmacol*. 2010;62(11):1622–1636. https://doi.org/10.1111/j.2042-7158.2010.01107.x.
30. Savla R, Browne J, Plassat V, Wasan KM, Wasan EK. Review and analysis of FDA approved drugs using lipid-based formulations. *Drug Dev Ind Pharm*. 2017;43(11):1743–1758. https://doi.org/10.1080/03639045.2017.1342654.
31. Bergström CAS, Larsson P. Computational prediction of drug solubility in water-based systems: qualitative and quantitative approaches used in the current drug discovery and development setting. *Int J Pharm*. 2018;540(1-2):185–193. https://doi.org/10.1016/j.ijpharm.2018.01.044.
32. Williams HD, Ford L, Igonin A, et al. Unlocking the full potential of lipid-based formulations using lipophilic salt/ionic liquid forms. *Adv Drug Deliv Rev*. 2019;142:75–90. https://doi.org/10.1016/j.addr.2019.05.008.
33. Holm R, Kuentz M, Ilie-Spiridon AR, Griffin BT. Lipid based formulations as supersaturating oral delivery systems: from current to future industrial applications. *Eur J Pharm Sci*. 2023;189. https://doi.org/10.1016/j.ejps.2023.106556.
34. Alskär LC, Porter CJH, Bergström CAS. Tools for early prediction of drug loading in lipid-based formulations. *Mol Pharm*. 2016;13(1):251–261. https://doi.org/10.1021/acs.molpharmaceut.5b00704.
35. Persson LC, Porter CJH, Charman WN, Bergström CAS. Computational prediction of drug solubility in lipid based formulation excipients. *Pharm Res*. 2013;30(12):3225–3237. https://doi.org/10.1007/s11095-013-1083-7.
36. Williams HD, Ford L, Lim S, et al. Transformation of biopharmaceutical classification system class I and III drugs into ionic liquids and lipophilic salts for enhanced developability using lipid formulations. *J Pharm Sci*. 2018;107(1):203–216. https://doi.org/10.1016/j.xphs.2017.05.019.
37. Ryan CD, Griffin BT, O'Shea JP. Rationalizing counterion selection for the development of lipophilic salts: a case study with Venetoclax. *Mol Pharm*. 2024;21(6):2981–2992. https://doi.org/10.1021/acs.molpharmaceut.4c00106.
38. Ford L, Tay E, Nguyen TH, et al. API ionic liquids: probing the effect of counterion structure on physical form and lipid solubility. *RSC Adv*. 2020;10(22):12788–12799. https://doi.org/10.1039/d0ra00386g.
39. Tay E, Nguyen TH, Ford L, et al. Ionic liquid forms of the antimalarial lumefantrine in combination with LFCS type IIIB lipid-based formulations preferentially increase lipid solubility, in vitro solubilization behavior and in vivo exposure. *Pharmaceutics*. 2020;12(1). https://doi.org/10.3390/pharmaceutics12010017.
40. Williams HD, Ford L, Han S, et al. Enhancing the oral absorption of kinase inhibitors using lipophilic salts and lipid-based formulations. *Mol Pharm*. 2018;15(12):5678–5696. https://doi.org/10.1021/acs.molpharmaceut.8b00858.
41. Musil F, Willatt MJ, Langovoy MA, Ceriotti M. Fast and accurate uncertainty estimation in chemical machine learning. *J Chem Theory Comput*. 2019;15(2):906–915. https://doi.org/10.1021/acs.jctc.8b00959.
42. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Digit Med*. 2022;5(1). https://doi.org/10.1038/s41746-022-00614-9.
43. FDA and CDER. *Opshena*. 2013.. Accessed 16 December 2025; https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/203505s015lbl.pdf.
44. Thangaraju P, Velmurugan H, SS TY. Drug flibanserin—in hypoactive sexual desire disorder. *Gynecol Obstetr Clin Med*. 2022;2(2):91–95. https://doi.org/10.1016/j.gocm.2022.04.003.
45. DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *Lancet*. 2020;395(10221):361–369. https://doi.org/10.1016/S0140-6736(19)33220-9.
46. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMa Intern Med*. 2016;176(12):1826. https://doi.org/10.1001/jamainternmed.2016.6008.