CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Computational pharmaceutics approaches to inform drug developability: focus on lipid-based formulations
Authors	Bennett-Lenane, Harriet
Publication date	2021-10-04
Original Citation	Bennett-Lenane, H. 2021. Computational pharmaceutics approaches to inform drug developability: focus on lipid-based formulations. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2021, Harriet Bennett-Lenane. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-02 17:52:54
Item downloaded from	https://hdl.handle.net/10468/13238



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann

National University of Ireland

Coláiste na hOllscoile Corcaigh

University College Cork

School of Pharmacy

Computational Pharmaceutics Approaches to Inform Drug Developability: Focus on Lipid-Based Formulations

Thesis Presented by

Harriet Bennett-Lenane, *B.Sc. (Pharm.), M.Pharm., M.P.S.I.*

In fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of

Prof. Brendan T. Griffin, *B.Sc. (Pharm.), Ph.D., M.P.S.I.*
and
Dr. Joseph P. O'Shea, *M.Pharm., Ph.D., M.P.S.I.*
and
Prof. Caitriona M. O'Driscoll, *B.Sc. (Pharm.), M.A., Ph.D., M.P.S.I.*

Head of School:
Prof. Brendan T. Griffin

October, 2021

Even when we find not what we seek, we find something as well worth seeking as what we missed.

Robert Boyle, Lismore Castle,

The Works of the Honourable Robert Boyle, 1772.

Table of Contents

Declaration **11**

Research Funding **11**

Author Publications **14**

Conference Presentations **16**

Concomitant Peer Reviewed Publications **17**

Chapter 1: Introduction: Challenges in Drug Development and Exploring Pathways for Improved Drug Developability **20**

Challenges in Drug Development Stemming from Modern Drug Discovery Approaches 21

Bio-Enabling Formulations 22

Lipid-Based Formulations 23

Solid Dispersions 25

Evolving Pre-Clinical Formulation Design and Selection 27

Drug Classification Systems: Moving from the Drug Substance to Understanding Developability 29

Can Structured Formulation Development Be Achieved Through Property Based Rules? 32

Formulation Decision Trees 36

Can a Retrospective Biopharmaceutical Analysis of Drug Properties Inform Future Success? 41

Methods 43

Dataset Selection 43

Compilation of Physicochemical Descriptors 44

Statistical Analysis 46

Results 47
Commercial Success to Date 47
Commercial Lipid-Based Formulations 47
Commercial Solid Dispersions 49
Commercial Products via Both Formulation Technologies 51
BDDCS Classifications 52
Retrospective Statistical Analysis of Properties of Commercialised LBF and SD Drug Compounds. 53
Property Trends Resulting from a Retrospective Statistical Analysis of Licensed LBF and SD Products. 57
Molecular Weight (MW) 57
Melting Point (T_m) 58
Lipophilicity (logP, clogP, $logD_{7.4}$) 60
Aqueous Solubility (logS) 61
Percentage Excreted Unchanged in Urine (%U) 62
Rotatable Bond Count (RB) 62
Hydrogen Bond Acceptors (HBA) 63
Hydrogen Bond Donors (HBD) 64
Polar Surface Area (PSA) 65
Lipinski Rule-of-5 Violations (Ro5) 65
Dosage Strength (pDose and MDS) 66
Non-Significant Properties 67
Properties of Drugs Commercialised via Both Bio-Enabling Formulation Technologies 68
Summary of Findings 69

Chapter 2: Introduction: Computational Tools to Inform Drug Developability 71

The Journey from “Classic” to “Modern” Expert Learning 72

Computational Modelling Opportunities Across the Pharmaceutical Industry..74
Big Opportunities from Big Data 76
Theory- versus Data-Driven Modelling 77
The Emerging Field of Computational Pharmaceutics 81
The Nuts and Bolts of Machine Learning 86
Machine Learning Algorithms..... 88
Multiple Linear Regression (MLR) 88
Principal Component Analysis (PCA) 89
Partial Least Squares Regression (PLS) and Partial Least Squares Discriminant Analysis (PLS-DA) 90
Support Vector Machines (SVM) 91
Artificial Neural Networks (ANN) 92
Steps in Machine Learning Model Development 95
Current Challenges to Wider Integration of Computational Pharmaceutics Tools 98

Thesis Objectives..... 100

Chapter 3: Machine Learning Methods for Prediction of Food Effects on Bioavailability: A Comparison of Support Vector Machines and Artificial Neural Networks..... 101
Abstract 102
Introduction 103
Methods 107
Database Collation..... 107
Food Effect Classification 108
Compilation of Physicochemical Descriptors 108
Statistical Analysis 109
BCS Classification..... 110
Machine Learning/Model Development..... 110

Support Vector Machine Classification 112
Artificial Neural Networks 113
Results 114
Analysis of Trends in Physicochemical Descriptors between Food Effect Classifications 114
Food Effect Prediction using BCS Classification 117
Applying ANN and SVM to Predict Food Effect Classification 118
Discussion 123
Conclusion 129

Chapter 4: Applying Computational Predictions of Biorelevant Solubility Ratio Upon Self-Emulsifying Lipid-Based Formulations Dispersion to Predict Dose Number 130
Abstract 131
Introduction 132
Methods 136
Dataset Selection 136
Formulations 138
Media Preparation 138
Media Characterisation: Media Droplet Size and Zeta Potential 139
Experimental Solubility Determination 139
Shake Flask Method 140
µDISS Profiler 140
Drug Physicochemical and Molecular Properties 141
Biopharmaceutical Data Analysis 141
Multivariate Data Analysis and Modelling Parameters 142
Solubility Equation for Predicting Biopharmaceutical Dose Number and DCS Class. 144
Predicting DCS Classifications of Commercial LBF Drugs 145

Results 146
SEDDS Characterisation on Biorelevant Dispersion 146
Solubility in Biorelevant SEDDS Dispersions – Comparison of SEDDS *Migylol812* and SEDDS *OliveOil* 147
Solubility Ratio Trends 148
Computational Prediction of Biorelevant Solubility Gain with SEDDS. 152
Enhanced Biorelevant Solubility Ratio Equation 154
Use of Predicted Solubility Ratios to Predict Drug DCS Class with SEDDS. . 154
Predicted DCS Classifications of Commercial LBF Drugs. 158
Discussion 160
Conclusion 168

Chapter 5: Application of Artificial Neural Networks to Predict the Apparent Degree of Supersaturation in Supersaturated Lipid-Based Formulations: A Pilot Study 169
Abstract 170
Introduction 171
Materials and Methods 174
Chemical and Materials 174
Formulations 174
Dataset Selection/Drug Physicochemical and Molecular Properties 175
Equilibrium Solubility Determination 175
Apparent Degree of Supersaturation (aDS) 176
Differential Scanning Calorimetry 177
Statistical Analysis 180
Partial Least Squares Regression (PLS) 180
Artificial Neural Networks (ANN) 181
Results 182
Comparing the Solubility of MC and LC-based LBF and sLBF 182

Apparent Degree of Supersaturation 183
Quantitatively Predicting aDS using PLS and ANN 184
Discussion 188
Conclusion 192

Chapter 6: Exploring Porcine Gastric and Intestinal Fluids using Microscopic and Solubility Estimates: Impact of Placebo Self-Emulsifying Drug Delivery System Administration to Inform Bio-Predictive in vitro Tools. 193
Abstract 194
Introduction 195
Materials and Methods 199
Materials 199
Gastric and Intestinal Fluid Collection 199
Cryogenic and Negative Stain Transmission Electron Microscopy Studies 201
Solubility Studies 202
RP-HPLC/UV Analysis 203
Statistical Analysis 204
In Silico Prediction 204
Results 205
pH Characterisation of Gastric and Intestinal Porcine Media with a SEDDS 205
Microscopic Evaluation of Fasted and Fed State Gastric and Intestinal Porcine Fluids 207
Microscopic Evaluation of Porcine Gastric and Intestinal Fluids Post Administration of a Placebo SEDDS 209
Fenofibrate Solubility in Pig Gastric and Intestinal Media Post Ingestion of a Placebo SEDDS 212
Exploration of *in vitro* SEDDS Screening Tool using Enhanced Biorelevant Media and Investigation of Appropriate Dilution Conditions. 214
Discussion 216

Conclusion 224

General Discussion 225

Overview and Summary 225

Model Performance and Lessons Learned from Computational Modelling 229

Model Performance Comparisons 229

Interpretability of the Modelling Results 233

Importance of Biorelevant Experimental Inputs for Informing and Confirming Computational and *In Vitro* Tools – Lessons Learned from the Pig Model. 235

Interplay Between Computational Modelling and Conventional Biopharmaceutics Related Classifications Systems 240

Model Integration and Computational Accessibility 242

Computationally Informed and Experimentally Confirmed Developability Testing to Reconsider the Drug Substance to Drug Product Paradigm 246

Overall Conclusion and Future Perspectives 257

Appendix 1: Chapter 1 259

Appendix 2: Chapter 3 273

Appendix 3: Chapter 4 282

Appendix 4: Chapter 5 286

Abbreviation List 296

References 300

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.

Author Contribution

All work was performed independently by the author, with the following exceptions:

Chapter 5

Lipid solubility studies for four drugs were competed by Marc Jouston in Janssen Research and Development, Beerse, Belgium.

Chapter 6

Cryo-TEM and Negative Stain TEM imaging of the porcine luminal fluids was conducted by Dr. Jacob R. Jørgensen in the Core Facility for Integrated Microscopy, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark. Prof. Brendan Griffin and Dr. Niklas Koehl collected the porcine luminal fluids.

Signed:  Date: 04/10/2021

Research Funding

This research was conducted with support from the Irish Research Council Post Graduate Scholarship Project number GOIPG/2018/883.

Acknowledgements

Of all the sections of this thesis it is perhaps surprising that one of the more challenging aspects has been writing the acknowledgements. You see, it is such a profound responsibility, in my opinion, to appropriately acknowledge and convey the gratitude I have for every person who made this thesis possible and contributed to my wonderful experience in the School of Pharmacy in UCC.

Firstly, I would like to thank my supervisors Brendan, Joey and Caitriona. Before embarking on this PhD, I could never have imagined how instrumental their support and guidance would be. By not only helping to produce this piece of work, but also in creating a working environment most PhD students could only wish for. This has been a fantastic experience from start to finish, which has only been enriched through their openness, co-operation, and hands-on approach. While Covid-19 meant that, for months on end, our interactions were purely virtual, I never felt less supported. That being said, I won't particularly miss having to listen back to my own speaking voice on a Microsoft Teams meeting recording! Outside academics, I will sincerely miss all the pre- and post-match analyses and I hope they think of me if ever pigs fly and Waterford win an All-Ireland.

To all the staff in the School of Pharmacy who have been so helpful and inviting towards me during the last 3 years. Special thanks to Aisha, Aine, Sonja, Abina, Kathleen, Michael and in particular Ken, who was always available for a friendly chat amid another HPLC mishap. Special mention must also be given to all those who helped carry out the work in this thesis. Throughout these three years I have been lucky enough to collaborate with numerous groups and researchers, far and wide, for whom I have the upmost admiration. Special thanks to Niklas and Laura for their wealth of

knowledge and support throughout my first year, as well as Patrick for his guidance, particularly during my enjoyably trips to Pion Inc. I will continue to work smarter not harder.

To the pharmaceutics lab group, Brian, Jacinta, Ayse, Khaled, Olivia, Ana Luiza and a list of people too long to mention, who cannot be forgotten. Thanks for all the lunches and chats, whether they be about research or not, and for your willingness to help no matter what my query. While many of the social aspects of this PhD experience were unfortunately cut short from March 2020 onwards, I will always hold fond memories of our time in the write-up area.

To Eoghan, thanks for sensing when it was time to bring out the hidden Percy Pig packets and keeping a constant stash of Pepsi Max available during this writing process. While 2020, in particular, didn't turn out to be all that we expected, I can't wait to dust off the travel journal again and experience the world and many more adventures that life has to offer with you. You'll always be in P1.

Finally, to Mam and Dad. Thank you for always being there and for stopping at nothing to provide me with the greatest opportunities that anyone person could want in life. As my formal journey with education comes to an end (for now!), it is somewhat fitting that it ends next door to the building you brought me home from 27 years ago. It's been a long road since my tears before school on a Friday in case I got one of my spellings wrong or the agony of repeating times tables for hours on end, but it is down to your support that I am where I am today.

Thank you all.

Author Publications

i. **Chapter 1:** Harriet Bennett-Lenane, Joseph P. O'Shea, Caitriona M. O'Driscoll, Brendan T. Griffin.

Review Article titled: "A Retrospective Biopharmaceutical Analysis of >800 Approved Oral Drug Products: Are Drug Properties of Solid Dispersions and Lipid-Based Formulations Distinctive?", published August 18th 2020 in The Journal of Pharmaceutical Sciences, Volume 109, Pages 3248-3261.

DOI: https://doi.org/10.1016/j.xphs.2020.08.008.

ii. **Chapter 3:** Harriet Bennett-Lenane, Brendan T. Griffin, Joseph P. O'Shea.

Research Article titled: "Machine Learning Methods for Prediction of Food Effects on Bioavailability: A Comparison of Support Vector Machines and Artificial Neural Networks ", published September 24th 2021 in the European Journal of Pharmaceutical Sciences, Volume 167, 106018.

DOI: https://doi.org/10.1016/j.ejps.2021.106018

iii. **Chapter 4:** Harriet Bennett-Lenane, Niklas J. Koehl, Patrick J. O'Dwyer, Karl J. Box, Joseph P. O'Shea, Brendan T. Griffin.

Research Article Entitled: "Applying Computational Predictions of Biorelevant Solubility Ratio upon Self-Emulsifying Lipid-Based Formulations Dispersion to Predict Dose Number.", published 31st October 2020 in The Journal of Pharmaceutical Sciences, Volume 110, Pages, 164-175.

DOI: https://doi.org/10.1016/j.xphs.2020.10.055.

iv. **Chapter 5:** Harriet Bennett-Lenane, Joseph P. O'Shea, Jack D. Murray, Alexandra-Roxana Ilie, René Holm, Martin Kuentz, Brendan T. Griffin.

Research Article Entitled: “Application of Artificial Neural Networks to Predict the Apparent Degree of Supersaturation in Supersaturated Lipid-Based Formulations: A Pilot Study.”, published 5th September 2021 in Pharmaceutics MDPI, Volume 13, Pages 1398.

DOI: https://doi.org/10.3390/pharmaceutics13091398

v. **Chapter 6:** Harriet Bennett-Lenane, Jacob R. Jørgensen, Niklas J. Koehl, Laura J. Henze, Joseph P. O’Shea, Anette Müllertz, Brendan T. Griffin. Research Article Entitled: “Exploring Porcine Gastric and Intestinal Fluids using Microscopic and Solubility Estimates: Impact of Placebo Self-Emulsifying Drug Delivery System Administration to Inform Bio-Predictive *in vitro* Tools.“, published 1st June 2021 in the European Journal of Pharmaceutical Sciences, Volume 161, 105778.

DOI: https://doi.org/10.1016/j.ejps.2021.105778.

Conference Presentations

i. Harriet Bennett-Lenane, NiklasJ Koehl, Caitriona M. O'Driscoll, Brendan T. Griffin.

Bridging the Food Gap: Can Lipid Based Formulations Overcome Food Dependant Oral Bioavailability?"

Poster presented at the 41st All Ireland Schools of Pharmacy Conference, Trinity College Dublin, Dublin, 17-18th April 2019 and at the School of Pharmacy Research Day, University College Cork, Cork, 23rd May 2019.

ii. Harriet Bennett-Lenane, Niklas J. Koehl, Patrick J. O'Dwyer, Karl J. Box, Joseph P. O'Shea, Brendan T. Griffin

"Applying Computational Prediction of Biorelevant Solubility in Dispersions of Self-Emulsifying Lipid Based-Formulations to Predict Dose Number"

Poster presented at AAPS PharmSci360, Virtual Event, 26th October – 5th of November 2020.

iii. Harriet Bennett-Lenane, Brendan T. Griffin, Joseph P. O'Shea.

"Applying Machine Learning to Predict Food Effects of Newly Licensed Oral Medicines".

Poster presented and oral poster blitz presentation at Targeting Therapeutics for Brain Disorders Virtual Conference, University College Cork, Cork, 2nd June 2021.

Concomitant Peer Reviewed Publications

i. Patrick J O' Dwyer, Karl J Box, Niklas J Koehl, Harriet Bennett-Lenane, Christos Reppas, Rene Holm, Martin Kuentz, Brendan T Griffin.

Research Article Entitled: "Novel Biphasic Lipolysis Method to Predict *in Vivo* Performance of Lipid-Based Formulations", published 6th August 2020 in Molecular Pharmaceutics 2020, Volume 17, Pages 3342-3352.

DOI: https://doi.org/10.1021/acs.molpharmaceut.0c00427.

ii. Laura J. Henze, Niklas J. Koehl, Harriet Bennett-Lenane, René Holm, Michael Grimm, Felix Schneider, Werner Weitschies, Mirko Koziolek, Brendan T. Griffin.

Research Article Entitled: "Characterization of gastrointestinal transit and luminal conditions in pigs using a telemetric motility capsule", published 1st January 2021 in the European Journal of Pharmaceutical Sciences, Volume 156, 105627.

DOI: https://doi.org/10.1016/j.ejps.2020.105627.

iii. Niklas J Koehl, Laura J Henze, Harriet Bennett-Lenane, Waleed Faisal, Daniel J Price, René Holm, Martin Kuentz, Brendan T Griffin.

Research Article Entitled: "In Silico, In Vitro, and In Vivo Evaluation of Precipitation Inhibitors in Supersaturated Lipid-Based Formulations of Venetoclax", published 23rd April 2021 in Molecular Pharmaceutics 2021, Volume 18, Pages 2174-2188.

DOI: https://doi.org/10.1021/acs.molpharmaceut.0c00645.

Abstract

Purpose Declining productivity in the face of increasing numbers of poorly water-soluble drugs has fast-tracked necessity for predictive tools which assess the delivery potential of bio-enabling formulations. However, there is a perceived risk associated with early-stage selection of bio-enabling formulations. Computational pharmaceutics is a growing area of research interest to support structured guidance in formulation strategy. Using data-driven modelling, a streamlined roadmap of computational possibilities for development scientists is possible. Accordingly, the aim of this thesis was to examine the application of machine learning (ML) computational modelling to inform candidate developability. Via prediction of both quality target product profile characteristics and formulation performance indicators for lipid-based formulations (LBF). In recognition of the fact that computational models will not entirely circumvent need for manual screening, a further aim was to explore if analysis of landrace pig gastrointestinal fluids could facilitate increased bio-predictive performance of *in vitro* tools for LBFs.

Methods Data-driven computational models using various ML algorithms, with both classification and regression outputs, were developed to predict food effect on bioavailability, solubility ratio (SR) upon self-emulsifying drug delivery system (SEDDS) dispersion and apparent degree of supersaturation (aDS) ratio in supersaturated LBFs (sLBF). Model performance was validated using test sets or comparisons to *ex vivo* results. Gastrointestinal fluids from the landrace pig were collected in the fasted state, fed state and post placebo SEDDS administration. E*x vivo* solubility analysis and microscopic imaging were completed using these fluids, where *in vitro* biorelevant dispersion screening with SEDDS using various dilution conditions was compared to the *ex vivo* results.

Results Firstly, this thesis demonstrated the applicability of ML for the prediction of a quality target product profile characteristic of interest in early development, namely food effect on bioavailability. Secondly, this thesis has advanced computational pharmaceutics to inform drug developability. Computational predictions of solubility gain upon SEDDS dispersion informed a biopharmaceutical dose number in intestinal fluids, which can be incorporated within the developability classification system (DCS) framework to inform drug developability. Thirdly, in recognition of the use of supersaturated LBFs to overcome dose loading limitations, this thesis has demonstrated how ML algorithms can predict the maximum dose loading upon thermal induced supersaturation. Moreover, increased understanding of the fate of SEDDS upon oral administration furthered the utility of the pig pre-clinical model, validated accuracy of *in silico* predictions and aided development of a bio-predictive *in vitro* screening tool for LBFs. Ultimately, it was demonstrated that the computational and *in vitro* tools developed in this thesis can be embedded within a wider refined drug substance to drug product development framework.

Conclusion This thesis highlighted how pharmaceutics datasets are amenable to ML. The ability of computational pharmaceutics to facilitate structured formulation decisions was demonstrated. As model development aided increased understanding of the investigated phenomena through their relationship to drug properties, this thesis identified the significant potential to be gained from early analysis of drug properties. Additionally, utility of the landrace pig model to inform increasingly bio-predictive *in vitro* screening tools was established. The proposed refining of the drug substance to drug product development framework demonstrated the significance of both computationally informed and experimentally confirmed aspects of drug developability decision-making.

Chapter 1: Introduction: Challenges in Drug Development and Exploring Pathways for Improved Drug Developability

This chapter contains material adapted from the following publication:

A Retrospective Biopharmaceutical Analysis of >800 Approved Oral Drug Products: Are Drug Properties of Solid Dispersions and Lipid-Based Formulations Distinctive?

Harriet Bennett-Lenane, Joseph P. O'Shea, Caitriona M. O'Driscoll, Brendan T. Griffin. Published August 18th, 2020 in The Journal of Pharmaceutical Sciences, Volume 109, Pages 3248-3261. DOI: https://doi.org/10.1016/j.xphs.2020.08.008.

Challenges in Drug Development Stemming from Modern Drug Discovery Approaches

Modern drug discovery approaches, including combinatorial chemistry, molecular modelling and high throughput screening have revolutionised drug candidate discovery in the pharmaceutical industry (1-3). However, modern drug candidates often display high lipophilicity, poor aqueous solubility and resultant reduced oral bioavailability (4, 5), as these properties are common negative penalties traded for high potency and selectivity for contemporary lipophilic binding pockets or drug targets (2, 6). The resulting poor aqueous solubility of drug candidates means that when the solubility of the drug in the gastrointestinal tract (GIT) is limited, the likelihood of solubility or the dissolution rate limiting oral bioavailability is high. As a result, the ultimate clinical usefulness of these compounds is limited as oral drugs generally need to be dissolved in order to permeate the intestinal membranes (7, 8). Accordingly, limited aqueous solubility of candidates represents a major challenge to successful transition from drug substance to drug product. Such a "poor solubility challenge" plays a significant role in both rising product attrition rates and rising financial investment in research and development (R&D). These challenges are exemplified by the fact that the number of new medicines approved for clinical use decreasing consistently from the 1950s-2010s (9), as it is estimated that anywhere from 40-90% of all compounds in development display limited aqueous solubility (10, 11).

As high attrition of clinical candidates continues to impact significantly on pharmaceutical R&D, drug "developability" is becoming an increasingly significant focus for formulation scientists. Drug developability encompasses a broad term involving evaluation of a potential therapeutic candidates' "drug-like" properties,

including both its physicochemical and biopharmaceutical characteristics, with a particular emphasis placed on the absorption, distribution, metabolism, excretion, and toxicology (ADMET) process (12, 13). In terms of achieving developability of these drug candidates, it appears that modification of lead candidates by designing-in developability characteristics, translating to reduced physicochemical liability (14), is not always sufficient, and more commonly successful development of these drugs lies with various bio-enabling formulation technologies (15, 16).

Bio-Enabling Formulations

Traditionally, conventional immediate release formulations have been sufficient for successful delivery of drugs with favourable absorption properties, leading to efficient formulation development and limited resource investment. However, the biopharmaceutical challenges and variability of poorly water-soluble drugs (PWSD) has limited the applicability of these approaches in place of bio-enabling formulation strategies (14, 17). Bio-enabling formulations enhance the drug dissolution and/or GIT solubility through various means of either altering the present drug form or its physicochemical properties or co-administering solubilising excipients to facilitate adequate solubility and dissolution characteristics along the GIT (16). While certain gaps in knowledge remain regarding the most effective employment of bio-enabling formulations (6, 15), they provide commercially scalable approaches that meet strict regulatory standards during the drug approval processes. Numerous bio-enabling strategies may be employed to overcome these biopharmaceutical limitations and achieve drug concentrations in excess of equilibrium solubility in GI fluids. These include: 1) particle size reduction (micronization/nanonization), 2) modification of crystal form (pharmaceutical salts, co-crystals, or polymorphs) 3)

complexation/solubilization (eg. surfactants, cyclodextrins or lipid-based formulations), or 4) development of a solid dispersion (17-19). As reference will be made to both solid dispersions (SD) and in particular, lipid-based formulations (LBF) throughout this thesis, a brief introduction of these strategies is now presented.

Lipid-Based Formulations

LBFs offer numerous biopharmaceutical advantages for PWSD resulting from over half a century's worth of research and investment. The original rationale for investigating lipid-based systems to improve absorption was the observation that numerous examples of PWSD demonstrated significantly favourable increases in oral bioavailability when co-administered in the fed state compared to pre-prandial conditions (14). The term LBF spans a wide range of formulations, composed of pure oils or mixtures of oils, surfactants and/or co solvents in various proportions as classified by the lipid formulation classification system (LFCS) (20, 21). Previous research has suggested that many of the marketed LBF products consist of Type II or III formulations, often referred to as self-emulsifying drug delivery systems (SEDDS) (22). These can spontaneously emulsify upon dispersion due to the presence of surfactants and hydrophilic excipients, decreasing reliance on endogenous lipid digestion to facilitate emulsification (6).

The administration of lipid excipients enhances the drug solubilisation capacity of the GI environment, stimulating endogenous bile acid secretion, leading to production of a mixture of solubilising colloidal structures composed of endogenous and exogenous lipids (23). These can effectively solubilise the PWSD (14, 24, 25) and the drug is retained either solubilised or in a transiently supersaturated state allowing for

increased absorption (14). The "spring and parachute" analogy applies here to the generation and prolongation of supersaturation where the "spring" involves the self-emulsifying properties of the LBF, incorporating the solubilised active substance (26), while "parachute" refers to formulation additives which increase stability, reducing drug precipitation *in vivo* (27) (Figure 1-1). LBFs are also biopharmaceutically advantageous regarding impact on intestinal permeability (28), metabolism (29) and lymphatic transport (30, 31).

LBFs have been traditionally employed for drugs which display poor aqueous solubility and high lipophilicity (logP). "Grease-ball" molecules are therefore commonly quoted as strong candidates for LBFs owing to their dominant lipophilic and relatively weak crystal lattice energy characteristics which favour solvation in lipids. Conversely, while "brick dust" molecules may display acceptable lipophilic character for lipid solvation, the high hydrophobic interactions of these drugs provide the ultimate limiting factor to acceptable dose loading levels as drug solubility is limited by a high crystal lattice energy (6). Resultantly, the prevalence of "brick dust" compounds present a challenge for development of drugs with more traditional LBF types. This has resulted in the establishment of a broadened landscape of both formulation compositions and excipient combinations, as well as emerging novel LBF applications, including supersaturated LBFs (sLBF). These are kinetically stable solutions containing a drug concentration above its thermodynamic solubility, where increased drug loads are achieved through thermally inducing supersaturation to overcome the crystal lattice energy (32-34). Finally, from a pharmaceutical manufacture standpoint, once acceptable manufacturing equipment is in place, large scale manufacture of LBFs is relatively low risk and less technologically demanding

which can usually be completed on a smaller scale than other delivery technologies (35, 36).

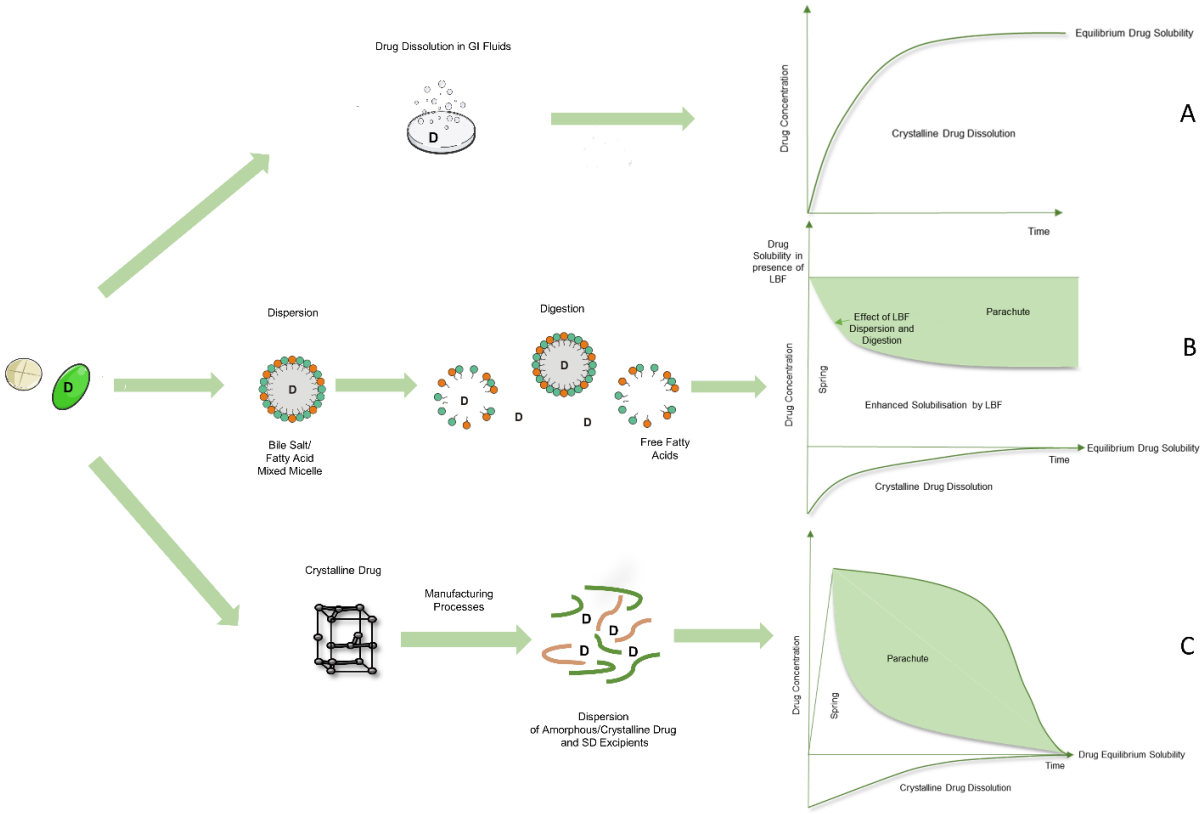


Figure 1-1: Visual representation of modes of action of A) conventional immediate release oral drug products, B) LBF products, C) SD products. Adapted from Feeney et al. (14) and Williams et al. (26).

Solid Dispersions

The merits of SDs to improve oral absorption has been demonstrated as far back as the 1960s (37, 38). SDs are generally two-component systems, containing one or more active substances dispersed in an inert matrix. Depending on the physical state of the carrier, SDs are classified as either crystalline or amorphous, while the API can be also be presented as amorphous or crystalline particles or as a molecular dispersion (39). SDs can facilitate increased solubility and dissolution through a reduction in API particle size, potentially to a molecular level, enhanced wettability and porosity, and altered drug crystalline state, preferably to an amorphous state (40). In its most commonly used form, a SD involves dispersion of drug in an amorphous polymer

matrix with drug present in the molecularly dispersed state (a glass solution) (41). This composition exploits the fact that the solubility of the dispersed or amorphous state can be much higher than comparative solubility of the most stable crystalline polymorph, thus, a supersaturated solution is more easily attained (6). Upon amorphisation, the impact of crystalline long range order on drug solubility and dissolution is largely reduced as intermolecular interactions are weaker and Gibbs free energy is increased (42, 43) Thus, SDs are considered useful for drugs which exhibit solid state limited solubility (i.e. 'brick dust' molecules), but can also be of merit for "grease ball" type molecules due to reduced particle size and increased hydrophilicity due to excipients (44, 45).

Compounds are classified as glass formers (Class II and III) or non-glass formers (Class I) according to Baird et al. (46). These classifications are made by their crystallisation behaviour during a cycle of heating/cooling/heating. The drugs most suited to the amorphous form are class III compounds which melt to form a stable glass and do not recrystallize during either cooling or a second heating. SD systems contain stored potential energy similar to a "spring" which when dispersed can release and forms a supersaturated state when exposed to the GIT (Figure 1-1). The innate thermodynamic instability of the supersaturated state may lead to precipitation or in the case of amorphous SD premature recrystallization, which historically limited the wider application of these formulations. A variety of excipients such as polymers can be utilised to act as a "parachute" in the prevention of precipitation or recrystallization and maintain the solubility advantage. Successive generations of SDs have been produced each providing updated and altered excipients such as polymers to maintain this amorphous solubility advantage or more recently facilitating sustained drug release (38, 39, 47). As the types of SDs continue to expand so does the methods used

to manufacture them. In general, these methods are mainly divided into “solvent-based” and “melt-based” methods (6, 39, 43), where the choice of manufacturing method is often based upon drug physicochemical properties, particularly for thermally unstable compounds which can render certain techniques more or less suitable.

Evolving Pre-Clinical Formulation Design and Selection

Oral formulation development must cope with both biopharmaceutical and technical challenges as well as limited timelines and resources. Therefore, the unavoidable necessity for development of such sophisticated bio-enabling delivery systems generally requires considerable resource investment. Historically, companies have shown marked differences in their decision-making approaches towards formulation strategy, with many companies predominantly focusing on techniques with which they are familiar. Many organisations have followed a pragmatic approach, deliberately focusing on speed, ease of implementation and minimization of cost expenditure (16, 48, 49). In these cases contributory factors in guiding formulation choice are often still dependant on in-house expertise (16), as physicochemical properties and specific biopharmaceutical limitations are often overlooked. This all means that biopharmaceutical investigations of the API and formulation testing at the different stages of drug discovery and development is still, at its core, an empirical process rather than based on in depth mechanistic understanding, and significant gaps remain (50). As a result, several formulation options are often developed in parallel using cumbersome, iterative formulation screening assays, a highly resource-intensive strategy (16). Clearly, significant focus ought to be placed on not only formulation development but also the level of structure that is available to guide and streamline

their selection. Such an increasingly decision-based approach where predictive tools are applied, can help to overcome any reticence to implement bio-enabling formulations and represent an attractive process in comparison with the reagent-expensive and time-consuming alternative of bench-based empirical design.

Over the last 15-20 years progress in understanding and capacity to predict the GI absorption process has led to the implementation and refinement of biopharmaceutical predictive tools of varying complexity to forecast drug behaviour (49, 51-55). As the predictive capacity of any tool depends on its ability to sufficiently simulate the dynamic GI environment, significant efforts are being made to refine this process by improving the efficiency and ensuring validation of current bio-predictive screening tools (54). For example, while traditionally, *in vitro* dissolution testing for new drug products were conducted using aqueous media or buffers, since its introduction to the drug development space in the last 20 years with subsequent refinement, dissolution and solubility in biorelevant media are now widely assessed to provide greater simulation of the human *in vivo* environment in either a pre- or post-prandial state (56, 57). Further advances are represented by the introduction of *in vitro* permeability models (58), pharmaceutical profiling (59), *in silico* PBPK models for integration of *in vitro* data and prediction of GI drug absorption (60, 61) and increased physiological understanding of the ability of animal models to closely mimic human *in vivo* exposure (62, 63). While a recent review of current practices exposed a significant reliance upon *in vivo* testing in animals which remains the mainstay of diagnosing biopharmaceutical performance (48). In addition to these methods, various well-established drug classification systems, which also concern drug properties including solubility and permeability, likewise provide important tools in guiding development.

Drug Classification Systems: Moving from the Drug Substance to Understanding Developability

The significance of various biopharmaceutical drug classifications to inform formulation development strategies cannot be overlooked. In particular, the biopharmaceutics classification system (BCS) represented a substantial shift in the drug development process by introducing the concept that drug classifications attained from *in vitro* methods could be utilised to guide formulation strategies. The BCS provides a diagnostic framework for characterisation of these properties which limit oral bioavailability and development, identifying the rate limiting step as being either solubility or permeability (64). PWSD typically fall into BCS Class II or IV and represent candidates for bio-enabling formulation strategies (Figure 1-2). Since its adoption the BCS has been used widely in the pharmaceutical industry to categorise drug candidates in early formulation development and facilitates BCS-based biowaivers for highly soluble compounds (BCS Class I/III) replacing *in vivo* bioequivalence studies for eligible candidates. A significant modification to the BCS is the biopharmaceutical drug disposition classification system (BDDCS). This system aims to predict drug disposition characteristics as a predictor of absorption, through assessment of drug metabolism in place of intestinal permeability, while also incorporating effects of metabolising enzymes and transporters *in vivo* and drug disposition in development (65). While, in recognition of the importance of investigating solubility and dissolution under physiologically relevant conditions to select appropriate formulations, and to reduce the risk of food effects, a modified BCS for intestinal absorption using biorelevant solubility media, has also been published (66).

It could be suggested that one of the most significant contributions to using an extension of the BCS to inform early evaluation of drug developability and oral formulation decisions was the developability classification system (DCS) introduced by Butler and Dressman (13). The DCS, like the BCS, uses estimates of solubility and permeability for classification (Figure 1-2). However, unlike the BCS, an estimate of *in vivo* intestinal solubility using biorelevant media (FaSSIF) is used, with an increased effectively available volume of 500 mL used for dose/solubility calculations. The DCS also considers the concept of a solubility limited absorbable dose (SLAD), which is the maximal dose that could potentially be absorbed, factoring in both biorelevant solubility in physiologically relevant fluid volumes in the GIT and the compensatory effects of solubility and permeability on dissolution *in vivo*. It also provides a means of estimating the critical particle size at which dissolution becomes rate-limiting to absorption. A clear advantage of this developability assessment tool is that a further differentiation is made within class II drugs, according to either a dissolution (IIa) or a solubility limitation (IIb) of drug absorption, where the SLAD represents the boundary between these subclasses. These subclasses are meaningful for formulation selection. Various dissolution limitations (class IIa) may be overcome with rather simple formulations based on particle size-reduction, while class IIb drugs typically require extensive formulation effort with bio-enabling formulations to overcome the solubility limitation to absorption.

While the DCS provided a solid foundation for a classification system focusing on early drug development, limitations for certain subcategories of compounds (e.g. weak bases) were acknowledged (13). As a result, the refined developability classification system (rDCS) was developed (67). The rDCS, offers two levels of investigations,

"standard investigations" for all compounds and "customised investigations" only when specific criteria are met. The first level of standard investigations, like the original DCS, with minor modifications, classifies the drug using the same DCS classes, according to its dose/solubility ratio and permeability. This provides a first impression of whether development of a drug may be challenging. Depending on the outcome of these standard investigations, various customised investigations using small scale *in vitro* assessments may be triggered when potential exists for supersaturation/precipitation (e.g., salts of acids, weak bases) or to investigate permeation versus dissolution-limited absorption. Incorporation of customised investigations in specific cases provides a more complete understanding of the potential *in vivo* performance of a compound and have strengthened the applicability of the DCS to the drug development setting (67). Overall, investigations based on biopharmaceutical drug class, using these various classification tools, provides a modern starting point for drug development strategies as they begin to shift focus from the drug substance to understanding developability. However, it remains clear that combined use of both traditional and emerging biopharmaceutical tools is required to achieve a structured pre-clinical development pipeline where only the most promising candidates are selected for testing.

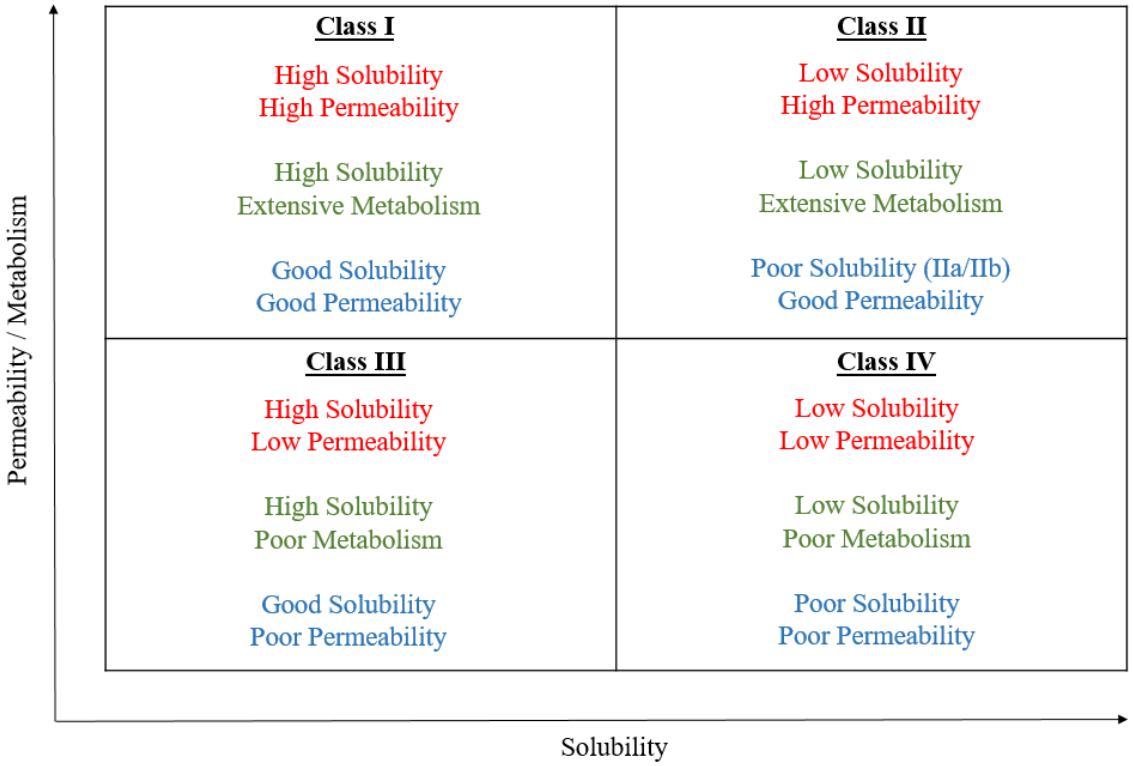


Figure 1-2: Schematic representation of the various classification parameters for drugs using the BCS, BDDCS and DCS classification systems. Red = BCS, Green = BDDCS, Blue = DCS. Drugs are further separated in DCS Class 2, IIa = dissolution rate limited, IIb = solubility limited. Scales and measurements per parameter are different depending on the classification system.

Can Structured Formulation Development Be Achieved Through Property Based Rules?

In spite of the growing numbers of biopharmaceutical tools and classification systems, scope exists for improved predictability in drug product development decisions (15, 16, 68). This is essential to streamline the both selection and development of bio-enabling formulations. So that questions regarding whether adequate formulation approaches can rescue compounds are answered at the earliest point, bypassing costly late-stage terminations. While commercial success has been achieved using various bio-enabling formulations for more than half a century, the selection of an optimal

strategy for new drug candidates is still a significant challenge for most pharmaceutical scientists. Accordingly, a significant effort has been made to adopt a more focused and structured approach based on improved scientific knowledge and expertise, with special reference made to drug physicochemical properties.

Successful development of oral drugs requires consideration of numerous factors, including physicochemical, biopharmaceutical and physiological determinants (69). While, the BCS indicates the importance of solubility and permeability as cornerstone biopharmaceutical parameters which determine oral absorption (64), physicochemical properties of the API also play a crucial role in GI absorption (70). These properties can significantly influence drug performance owing to their impact on dissolution, precipitation, permeability and food effects. As a result, since the 1990s significant interest has been placed on efforts to link drug properties with *in vivo* performance (54), to increase understanding of how molecular properties and descriptors might correlate with a better chance of a candidate being like a "drug". The term "drug-likeness" has been used to used describe the concept of acquiring as much relevant information on the structural features and physicochemical properties of compounds to help discover new drugs which maximally resemble existing drugs, with respect to these key physicochemical and biological properties (71). Certain parameters have been considered to be associated with this likelihood or "drug-likeness" which include but are not limited to; molecular weight (MW), the calculated logarithm of the octanol−water partition coefficient (clogP), number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), topological polar surface area (TPSA), rotatable bonds (RB), and number of aromatic rings (70).

Overtime, different property-based rules have emerged which identify molecular properties perceived to increase the probability oral absorption, occasionally defined as "drug-ability". Of these, the most recognisable is Lipinski et al.'s Rule-of-Five (Ro5) (72). The Ro5 was developed as a guide to design molecules with the potential for high oral bioavailability, stemming from a study conducted at Pfizer in the late 1990s regarding the favourable absorption properties of several thousand orally administered drugs and clinical candidates. As a result, cut-offs for 4 key properties were calculated for which 90% of the molecules studied were compliant. Each threshold was a multiple of 5. It was deemed that molecules which exhibited the best solubility and permeability were found to have: a MW below 500 Da, clogP smaller than 5, no more than 5 HBD and no more than 10 HBA. This simple rule states that when a compound exhibited two or more Ro5 violations it is considered problematic and poor absorption or permeation is more likely. The inclusion of the 4 properties was justified by the original authors as increasing MW has been previously related in literature to poorer intestinal permeability (73). Lipophilicity, expressed as clogP, is a widely utilised property important for solubility but also previously correlated to membrane permeability measurements for compounds which are passively transported (74, 75). While too many hydrogen bond donor and acceptor groups hinder permeability across a membrane bi-layer (76).

Even though the Ro5 was designed to provide chemists with a simple means of predicting potential problems with solubility and permeability, as these factors strongly influence drug absorption (77), limitations of its subsequent adoption as a rule of thumb for druglike properties are acknowledged. For example, both natural products and compound classes that are substrates of biological transporters are

exceptions to the rule (72). Furthermore, subsequent examination of oral drug parameters has demonstrated parameters such as clogP and HBD have remained constant while the cut-offs for parameters such as MW and HBA have increased substantially over the past 20 years (70, 78, 79). This has prompted some researchers to place into question the hypothesis that MW is a "drug-like property" (70). Conversely, drug lipophilicity, which has been demonstrated to be changing less over time than other physical properties, has been suggested as an especially important drug-like property (79). Additionally, recent studies on drug absorption and developabillity have also highlighted the importance of polar surface area (PSA), logD (octanol–water partition coefficient at various pH values), number of rotatable bonds (RB) and the number of aromatic rings (79-81). Overall, despite limitations of this property filter, its conceptual simplicity and ease of calculation has made it the leading measure of drug-likeness and it is still applied today, 20 years after its conception, though perhaps without rigid adherence.

In addition to the Ro5, the "Oral PhysChem Score" uses a traffic light system based on drug values of MW, clogP, RB, calculated aqueous solubility and TPSA (82). Where a lower summed value of the five traffic light properties indicates improved biopharmaceutical properties of a candidate. While the Ro5 remains the most widely known rule for "drug-like" properties, the Rule of Three (Ro3), has been adopted into the concept of "lead-like" properties in fragment-based drug discovery, indicating successful hits exhibit particular physicochemical properties of MW < 300, clogP < 3, number of HBD and HBA < 3 each, and number of rotatable bonds <3 (83). In addition, the Ro3 study suggested PSA (≤60) might also be useful criteria when constructing fragment libraries for efficient lead discovery.

Overall, several rule-based systems have been developed to forecast absorption and drug-likeness. As these guides are successful in the candidate identification space, the point must be raised if similar approaches based on drug properties could be employed to inform formulators about hurdles to expect during formulation development. Undoubtedly, property-based rules that direct structure to formulation selection strategy through use of properties which are available during pre-formulation, would be highly advantageous.

Formulation Decision Trees

At present, use of drug properties to provide organisational structure in the pre-formulation phase of development is primarily seen in use of instructive formulation decision trees (DTs). These flowchart DTs link drug properties with decisions on the preferred formulation technology for a particular drug. They have aided a shift in industrial product development from an essentially empirical approach to a more rational decision-based approach (16, 48) and are grounded on the biopharmaceutical risks of the API. Many companies have developed internal "yes/no" DTs based primarily on prior organizational experience. Typically, these DTs concern various biopharmaceutical properties, combining *in vitro* and *in vivo* data with *in silico* modelling to suggest suitable cut-offs. Typically, various categories of DTs are employed depending on the stage of formulation selection. Firstly, the formulation scientist must make the distinction between the conventional versus non-conventional nature of the formulation required. Thus, the main focus of these initial DTs is to provide an indication of whether a bio-enabling formulation strategy ought to be developed for clinical studies, which represents an important distinction as both the developmental timeline and cost requirements will be dependent on this necessity.

Approaches followed may be different, but a stepwise approach considering a combination of crucial elements such as the dose to be administered orally in relation to the API solubility in biorelevant media (dose number), aqueous solubility or data from relative bioavailability studies of the API in suspension versus solution using rats or beagle dogs are common (16, 35, 48, 53, 84). Additionally, the significance of the aforementioned biopharmaceutical drug classification systems to formulation strategies is also evident as formulation development plans often hinge on the BCS classification of a drug, which is used as a starting point for decision making in numerous DTs including examples from Boehringer Ingelheim, Janssen and Bristol Myer-Squibb (48, 85). While other DTs have utilised DCS classifications to differentiate drugs requiring either dissolution or solubility enhancing strategies using the SLAD principle (86).

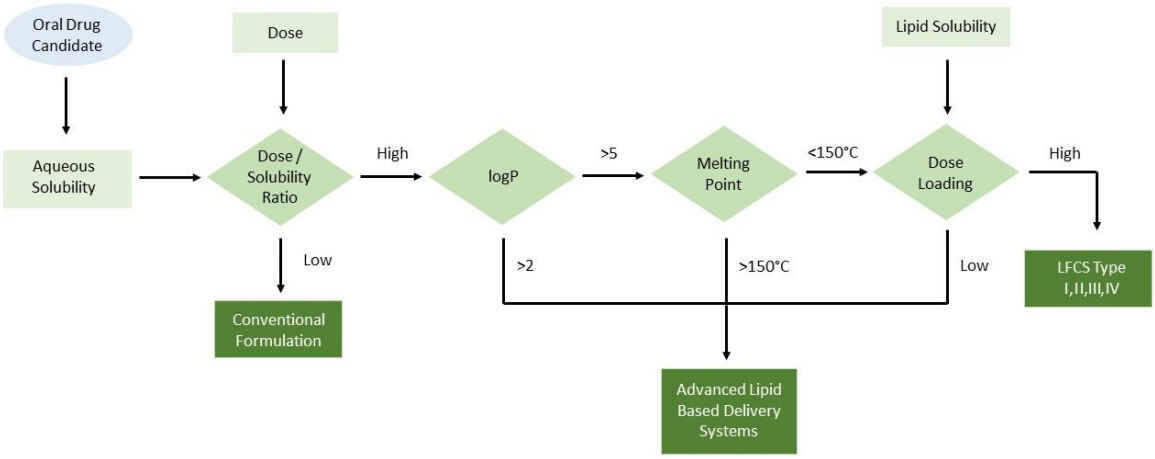


Figure 1-3: Exemplary formulation decision tree flow chart for the selection of lipid-based formulations.

After identification of a requirement for a bio-enabling approach, the next step typically involves decisions relating to the most appropriate class of bio-enabling formulation to be implemented. Here, DTs have also proven useful developability assessments. As an example, in terms of LBF systems, previously published DTs have included multiple decision nodes properties including the extent of drug solubility in

aqueous media or the lipid vehicle, melting point, logP or human dose (Table 1-1 and Figure 1-3) (16, 84, 86). While a further previous example of a structured development for LBFs emphasised excipient solubility screening and considerations of the biological effects of various excipients, before proceeding to various *in vitro* tests including dispersion and *in vitro* lipolysis in response to the significance of *in vivo* processing to overall performance (87). This significance of dispersion and digestion is echoed by the lipid formulation performance classification system (88). Using this informative tool formulation performance is graded upon 4 grades A to D depending on their ability to withstand drug precipitation and maintain solubilisation under various *in vitro* conditions i.e., dispersion in simulated GI fluids and digestion under simulated "normal" and "stressed" intestinal digestion conditions. As a result, increasingly effective formulation discrimination and performance grading can be achieved in drug product development (88). Furthermore, DTs have been developed which indicate appropriate candidates for SD formulations (Table 1-1). Among these examples, significant characteristics for decision making included glass forming ability, glass transition temperature, melting point, heat stability, presence of ionisable groups, MW, torsional bonds, HBA and PSA (16, 85, 86).

Table 1-1: Published Examples of Formulation Decision Trees for LBF and SD Bio-Enabling Formulation Selection

Author	Formulation Type	Decision Node Properties Included
Kuentz et al. (16)	SD	Glass Forming Ability Melting Point Degree of Maximum Supersaturation Heat Stability Solubility in Volatile Solvents
Kuentz et al. (16)	LBF	Aqueous Solubility Relative to Dose logP Melting Point Dose Solubility in Glycerides Solubility in Polyethylene Glycol or Propylene Glycol Derivatives
Kuentz and Imanidis (86)	LBF	BCS Class DCS Class Dose Melting Point logP
Kuentz and Imanidis (86)	SD	BCS Class DCS Class Dose Melting Point logP Molecular Weight Number of Torsional Bonds Number of Hydrogen Bond Acceptors Polar Surface Area
Ku (85)	SD	Ionisation Potential Dose Volume Solubility in Vehicle Solution Stability
Rabinow (84)	LBF	Potential for Salt Formation Aqueous Solubility logP Melting Point Dose

Overall, DTs have emerged as increasingly standard biopharmaceutics predictive practises in the formulation characterisation toolkit, making use of currently available knowledge. However, limitations are apparent. Firstly, whilst use of DTs have been suggested to significantly reduce the demand for animal studies in line with the 3R initiative (49, 89), depending on the specific DT certain *in vitro* and *in vivo* testing will be initially required to identify a suitable formulation strategy. Meaning that even though a more structured approach is provided to decision making, the practical requirements for physical formulation screening are not lessened to any meaningful degree. Furthermore, it could be suggested that they provide a sometimes oversimplified or biased view of the formulation development process and frequently omit all potentially suitable formulation strategies, providing only a reflection of tendency for formulation suitability. Their impact is also limited by complexity in decision nodes numbers. Lastly and perhaps most significantly, while DTs from different companies are typically compiled of similar key elements, each decision node has its own rationale making sense in the context of the specific settings of each company, which may not be universally applicable or undisclosed. This make it difficult to assess which is the superior approach or to define an overarching DT/strategy that would work across companies. Even though limitations of DTs are apparent, more generally the concept of anticipating "successful formulation potential" from drug properties and the use of property-based rules is undoubtedly gaining significant interest.

Can a Retrospective Biopharmaceutical Analysis of Drug Properties Inform Future Success?

Upon reflection of the potential for such DTs to inform developability, the opportunity for broader applicability of property-based rules to guide bio-enabling formulation selection has prompted interest from this author regarding other drug properties-based analyses to predict formulation suitability. Stemming from the fact that DTs in general provide crude estimates and are fundamentally built upon previous experience, it was hypothesised if an alternative "experience" based expert learning approach could be adopted through the retrospective analysis of drugs which have reached commercial success using certain bio-enabling approaches. Consequently, this could provide a statistics-based assessment of drug properties which favour certain formulation types, aiming to improve predictability in the formulation development process. As a result, the aim of the following work was to provide a retrospective, top-down, analysis of the current landscape of commercial products using certain bio-enabling approaches. Identifying which drug properties are likely to identify successful delivery technologies at an earlier stage in development. This analysis focuses on the commercial utility of the two most encountered bio-enabling formulation approaches; LBF and SD, due to the extensive reports in the literature on their capacity to enhance oral delivery, and numerous examples of commercial successes as licensed drug products in clinical use.

As part of this analysis an up-to-date and comprehensive list of commercially available LBF and SD formulations is provided, trends in the type of drugs and formulations currently reaching the marketplace are discussed and key physicochemical and biopharmaceutical predictors of successful formulation development are identified. In order to achieve these aims, the commercial examples to date of drug products

formulated as either SD or LBF is examined and classified according to BDDCS class of the formulated active substance. Selected physicochemical characteristics and molecular properties of these commercial drugs are statistically analysed and compared to a list of compounds not produced via either technology. The aim of this analysis is to explore which drug properties signal suitability of a drug for LBFs or SDs, or moreover, properties which potentially distinguish between them. Thus, attempting to bridge a gap in current drug development, involving widespread use of drug-likeness filters and ADME optimisation to guide drug discovery and refine drug candidate selection. While many merits exist for their use, there also exists a risk that current filters may be overly conservative and conceptually simplistic. As increasing numbers of drugs emerge beyond the preferred chemical space it could be argued that complementary use of "formulation-likeness filters" in such instances could inform developers of bio-enabling technologies which may be appropriate, based on properties of their drug candidate, simultaneously analysing potential for success in terms of both drug likeness and bio-enabling potential. As the numbers of drug compounds using both LBF and SD in licensed commercial products continues to grow, so too does the database of information regarding suitable drugs compatible for such systems. This data bank could guide future commercial success of LBF and SD products, reflecting backwards in order to move forwards in the "bio-enabling" field with confidence.

Methods

Dataset Selection

An original databank of approximately 1000 drug compounds was collated from previous literature sources (90, 91) using the BDDCS classification and an in house database of oral drug compounds commercially approved by the EMA and FDA between 2010 and 2017 (92). Where information regarding BDDCS classification was not available, a drug's BCS classification was used as a surrogate due to the same parameter of solubility being used in both classifications. This master databank was split into three, namely, drugs commercially developed as LBF, SD and Others i.e., not commercially developed via either technology. LBF and SD drugs were identified from previous literature referencing commercial products (6, 42, 93-100), along with analysis of the online databases of the US and EU respective drug licensing authorities (Food and Drug Administration, European Medicines Agency, Health Products Regulatory Authority of Ireland) where dosage, licencing and excipient information regarding all products was also then obtained. Where a product was identified in peer reviewed literature but was not authorised in these three areas another national authority was investigated to establish if the product had been commercialised. A LBF was defined as Class I-IV of the Lipid Formulation Classification System (20, 21). All types of SDs were considered based on description in product or published literature that the product is a SD (i.e., both amorphous versus crystalline API dispersed in amorphous carriers.) Omega-3-acid ethyl esters, Florfenicol and Silibinin were removed from the database due to the lack of drug property data available. Exclusion criteria for the Others list included any drugs used in LBF or SD commercial products, active metabolites and non-orally delivered drugs. The final datasets contained 49 drugs grouped as LBF, 37 as SD and 763 as Others drugs. When including only poorly

soluble BDDCS Class 2/4 drugs there remained 38 drugs grouped as LBF, 30 as SD and 307 as Others drugs.

Compilation of Physicochemical Descriptors

Physicochemical properties to be assessed were identified and compiled from the literature publication "BDDCS Applied to Over 900 Drugs" (91). Physicochemical and molecular properties for the drugs not listed in Benet *et al.* above were obtained from PubChem, DrugBank or ADMET Predictor 9.5 (Simulations Plus, USA). The final properties of the drugs analysed included: Molecular Weight (MW), Maximum Dosage Strength (MDS), Hydrogen Bond Acceptors (HBA), Hydrogen Bond Donors (HBD), Polar Surface Area (PSA), Measured Partition Coefficient (logP), Calculated Partition Coefficient (clogP), Percentage Excreted Unchanged in Urine (U%), pDose, Logarithm of Aqueous Solubility (logS), Partition Coefficient at pH 7.4 ($logD_{7.4}$) , Rule-of-5 Violations (Ro5), pKa (Strongest Acidic), Melting Point (T_m) and Rotatable Bonds (RB). These are defined in Table 1-2.

Table 1-2: Definitions of the drug molecular properties and physicochemical characteristics analysed in the statistical analysis.

Property	Abbreviation	Definition
clogP	clogP	Logarithm of a molecules partition coefficient between n-octanol and using the method of Leo.
Hydrogen Bond Acceptors	HBA	Electronegative ion or molecule that must possess a lone electron pair in order to form a hydrogen bond.
Hydrogen Bond Donors	HBD	Heteroatom with at least one bonded hydrogen.
$logD_{7.4}$	$logD_{7.4}$	Partition coefficient of a drug at pH 7.4. This pH is utilised as this is the physiological pH of blood serum.
logP	logP	The measured partition coefficient of a molecule between an aqueous and lipophilic phase (n-octanol/water).
logS (mol/L)	logS	The 10-based logarithm of the solubility of a molecule mol/L.
Maximum Dosage Strength (mg)	MDS	The highest dosage strength licensed for a drug.
Melting Point (C°)	T_m	Temperature at which a solid changes state from solid to liquid.
Molecular Weight (g/mol)	MW	Molecular Mass of a drug.
pDose (mol/L)	pDose	$-\log_{10}$(Maximum Dose Strength) (molar).
Percentage Excreted Unchanged in Urine (%)	%U	The proportion of drug unchanged in the body and excreted in the urine.
pKa (Strongest Acidic)	pKa (Strongest Acidic)	The pH at which the drug is completely balanced between the charged and uncharged form. Strongest acidic refers to the strongest acidic group in the molecule.
Polar Surface Area ($Å^2$)	PSA	The sum of the fractional contributions to the surface area of all nitrogen and oxygen atoms calculated using the method of Clark.
Rotatable Bonds	RB	Any single bond, not in a ring, bound to a nonterminal heavy (i.e., non-hydrogen) atom.
Rule of Five Violations	Ro5	Number of Lipinski's Rule-of-Five violations which predicts poor absorption or permeation.

Statistical Analysis

A stepwise statistical analysis approach was adopted using SPSS (IBM Corporation, US). Frequency distributions of the variables were graphed for each of the three groups and normality was checked visually with Q-Q and P-P plots. Ratios of samples sizes between the 3 groups were obtained. Variances of the datasets were analysed and compared to Levene's Test for Equality of Variances. A p-value <0.05 indicated a violation of equal variance. The null hypotheses were that no differences were seen in a drug property between drug groups. Three separate comparison were made i.e., LBF vs SD; LBF vs Others; SD vs Others rather than a three-group comparison, using for example ANOVA. This enabled use of the most appropriate comparison method based on assessment of data normality and equality of variance in each group and is in line with the null hypotheses identified. Comparison between groups were made using the t-test, Welch's test, Bootstrap independent samples test (5000 samples) or Chi-Square test, all 2-sided, where appropriate. Rule-of-5 violations was recoded to a category variable or ≤1 or >1 violation and Chi-Square tests were used to test independence of this categorical variable. If 1 or more cells had an expected count below 5, Fisher's exact test was employed. A p-value of 0.05 was used as the significance level for all tests. Finally, in order to analyse only PWSD, subsets of the three datasets were created containing only BDDCS Class 2/4 drugs and the statistical analysis described above was repeated.

Results

Commercial Success to Date

We envisage a gap in the literature in terms of a comprehensive list of drugs which have been commercially developed as either LBF or SD, to provide a true measure of clinical development success. Information involving product names, drug compounds and excipients used, dosage forms and strengths was collated (Appendix 1 Table 1-5, 1-6). Some products have been subsequently withdrawn from the commercial market however, all products were licensed at one point in time.

Commercial Lipid-Based Formulations

LBF products have been successfully authorised internationally since the 1940s. Early examples of commercial products consisted of Type I formulations of the LFCS e.g. Drisdol® (101). As years progressed interest in self-emulsifying systems intensified (14) and resulted in a large surge in increasingly complex Type III and IV LBF products in the 1980s-1990s (95). Review of the published literature and online databases of drug product regulatory authorities in the US and EU identified 67 commercial LBF products. As illustrated in Figure 1-4, a higher number of the LBF products have been authorised in the US (47/67) compared to the EU (26/67). Differences in the number of marketed products could represent strategic commercial decisions based on factors such as level of clinical demand or regulatory burden.

In a small number of cases more than one dosage form e.g., capsule, and oral solution, have been produced for the same drug product (6/67 products). In comparison, multiple dosage strengths have been licensed almost half of the products (28/67 products). It was observed that soft gelatin capsules dominantly account for the most

popular LBF product dosage form (40/67), followed by oral solutions (10/67), hard capsules (10/67) and oral suspensions (1/67) (Figure 1-5). There are also 6 products which are controlled release, demonstrating a further drug delivery advantage of LBFs. These are extended-release capsules (3/67), extended-release suspension (1/67) prolonged release capsule (1/67) and sustained release granules (1/67). Clearly, soft gelatin capsules represent the more prevalent dosage form as they can safely encapsulate liquid dosage forms in comparison to hard capsules. While there has been successful suspensions produced (102), solutions remain the most popular approach for commercial products according to our analysis. In terms of year of authorisation, it can be seen (Figure 1-6) that the period of 2000-2009 contained the highest number of commercial LBF approvals (37%). As such, combining the 1990s and 2000s accounts for 63% of all commercial LBF products. However, this spike in approvals did not continue into the period since 2010 where only 9% of all LBF products have been commercialised. Overall, the findings here are comparable to analysis examining growth in the number of LBF/SEDDS publications in PubMed from 1966 to 2016 where they saw a large surge of publication numbers from the mid-1990s (14). Finally, a number of the listed products have been either discontinued or withdrawn from the market (12/67). No trends were evident where the reasons for withdrawal were linked to reasons of efficacy, safety nor stability. In the majority of cases a lack of clinical demand or switch to another dosage form was cited by the manufacturer.

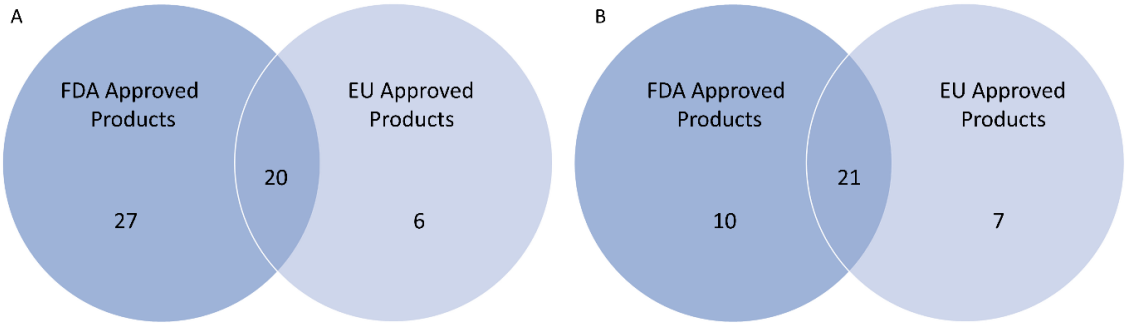


Figure 1-4: Venn diagrams illustrating the numbers of LBF (A) and SD (B) commercial products authorised by the FDA and EU (EMA and HPRA).

Commercial Solid Dispersions

The earliest example of a commercial SD product is Cesamet® (Nabilone) from 1982 (103). Overall, 39 commercial SD products were identified. Four of these have been marketed under a different brand name in a different region (Certican® = Zortress®, Incivek® = Incivo®, Cokiera® = Viekira XR®, Galvumet® = Eucreas®). Compared to LBFs, commercial SDs form a smaller number of licensed products, which may reflect that LBF products were a more established commercial pathway in the 1980's and 1990's, relative to SDs (14, 104). As an example, the first LBF was approved over 40 years before the first SD commercial products (Drisdol®, 1941 and Cesamet®, 1982). When commercial SD products manufacturing methods were analysed, we found the majority of products were produced via either spray drying or melt extrusion methods in line with previous research analysis (44).

The widespread global market for SD products is apparent. From Figure 1-4 close to 50% of SD commercial products are authorised in both the United States and EU markets. Multiple dosage strengths were seen for a majority of products (23/39), similar to LBFs, potentially due to scalability and manufacture of dose proportional

preparations of SDs. In terms of dosage forms immediate release tablets are most popular (27/39) (Figure 1-5). While capsules (4/39) and granules for oral suspension (1/39) are also seen, as well as controlled release tablets and capsules, in the form of extended, delayed or prolonged release. 5/39 identified SD products have been either discontinued or withdrawn from the market. Upon review no evidence could be found to suggest the majority of removals were due to efficacy or safety issues and were voluntary due to declining clinical demand or use of alternative dosage forms. Conversely, in the case of Rezulin® (Trogslitazone), its removal was linked to the development severe idiosyncratic hepatocellular injury (105). However, this is due to the drugs intrinsic toxicity rather than lack of effective formulation delivery.

In contrast to only 9% of LBF products, 54% of SD commercial products have been authorised since 2010 (Figure 1-6), demonstrating a sharp growing development trend toward SDs in recent years. It has previously been suggested that SD formulation technologies have been embraced to a much greater extent since 2012 (44), with comparative spikes in terms of related research articles seen from 2010-2015 (96). As evidence of the commercial success of SD technology, Harvoni® (Gilead Sciences, Inc.), containing Ledipasvir and Sofosbuvir, used to treat chronic Hepatitis C was second in the blockbuster list of drugs ranked by sales revenue in 2015 (106).

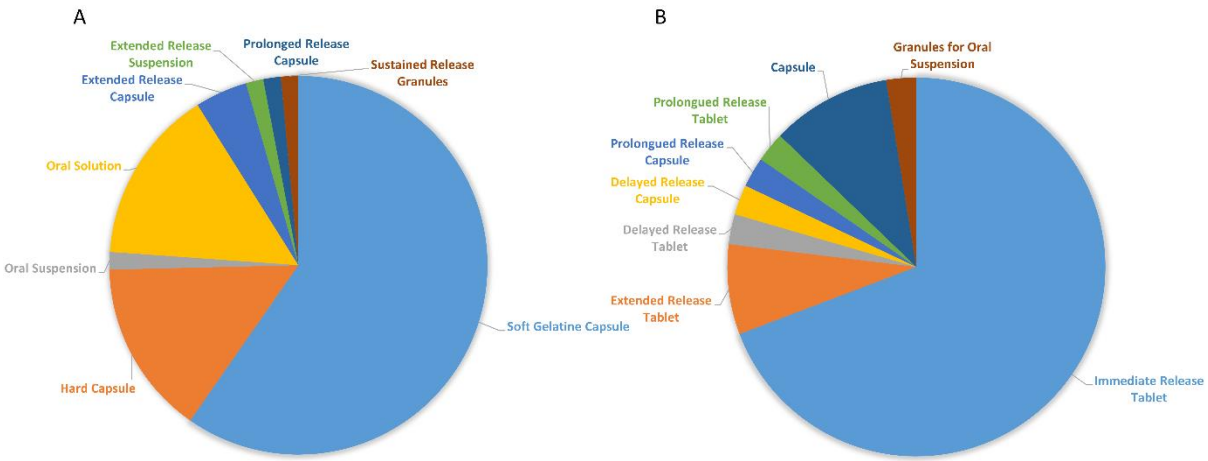


Figure 1-5: Pie charts illustrating the different dosage forms products for LBF (A) and SD (B) commercial products.

Commercial Products via Both Formulation Technologies

Four drugs have been commercially produced via both LBF and SD technologies. These are Fenofibrate, Lopinavir, Ritonavir and Nimodipine. In the case of Lopinavir, it was originally produced in combination with Ritonavir in Kaletra® as an LBF capsule and subsequently replaced by AbbVie Inc.© with the SD tablet form exhibiting a higher dose loading capacity. This resulted in a reduced pill burden and aided compliance while also providing the added advantage of absence of food effect (107). Similarly, Ritonavir has also been commercialised as both a SD and LBF in Norvir® (108). In this case, original liquid filled capsules containing Ritonavir in an ethanol, surfactant and water-based solution were withdrawn from the market due to discovery of a previously unknown polymorph, leading to a significant decline in drug solubility and potential for poor bioavailability (109, 110). When this original form was removed from the market, patients were encouraged to switch to the oral liquid form. In 1999, AbbVie Inc. (previously Abbott), applied for approval of an LBF soft gelatine capsule form overcoming this stability problem which required refrigeration. Ultimately in 2010, this LBF form was replaced by an SD 100 mg tablet which overcame the

requirement for refrigeration, which improves convenience. Therefore, in two cases, choices of both LBF and SDs were largely based on commercial strategies (Fenofibrate and Nimodipine), whereas for Lopinavir and Ritonavir, initially the more established formulation strategy of LBFs were launched, however, due to problems with dose loading and stability were ultimately replaced with SDs. Overall, this relatively small overlap of drugs produced by both technologies observed, could suggest existence of distinctive drug properties which render a drug candidate more suitable for SD delivery over LBF delivery or vice versa.

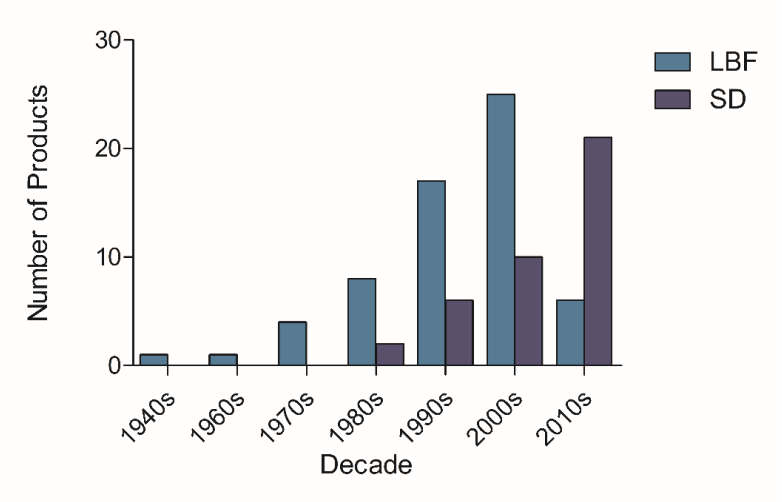


Figure 1-6: Grouped bar chart illustrating the number of SD and LBF commercial products authorised by decade from 1940.

BDDCS Classifications

The three drug sets were grouped according to BDDCS classification. These visual representations are found in Figure 1-7. As expected, the highest numbers of LBF (76%) and SD (60%) drugs in commercial products belong to BDDCS Class 2. Also, as anticipated, the second highest proportion of SD commercially used drugs come from BDDCS Class 4. In contrast, the second highest proportion of LBF drugs were

found to be BDDCS Class 1 which indicates that, not only solubility limited compounds are successfully commercialised via LBFs. This most likely reflects a strategic commercial decision, as opposed to a strategy to address a solubility or permeability limitation, and may reflect that the large scale manufacture of LBFs are generally well established, and require relatively lower technologically input compared to other more expensive bio-enabling platforms such as SDs (35).

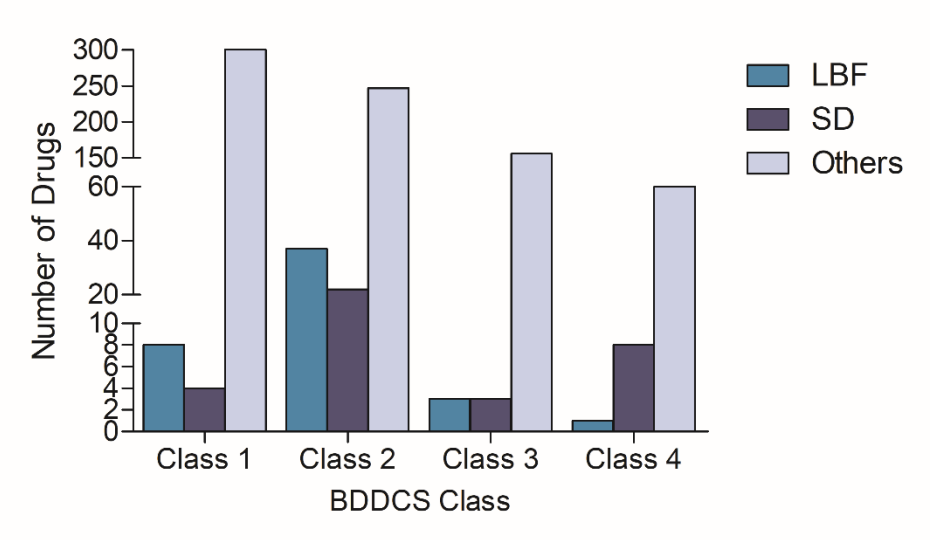


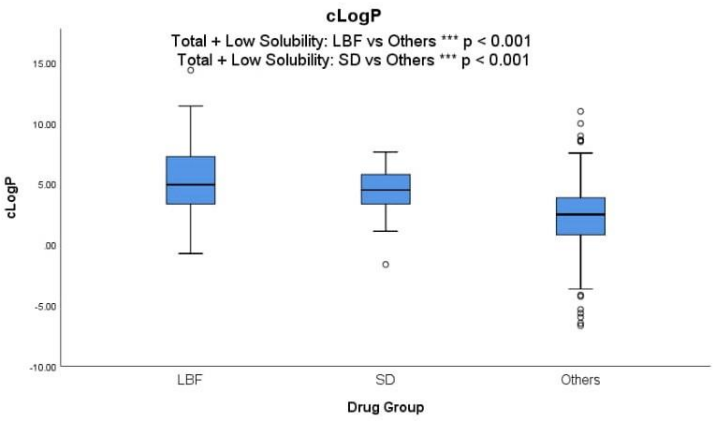
Figure 1-7: Visual representation of the proportion of drug per dataset of LBF, SD and Others drugs in BDDCS Class 1-4.

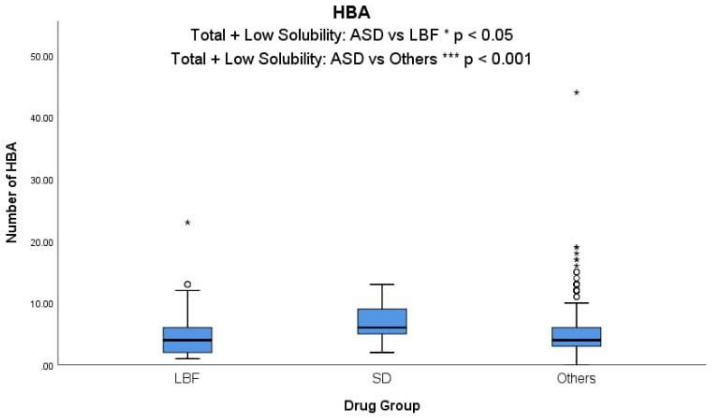
Retrospective Statistical Analysis of Properties of Commercialised LBF and SD Drug Compounds.

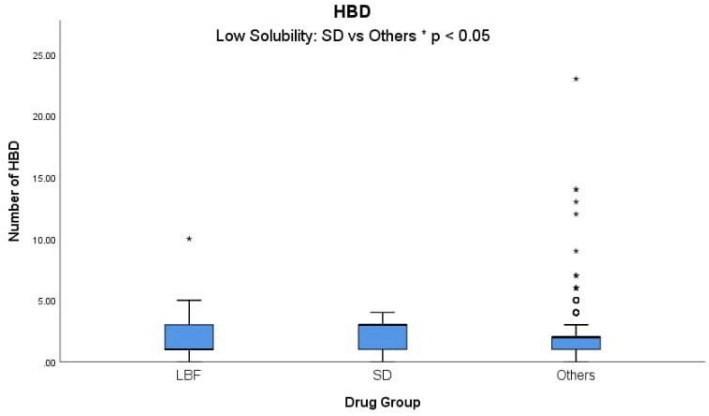
Molecular properties of drugs previously commercialised using LBF and SD formulation technologies were statistically compared with properties of drug substances not commercialised via either technology. Tabular results of the statistical analysis are shown in the Appendix (Appendix 1 Table 1-1 to 1-4). A visual representation of significant differences obtained is illustrated in Figure 1-8.

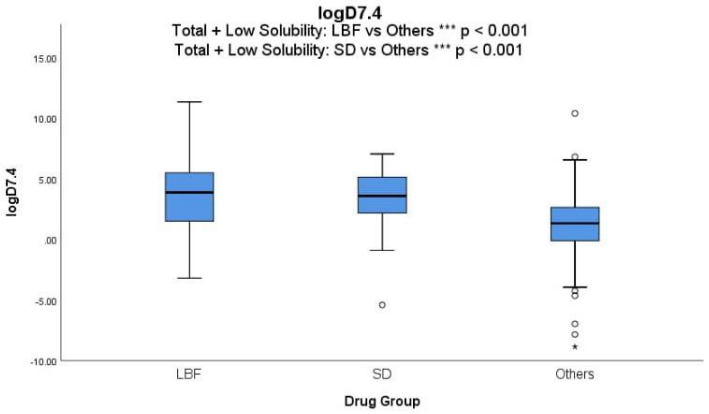
Upon analysis of all BDDCS classes, 8/15 properties were significantly different between the LBF and Others datasets, namely MW, logP, %U, logS, $logD_{7.4}$, Ro5, T_m and clogP. In addition to these 8 properties HBA, RB and PSA were also found to be significantly different between the SD versus Others datasets. Therefore, these properties can be predictive of suitability for commercial success via LBF or SD technologies according to the current commercial climate of both sets of drugs. While no clear trends for the properties of pKa (strongest acidic), MDS and pDose were differentiated between groups, thus, these properties did not appear useful in predicting suitability nor indicative of unsuitability for either formulation type. Between LBF and SD datasets significant differences in drug properties were observed as SDs displayed significantly higher mean HBA, RB, MW and PSA, compared to LBFs.

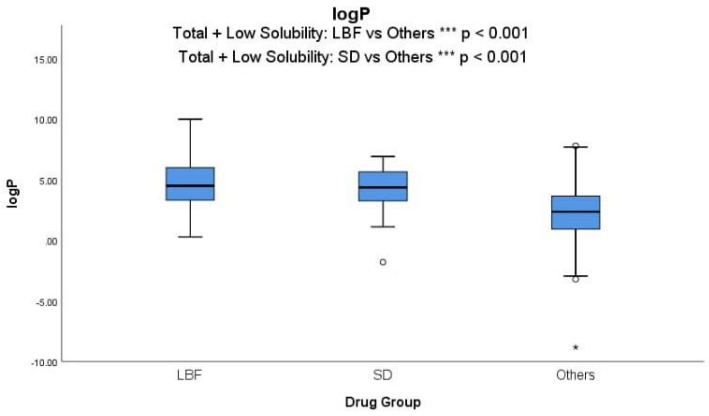
Subsequently, a subset analysis was performed on BDDCS Class 2/4 drugs (low solubility) to explore whether results would be altered by excluding high solubility drugs, typically delivered using conventional methods. This subset decreased the numbers in the LBF group by 22% (n = 38), the SD group by 19% (n = 30) and the Others group by 60% (n = 307). In terms of comparisons between LBF versus Others within this low solubility datasets, this resulted in the parameters of Ro5 ($p = 0.086$), MW ($p = 0.129$) and T_m ($p = 0.051$) being no longer significant, albeit marginally in the case of T_m. Conversely, differences in both MDS (** $p = 0.006$) and pDose (* $p = 0.026$) between LBF and Others gained significance in the low solubility dataset. In terms of comparisons between SD and Others, the low solubility subset did not result in loss of significance to any observation, while MDS (** $p = 0.003$), pDose (* $p = 0.037$) and HBD (* $p = 0.03$) also gained significance.

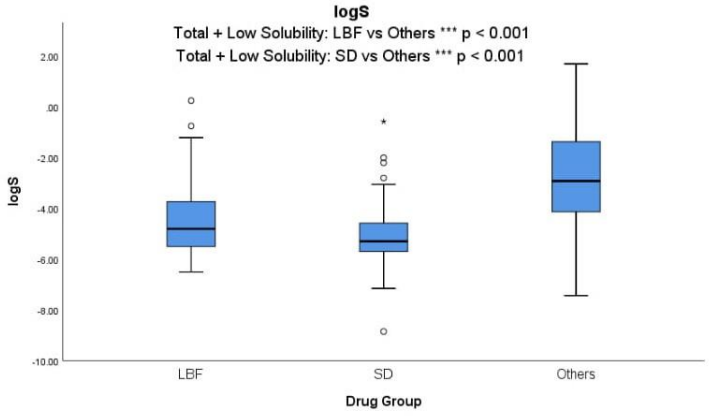

cLogP
Total + Low Solubility: LBF vs Others *** p < 0.001
Total + Low Solubility: SD vs Others *** p < 0.001
cLogP
15.00
10.00
5.00
.00
-5.00
-10.00
LBF
SD
Others
Drug Group

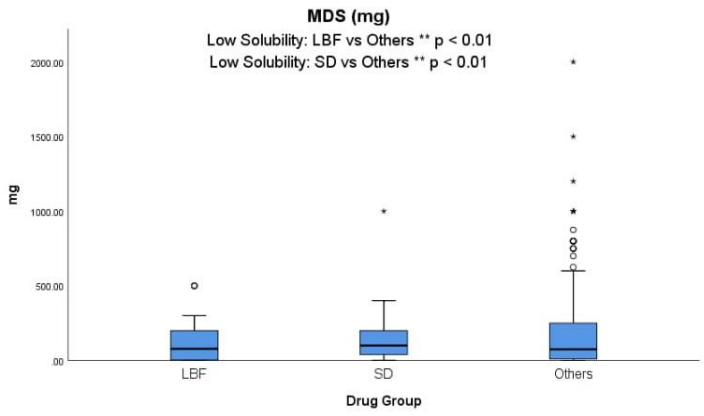

HBA
Total + Low Solubility: ASD vs LBF * p < 0.05
Total + Low Solubility: ASD vs Others *** p < 0.001
Number of HBA
50.00
40.00
30.00
20.00
10.00
.00
LBF
SD
Others
Drug Group

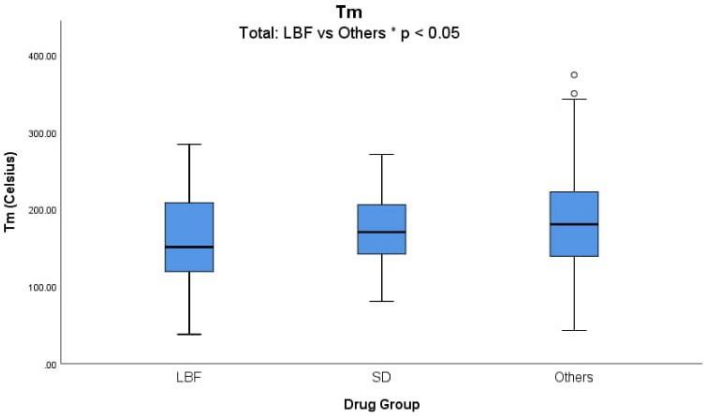

HBD
Low Solubility: SD vs Others * p < 0.05
Number of HBD
25.00
20.00
15.00
10.00
5.00
.00
LBF
SD
Others
Drug Group


logD7.4
Total + Low Solubility: LBF vs Others *** p < 0.001
Total + Low Solubility: SD vs Others *** p < 0.001
logD7.4
15.00
10.00
5.00
.00
-5.00
-10.00
LBF
SD
Others
Drug Group


logP
Total + Low Solubility: LBF vs Others *** p < 0.001
Total + Low Solubility: SD vs Others *** p < 0.001
logP
15.00
10.00
5.00
.00
-5.00
-10.00
LBF
SD
Others
Drug Group


logS
Total + Low Solubility: LBF vs Others *** p < 0.001
Total + Low Solubility: SD vs Others *** p < 0.001
logS
2.00
.00
-2.00
-4.00
-6.00
-8.00
-10.00
LBF
SD
Others
Drug Group


MDS (mg)
Low Solubility: LBF vs Others ** p < 0.01
Low Solubility: SD vs Others ** p < 0.01
mg
2000.00
1500.00
1000.00
500.00
.00
LBF
SD
Others
Drug Group


Tm
Total: LBF vs Others * p < 0.05
Tm (Celsius)
400.00
300.00
200.00
100.00
.00
LBF
SD
Others
Drug Group

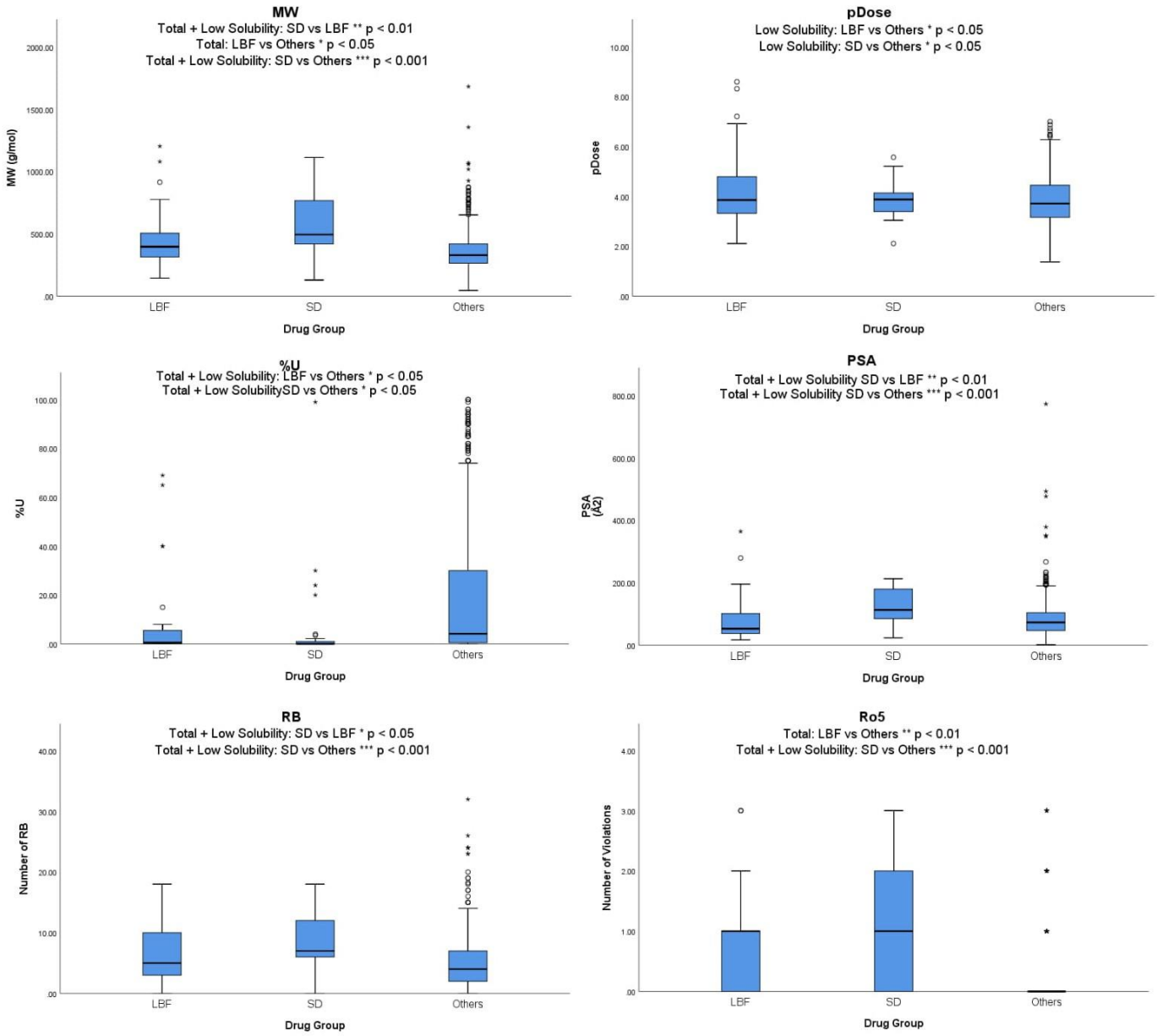


Figure 1-8: Visual representation of the statistically significant differences found between LBF, SD and Others. p-values for the statistically significant pairwise comparisons are shown. "Total" refers to analysis with all BDDCS Classes. "Low Solubility" refers to analysis of only BDDCS Class 2/4. When both "Total" and "Low Solubility" are stated p-value refers to the "Total" result. The dark line in the middle of the boxes is the median. The bottom and top of the box indicates the 25th (Q1) and 75th percentile (Q3). The T-bars are inner fences/whiskers which extend to 1.5 times the box height. The circles are outliers that do not fall in the inner fences. The asterisks are extreme outliers which have values greater than three times the height of the boxes

Property Trends Resulting from a Retrospective Statistical Analysis of Licensed LBF and SD Products.

Based on the statistical analysis of formulation types by drug properties the following general trends were observed in this work.

Molecular Weight (MW)

Drugs commercialised as both LBF and SD pharmaceutical products displayed significantly larger MW compared to those commercialised via conventional formulation approaches (i.e., Others). Comparatively, SD displayed significantly greater mean MW (586.6g/mol) versus LBF (448.2g/mol) suggesting that while both LBF and SD express potential to accommodate high MW drugs, SD approach may offer greater opportunities at the higher MW range. Additionally, only LBF, not SD drugs, lost significance versus Others when a low solubility dataset was analysed, suggesting that as MW increases any benefits LBF confer for PWSD are not as prevalent and preference for SD platforms prevails.

These results reflect drug development trends over recent decades of increasing MW of drug molecules in drug development pipelines (79, 111, 112). In the two last decades, there has been consistent trends for higher MW drugs being brought to market, exemplified when in 2016 and 2017 for the first time, average MW for new FDA approved oral drugs exceeded 500g/mol (70), with widespread increases in MW observed not merely due to approval of a small proportion of very high MW drugs. Such trends fall outside both the Lipinski Ro5 and the Ro3 for fragment based drug discovery (83). Resultantly, this sharp increase has prompted questioning regarding the justification of MW as a property of "drug-likeness" (70).

The trend for high MW observed here should be considered in line with the earlier reported trend for increasing use of SD approaches in the last decade. It is unclear whether these reflect independent trends in technological advances of both SD and increasing drug candidate MW or complementarity of both. However, it is clear that SDs offer a more commercially successful track record for high MW drugs. As most recently evidenced by the high MW antiviral, enzyme inhibitor drugs being delivered commercially in this manner e.g. Cokiera®, Epclusa®, Zelboraf®. These results are broadly supportive of the general rule of thumb that molecules with a MW of >300 g/mol can more easily be transformed into an amorphous state (113). Here, we uncovered only 2/37 drugs commercialised as SDs with MW <300 g/mol. It has also been suggested that comparatively high MW increases glass forming ability (GFA) of a drug (46, 113). While a higher solubility advantage was also demonstrated for higher MW drugs as a result of *in silico* predictive modelling of the amorphous solubility advantage (86). Resultantly, from our analysis MW provides a distinguishing property for potential commercial success between LBFs and SDs at the higher end of the MW scale.

Melting Point (T_m)

A significantly smaller mean T_m was found for LBF drugs (160.81°C) vs Others (181.18°C). This significance was lost, albeit marginally, when a low solubility dataset was analysed. When the variances of T_m among groups was analysed, the smallest spread of values was found amongst the SD group. While the lowest T_m values for LBF and Others groups respectively were 38°C and 43°C, the lowest T_m of a drug produced as a SD was approximately double these figures (80.5°C). T_m is often cited as an important drug characteristic influencing solubility in lipid vehicles, as an

indicator of the energy required to break intermolecular bonds and overcome the crystal lattice energy. Drugs possessing a high crystal lattice energy along with a moderate logP value (>2) are termed "brick dust" (102), typically possessing poor solubility in lipids due to limited capacity to dissociate from the solid form and are not ideal candidates for LBFs (6, 14). Previous work has demonstrated that addition of T_m improved computational predictions of drug solubility in triglyceride vehicles (114). It has been reported that for reasonable solubility in lipid vehicles, a low to intermediate T_m was preferable, and a T_m <150˚C was proposed as a baseline for the selection of LBFs as potential enabling formulation approaches (86, 115-117). However, in this analysis more than half (i.e., 55%) of commercially licensed LBFs exceeded this commonly recommended value of 150°C. A subset analysis revealed however, that the mean maximum dosage strength was significantly lower for drugs exceeding this value (i.e., 148.62 mg for drugs <150°C compared to 81.48 mg for >150°C). Overall, this would suggest that while low to intermediate T_m may be still be recommended, particularly for higher dose products, in the case of low dose/highly potent drugs, a T_m in excess of 150°C may not be limiting.

T_m was not observed to be a predictor of SD commercial success. This was unexpected as T_m was previously demonstrated to be an important predictor for the solubility advantage for amorphous drugs (86), in addition to differentiating between GFA classifications of compounds (46). T_m can also dictate the type of manufacturing method suitable for a particular SD commercial product due to heat unstable components and risks of chemical degradation (44), as well as being related to their glass transition temperature (118).

Lipophilicity (logP, clogP, $logD_{7.4}$)

Lipophilicity remains an important property of drug candidates in development over the last 15-20 years, due in part to the lipophilic molecular requirements of new drug targets (17, 119). It is thought to be correlated with MW, yet it appears to be changing less overtime than other drug properties (70, 79). A 2016 analysis of 1620 molecules patented around that time uncovered that approximately 50% had ligands displaying mean logP ≥4 (2). As such, Leeson and Springthorpe have even suggested lipophilicty to be the most important drug property, where high lipophilicity can result in increased risks of multiple target binding and potential toxicology (79). As expected LBF commercialised drugs displayed significantly higher measured logP, clogP and $logD_{7.4}$ values than drugs compounds in the Others dataset. High lipophilicity would be expected to facilitate sufficient drug loading capacity in lipid vehicles. It is commonly reported that "grease ball" drug molecules, displaying high lipophilicity and relatively low T_m are good candidates for LBFs (3), while the ability to facilitate lymphatic uptake by LBFs is optimised for highly lipophilic drugs (logP > 5) (120). Overall, this finding suggests that drugs with logP values of approximately 4–5 are good candidates for commercial LBFs due to the mean logP value of 4.7 observed. Previously, Pouton and Porter have suggested that a logP >5 demonstrates suitability for LBF as such drug compounds are incorporated into mixed micelles and absorbed efficiently (22). Interestingly, the greatest variance in logP values was also found in the LBF group. This could be related to the diverse range of classes of LBF available (20), where differing quantities of lipophilic and hydrophilic excipients in the formulation offers greater versatility for incorporating drugs across a range of lipophilicities.

While SDs did display significantly higher lipophilicity than Others, LBFs and SDs could not be separated in terms of this parameter. This reflects analysis by Ditzinger

et al. where 66% of SDs in literature displayed logP values of 2-6 (6). Previously, a logD cut off of ≤2.7 was suggested for SD over LBF formulation class suitability (121). However, our findings suggest that while lipophilicity provides potential to isolate drugs with potential for commercial success via LBF or SD delivery technologies, it does not differentiate between them. For example, earlier case studies of Kaletra®, and Norvir® containing highly lipophilic drugs (clogP ≥4.7) demonstrate that such drugs can be produced successfully as both LBFs and SDs. In these cases, despite high lipophilicity, the SD forms were ultimately more commercially favourable. While these provide just two examples, overall, these findings appear to challenge the commonly held belief that drugs with high logP values are more suited for LBFs and perhaps, begging the question if our rationale for assessing the utility of LBFs may be overly simplistic. As such, while previous research has demonstrated that the renowned ability of LBFs to eliminate the food effect does not always stand to scrutiny (92), the current results have also demonstrated that LBFs cannot be differentiated from SDs in terms of lipophilicity.

Aqueous Solubility (logS)

As expected, among the total dataset of drugs, aqueous solubility (expressed as logS) displayed a significantly lower solubility for both LBF and SD drugs versus Others. Interestingly, when excluding high solubility drugs from the dataset and reanalysed using only low solubility drugs, significances remained. This indicates that even within PWSD classes, LBF and SD technologies offer the opportunity to facilitate commercial development as oral drug products. In relating lipophilicity and hydrophilicity, Bergstrom et al. have previously suggested that a logP >3 is an indicator of reduced interaction with aqueous solvents (3). In this analysis, our mean

logP values for commercial LBFs (4.66) and SDs (4.16) were both found above this value. Such results are expected as both formulation technologies present a potential delivery solution for drugs encompassing the "poor solubility challenge".

Percentage Excreted Unchanged in Urine (%U)

Percentage drug excreted in urine also distinguished drugs suitable for both LBF and SD but not between the two delivery techniques. A significantly lower percentage of both LBF and SD drugs were excreted in urine compared to the Others dataset. This is not unexpected as drugs excreted in the urine unchanged are typically highly water soluble whereas PWSDs require metabolism into metabolites which are likely more polar and readily excreted (122). However, a range of factors may influence the predictive ability of this property, including need for a bioavailability factor for orally delivered drugs, coupled with the fact that that certain drugs or active metabolites may be excreted unchanged in bile not urine (91). This property demonstrated that SD and LBF drugs are less hydrophilic than Others, similar to our previous result of their higher lipophilicity and lower aqueous solubility.

Rotatable Bond Count (RB)

SD commercialised products displayed significantly higher mean RB count than both LBF and Others. Once again reflecting current trends in drug candidates, as bulk physical properties including MW and RB count have increased with time (79). This finding compliments previous observations that compounds exhibiting high amorphous stability contain higher numbers of RBs (123). Baird et al. have suggested that higher RB and molecular flexibility decreases probability of being incorporated

into an ordered crystalline structure (46), and demonstrated that both high MW and high RBs are indicative of higher GFA and lower crystallisation tendency (i.e. Class III GFA). Elsewhere, the number of RB, providing a measure of molecular flexibility, has been suggested by Kuentz et al. to positively influence the amorphous solubility advantage of a drug (86). Comparatively higher RB (e.g., 5-10) were indicative of suitability for a SD formulation approach, and at a mechanistic level this most likely reflects the ability of good glass forming drugs to display prolonged supersaturation, relative to poor glass former which are at greater risk of precipitation from supersaturated solutions. It is also noteworthy that molecular flexibility was not predictive of a LBF approach. Again, at a mechanistic level LBF increase drug concentrations via promotion of solubilisation in the intraluminal fluids and hence the ability of the inherent amorphous stability of the drug is not a considered to be a factor influencing performance.

Hydrogen Bond Acceptors (HBA)

HBA count was observed to be a property which distinguished between suitability of SD commercial drugs versus both LBFs and others, with a significantly higher mean HBA found for SD drugs (i.e., 6.87). The importance of HBA count is reflected in the fact that more than double (24%) of SD drugs had greater than 10 HBA compared to LBF drugs (10%). Furthermore, when comparing only low solubility drugs the significance of the differences between SD and both LBFs and Others was strengthened.

Hydrogen bonding interactions increase both stability and rigidity of the amorphous state by the formation of poorly packed aggregates which render crystal formation

increasingly difficult (123). Number of HBA has previously been significant in modelling both the potential for crystallisation of a drug, based on GFA class (124), as well as prediction of the solubility advantage for amorphous drugs (86). In the latter, the number of HBAs was the most important descriptor after MW in amorphous solubility advantage prediction. Additionally, hydrogen bonding between the API and polymer excipients is an important feature aiding polymers to inhibit drug crystallisation and promote amorphous stability. Hydrogen bonding between the two have been observed in dispersions displaying lower tendency and highest resistance to crystallisation (125, 126). Second, third and fourth generation SD utilise polymer carriers, either alone or in the presence of other polymers or surfactants (6). In this analysis, polymers were found to be the most widely used excipients in commercial SDs for both crystalline and amorphous based SD.

Hydrogen Bond Donors (HBD)

Both HBD and HBA counts are important with regard to Lipinski Rule-of-5 violations, amorphous stability and hydrogen bonding interactions between polymeric stabilisers and drugs. However, in this case, HBD was not found to be a property distinguishable between LBF, SD or Others in our analysis of the full datasets. However, when only low solubility drugs were analysed, a significant difference was observed between SD and Others. Previously, amorphous stability was found to be moderately correlated with the number of HBDs upon previous examination of a group of PWSDs (123) and positively correlated with MW ($r^2 = 0.70$), as discussed previously in this analysis to be influential. Thus, intensifying the significance of hydrogen bonding capacity in distinguishing suitability of drugs for SD commercial success.

Polar Surface Area (PSA)

The importance of hydrogen bonding capacity was once again reflected in the fact that PSA distinguished suitability of drugs for commercial SDs versus both LBFs and Others. Significantly higher mean values were found for the SD dataset (125.92 $Å^2$), versus LBF (79.68 $Å^2$) and Others (81.48 $Å^2$) which retained significance when only low solubility drugs were compared. The spread of values was also the smallest for SD drugs. Comparatively, drug development trends indicate the mean PSA of drugs has been increasingly significantly through the years (70, 79). However, it is important to bear in mind that correlation does not imply causation as in this case, the increasing prevalence of new drug candidates displaying higher PSA as well as increasing use of SD technologies could represent independent trends in both cases or reflect complementarity of both. PSA was previously determined a significant descriptor in *in silico* modelling long term amorphous stability (123) and amorphous solubility gain (86). For the later, the authors suggested a comparatively higher value for PSA as a property to prompt consideration for SD delivery. They found a range of 60-140 $Å^2$ being indicative of a high amorphous solubility gain. In our analysis, the mean PSA for SD commercial drugs was 125.9 $Å^2$, thus, within this range.

Lipinski Rule-of-5 Violations (Ro5)

We observed a significant association between drug group and prevalence of Ro5 violations. This 'drug likeness filter' states that, in general, an orally active drug has no more than one violation. Thus, in our analysis we used a cut-off of ≤1 (0 or 1) or >1 violation (2, 3 or 4). After this discrete numerical variable was recoded to a categorical variable, we observed both LBF and SD to be significantly different from Others in terms of Ro5 violations ($p^{**} < 0.01$, $p^{***} < 0.001$). As such, 30% of SD

and 18% of LBF commercial drugs displayed >1 violation compared to 6% of Others (Appendix 1 Table 1-2). Without question, the higher Ro5 violations observed mirrors the growing number of beyond Ro5 drugs candidates being produced in the search for biological selectivity for emerging biological targets. It has previously been observed that only approximately 50% of all drug targets appear accessible by compounds within the Ro5 chemical space (127). As such, extended Ro5 (eRo5) and beyond Ro5 (bRo5) compounds refer to those outside this defined chemical space (3). Perhaps suggestive that standard drug likeness filters may appear overly conservative as more and more non Ro5 compliant compounds reach commercial development. As mentioned previously, complementary use of formulation likeness filters may provide accurate predictions of formulation success for such troublesome drug candidates, as commercial success has been already demonstrated through LBF and SD approaches.

Dosage Strength (pDose and MDS)

Although, the LBF dataset demonstrated the lowest mean MDS (118.59mg) and the smallest first quartile value among the three groups, no significant differences were observed between the three groups. Conversely, upon comparison of only low solubility drugs, both LBF and SD drugs demonstrated significantly lower MDS compared to Others ($p^{**} < 0.01$, $p^{**} < 0.01$). Any lower dosage levels could refer to higher potency where smaller doses are required. While conversely PWSD not formulated by enabling formulations may require dosage increases to compensate for low bioavailability. A dose of <100 mg has previously been suggested as a significant factor to consider lipid-based drug delivery systems to dissolve the full dose. To overcome this perceived dose limitation LBF suspensions, along with the avocation of chase dosing (128) and use of ionic liquids have been suggested (129). Previously,

suitable drugs for LBF delivery have been proposed to be low dose drugs such as hormones, cytotoxic drugs or prolonged therapy drugs requiring dose titrations (35). Linking to this, two of the BDDCS class 1 drugs utilising LBFs commercially consisted of Vitamin D and its active metabolite with dosage levels in the microgram range (One-Alpha®, Thorens®, Uvedose®). Thus, dosage strength may also be a factor for previous observation that the second highest proportion of LBF commercial drugs are BDDCS class 1.

We also examined dosage strength in terms of pDose. When only low solubility drugs were analysed both LBF and SD drugs displayed significantly smaller doses compared to Others ($p^* < 0.05$, $p^* < 0.05$). This was somewhat unexpected as a stated advantage of SDs over LBFs is in general, the potential for much higher dosage levels, as high API-to-polymer ratios can offer higher drug loadings, echoing the commercial product Kaletra® resulting in a decreased pill burden. However, this could be affected by whether a crystalline or amorphous-based solid dispersion is produced. Instability of the amorphous form or presence/absence of polymers could alter drug loading capacities of amorphous-based solid dispersions.

Non-Significant Properties

No trends in pKa were established. However, a previous meta-analysis of 61 articles regarding supersaturating drug delivery systems including SD and LBFs between 2010-2015 revealed weakly acidic drugs demonstrated the highest improvement in the oral bioavailability-related parameters in comparison to weakly basic or neutral drugs (130). However, more extensive research is required as any effect of drug ionisation is difficult to analyse.

Properties of Drugs Commercialised via Both Bio-Enabling Formulation Technologies

As stated previously, four drugs have been commercially developed using both LBF and SD technologies. These drugs are Fenofibrate, Nimodipine, Ritonavir and Lopinavir. Two drugs displayed >1 Ro5 violation and all four were BDDCS class 2. Mean logP, clogP and $logD_{7.4}$ values for these drugs were high with all drugs displaying low aqueous solubility. With regard to T_m, only one drug, Lopinavir, had a T_m above the aforementioned cut off for LBFs of 150°C (174.5°C). Thus, it can be suggested that for a drug to act as a commercial candidate for success via both technologies it should display an intermediate T_m (e.g., ~150°C) to increase likely solubility in the lipid system. Three of the four drugs displayed ≥10 RBs and PSA >120 $Å^2$. Thus, while these properties reflect suitability for SDs, they do not, in practice, limit the commercial potential of drugs for success with LBFs. MW ranged from 360.83–720.946 g/mol, demonstrating the ability of both technologies to accommodate drugs with a wide range of MW. The average number of HBA and HBD were similar to our previous values and mean %U was low (1.58%). Overall, it appears clear from the current commercial portfolio of products, that PWSD displaying Ro5 violations, higher PSAs, a high RB count, mid-range T_m, high HBA and HBD count and a low %U, provide potential candidates for commercial development with both LBF and SD technologies. While in terms of drug properties which can distinguish between LBF and SD platforms in terms of commercial success, this review has demonstrated that drug MW, PSA, RB and HBA count show significant differences between current LBF and SD commercial products.

Summary of Findings

Prompted by the increasing application of DTs to provide property-based rules to formulation selection, this work examined physicochemical and molecular properties of the current commercial portfolio of drug products using LBF and SD formulations. A database of drugs commercially developed as LBFs and SDs was reviewed, prevalence of BDDCS class was determined and retrospective trends in drugs properties uncovered. It was established that drug properties could distinguish not only LBF and SD bio-enabled commercial drugs from Others but also distinguish between commercially successful LBF and SD drugs. The latter involved drug properties of MW, RB, HBA and PSA, indicating importance of size, molecular flexibility and hydrogen bonding capacity in formulation of SDs. In terms of well-established drug likeness filters, >1 violation of Lipinski's Ro5 was seen to be 5 and 3 times more prevalent for SD and LBF drugs, respectively, versus Others. While the T_m of 55% of commercial LBF drugs exceeded the often reported cut off of 150°C. A general trend toward increasing commercial development of SD formulations in recent years was observed. Encouragingly, many of the significant properties established reflect drug discovery trends of recent years, providing a positive outlook for potential of bio-enabling formulations to overcome solubility limitations. Furthermore, all drug properties included in the "Oral PhysChem Score" system i.e. MW, clogP, RB, solubility and PSA, indicative of bio-pharmaceutical performance of a drug, were found to be significant in this analysis (82).

This is not a definitive nor exhaustive list, drugs which do not fit some properties mentioned may be successfully developed in the future and certain properties not deemed significant do have their part to play. Moreover, as the numbers of drugs encompassing commercial LBF and SD products continues to grow, alterations to

these trends may develop as certain properties may emerge or become more influential over time. Additionally, it must also be acknowledged that other considerations such as drug efficacy, safety, instability or pharmaceutical commercial interest/priorities will also influence potential for commercial success. Utilizing and updating trends going forward can aid the continued growth of both LBF and SD commercial products. The results of this analysis outlines how retrospective analysis of commercialised drugs can lead to increased understanding of the properties of interest signally formulation suitability without requirements for manual testing. Results indicated that a limited number of drug properties, most of which are usually already available to the formulation scientist in early development, is likely sufficient to establish a statistically useful map of relationships between compound features and bio-enabling formulation categories. Demonstrating that retrospective assessments using existing knowledge available in early development and formulation likeness filters possess capacity to inform potential developability, either as a LBF or SD commercial product, based on previously successfully drug candidates and success stories over the last few decades. As such, if trends of increasing MW, lipophilic, flexible, beyond Ro5 candidates continue to stem from the discovery pipeline, the need for such bio-enabling formulations will also increase.





Bennett-Lenane,H. 2021. Computational pharmaceutics approaches to inform drug developability: focus on lipid-based formulations. PhD Thesis, University College Cork.

Please note that Chapter 2 (pp. 71-100) is unavailable due to a request by the author.

CORA Cork Open Research Archive http://cora.ucc.ie

Chapter 3: Machine Learning Methods for Prediction of Food Effects on Bioavailability: A Comparison of Support Vector Machines and Artificial Neural Networks

Harriet Bennett-Lenane, Brendan T. Griffin, Joseph P. O'Shea

School of Pharmacy, University College Cork, Cork, Ireland

Published in: *European Journal of Pharmaceutical Sciences (2021), 167.*

DOI: 10.1016/j.ejps.2021.106018

Abstract

Purpose: Despite countless advances in recent decades across various *in vitro*, *in vivo* and *in silico* tools, anticipation of whether a drug will show a human food effect (FE) remains challenging. One means to predict potential FE involves probing any dependence between FE and drug properties. Accordingly, this study explored the potential for two machine learning (ML) algorithms to predict likely FE.

Methods: Using a collated database of drugs licensed from 2016-2020, drugs were classified into three groups; positive, negative or no FE. Greater than 250 drug properties were predicted for each drug which were used to train predictive models using support vector machine (SVM) and artificial neural network (ANN) algorithms.

Results: When compared, ANN outperformed SVM for FE classification upon training (82%, 72%) and testing (72%, 69%). Both models demonstrated higher FE prediction accuracy than the biopharmaceutics classification system (BCS) (46%). This exploratory work provided new insights into the connection between FE and drug properties as the Octanol Water Partition Coefficient (S+logP), Number of Hydrogen Bond Donors (HBD), Topological Polar Surface Area (T_PSA) and Dose (mg) were all significant for prediction.

Conclusion: This study demonstrated the utility of ML to facilitate early anticipation of likely FE in pre-clinical development using four well-known drug properties.

Introduction

It is widely recognised that concomitant administration of oral dosage forms with food can alter drug pharmacokinetic profiles (92, 189, 190). As oral dosage forms are both widely and often chronically administered, understanding of the biological processes triggered by food consumption and its complex and drug-specific impact on oral bioavailability is vital. The numerous underlying mechanisms by which food exerts this effect on drug absorption include physiological changes in pH, gastric emptying times, fluid volumes, bile salt concentrations and intestinal enzyme activity, in addition to specific food effects including binding, metabolism or interference with transporters (92, 190, 191). The clinical consequences of these changes are assessed through comparison of pharmacokinetic parameters describing the rate and extent of bioavailability i.e. peak plasma concentration (C_{max}), time to peak plasma concentration (T_{max}) and area under the curve (AUC) in both the fed and fasted state (192). A food effect (FE) is defined as when the 90% confidence intervals for the ratio of population geometric means, based on log transformed data, for either $AUC_{0\rightarrow\infty}$ or C_{max} fall outside the 80–125% bioequivalence limits relative to the same formulation administered in the fasted stated (192). These FE studies are subject to stringent regulatory requirements (192, 193), and the consequences of food-mediated effects on bioavailability have been widely reported (190, 194-199).

Previous research that 40% of drugs licensed between 2010 and 2017 displayed significant FE (92) suggests that within the current drug development paradigm, anticipation of the impact of food on drug absorption is pertinent. Moreover, in addition to guiding the design of improved formulations that are FE resistant, this information is fundamental to optimise exposure of medicines with narrow therapeutic ranges in a clinical setting, to meet strict fed state bioequivalence study requirements

of international regulatory authorities and reduce costs associated with product failures due to variability in exposure (92, 95, 190, 194). Consequently, the ability to predict and anticipate FE is of immense value to drug development. To date, extensive mechanistic tools spanning a wide range *in vitro*, *in vivo* and *in silico* methods to predict FE have been described in literature (3, 200-203). Typically, drug performance under fasted and fed conditions are anticipated *in vitro* with dissolution tests in biorelevant media mimicking the human gastrointestinal environment (204-206), while *in vivo* predictions using various animal models (200, 202, 207), including canine and porcine examples have also been employed, along with drug solubility testing in aspirated intestinal fluids (201, 208). However, despite the varying levels of success achieved by these methods, the multifaceted factors associated with drug, meal type and physiological conditions mean that, as of yet, no universally comprehensive approach for FE prediction has been found and that all current models exhibit limitations.

Owing to the limitations of *in vitro* methods, and the challenges in performing *in vivo* pharmacokinetic studies, *in silico* methods have emerged as the "go-to" approaches for predictive biopharmaceutics (49). Drug classification tools comprising rules of thumb based on drug biopharmaceutical properties, and both the biopharmaceutics and biopharmaceutical drug disposition classification systems (BCS and BDDCS), provide simple guides to anticipate FE based upon related drug properties (200, 209, 210). While such approaches provide a readily accessible prediction of likely FE, the relatively low accuracy and precision of these predictions has necessitated the development of mechanistic and data driven *in silico* models of FE (196). Physiologically-based pharmacokinetic (PBPK) models lie at the centre of this diversifying modelling and simulation (M&S) toolbox, and have achieved

increasingly accurate predictions of FE (61, 153, 169, 211). Using mathematical equations to model physiological processes and anatomical parameters, these compartment-based absorption models mechanistically simulate a drugs plasma concentration-time profile in the fasted and fed states and can be applied in both pre-clinical studies with further potential application as decision-making aids for regulatory agencies (212). While PBPK models require comprehensive data about the physiological, biochemical, and physicochemical processes that occur, data-driven modelling tools which establish statistical relationships between FE and drug molecular properties, signify complementary additions to this ever-expanding M&S toolbox without the need for such. To date several attempts, spanning the last three decades, have been made to probe dependence between FE and drug physicochemical or molecular properties. These include linear correlations of AUC ratio with individual drug properties (213) and a tool for computational biopharmaceutical profiling of ligands, based on predicted increases in fed state simulated intestinal fluid (FeSSIF) solubility to signal a positive FE (3). Finally, qualitative computational models predicting if a drug displays a positive, negative, or no FE according to the $AUC_{fed/fasted}$ ratio, have also been published, using logistic regression (214) and more recently decision tree analysis (215).

Consequently, while modelling efforts to correlate the effect of food on AUC with drug properties exist, the application of multiple machine learning (ML) classification algorithms to predict FE remains unexplored. Support vector machine (SVM) classification and artificial neural network (ANN) algorithms have been gaining interest across various facets of drug design and development, supporting streamlined and decision-based pre-clinical testing (124, 167, 216-220). SVM is a pattern recognition method widely used in data mining which works by finding an optimal

separation line (hyperplane) which accurately separates and maximises the margins between two or more classes. Separation of non-linear data is also achievable through kernel functions which map the original data to a higher-dimensional "feature space" facilitating linear separation, as previously described (173, 183). While as a ML algorithm, ANN detects complex non-linear relationships between datasets, by mimicking basic human biological information processing methods. Adopting a general structure consisting of an input layer, hidden layer(s), output layer and using activation functions and connection weights, nodes (artificial neurons) from the input layer send data to the hidden and output layers through weighted connections ("synapses") to predict Y, as described in detail previously (172, 185).

Using these ML algorithms, the broad objective of this work sought to develop an *in silico* model which could identify important drug properties for accurate FE prediction. Here, drugs were classified as displaying a positive, negative or no FE according to a change in the extent of drug absorption (AUC) in the fed versus fasted states. Using a collated database of newly licensed drugs from 2016-2020, both SVM and ANN were employed to explore any relationship between the three FE classes and drug properties. The study design facilitated investigation into the prevalence of FE among drugs licensed in the last 5 years, while also assessing if either ML algorithm or the BCS classification tool provided the highest prediction accuracy for this dataset. Accordingly, this study provides the first evidence of the capacity of ANN and SVM to facilitate early anticipation of likely FE.

Methods

Database Collation

The drug database used in this study was obtained from drug products licensed in the European Medicines Agency (EMA) and Food and Drug Authority (FDA) between January 2016 and December 2020. The database of original new drug applications (NDA's) on the FDA website was searched by month from January 2016 to December 2020 along with the European Public Assessment Report (EPAR) of the EMA licensed products each year from 2016-2020. Exclusion criteria were any non-oral product, any biological product, any modified release product, any product where no FE information was available, any generic of a previously authorised product or any authorisation submission referring purely to changes of product indication. The products eligible for analysis were new molecular entities (NME), new combination products and reformulations of products which have been previously marketed. General information recorded included year of licensing, generic name, commercial name and any label restriction of the drug administration regarding food. Information regarding FE on absorption was obtained from the product EPAR or FDA label for each product where the ratios of $AUC_{fed/fasted}$ were recorded (Appendix 2 Table 2-1). In cases where the documentation stated a product showed no change or a non-significant change in $AUC_{fed/fasted}$, with no values or ratios provided, a value of 1 was assigned. As FE information was obtained from regulatory submissions only, variables related to meal composition that might affect the interpretation of results were minimised.

Food Effect Classification

In this study drugs were predicted to belong to one of 3 FE classes designated according to the change in the extent of drug absorption in the fed versus fasted state ($AUC_{fed/fasted}$) alone, with "positive" referring to significantly increased and "negative" referring to significantly decreased extent of drug absorption in the fed state. This lone classification parameter was chosen as information regarding C_{max} was more frequently omitted from EPARs and FDA drug labels. Previous comparative studies also used this classification criterion, and the toxicity, efficacy and clinical significance of numerous drugs including those which are chronically dosed, correlates better with total exposure (AUC) than C_{max} (92, 213). The FE ratio ($AUC_{fed/fasted}$) was obtained for all drugs in the final dataset (141 drugs). A positive or negative FE was considered significant if the ratio fell outside 80–125% in reference to the currently accepted 90% CI for the ratio of population geometric means between fed and fasted treatments for concluding a lack of food-effect (192). Drugs with $AUC_{fed/fasted}$ >1.25 were classified with a "positive FE", $AUC_{fed/fasted}$ between 0.8-1.25 were deemed to have "no FE" and $AUC_{fed/fasted}$ <0.8 a "negative FE". The final database of 141 drug compounds consisted of 44 Positive FE, 80 No FE and 17 Negative FE drugs.

Compilation of Physicochemical Descriptors

More than 250 descriptors for each drug were obtained from ADMET Predictor 9.5 (Simulations Plus, USA). Molecular structures were acquired as smiles from PubChem and used as inputs for the ADMET Predictor software (Version 9.5, Simulations Plus, California, USA) to calculate the molecular descriptors. A conscious effort was made to ensure repetition of drug properties found to be significantly

correlated with FE in previously published reports, to facilitate comparisons (213-215). The drug dosage strength used in the FE bioequivalence study for each drug was obtained from the EPAR or FDA product label respectively. A maximum absorbable dose (MAD) was calculated using a predicted fasted state simulated intestinal fluid (FaSSIF) solubility, again employing ADMET Predictor, using the equation described previously (221). The dose/solubility ratio was calculated using the dose as described above divided by an aqueous drug solubility predicted from ADMET Predictor.

Statistical Analysis

Prior to employment of ML, to analyse any linear univariate correlations between FE classification and selected drug properties, a stepwise statistical analysis approach, as described previously (222), was adopted using SPSS (IBM Corporation, US) on the full drug database. It was hoped that these preliminary results could inform which properties may be significant for ML prediction. In brief, frequency distributions of the variables were graphed for each FE classification (positive, negative, no FE) and normality was checked visually with Q-Q and P-P plots. Ratios of samples sizes between the 3 groups were obtained. Variances of the datasets were analysed and compared to Levene's Test for Equality of Variances. A p-value <0.05 indicated a violation of equal variance. The null hypotheses were that no statistical differences were seen in a drug property between drug classes. Three separate comparisons were made i.e., negative versus no FE, no FE versus positive, positive versus negative. Comparison between groups were made using the t-test, Welch's test or Bootstrap independent samples test (5000 samples.) A p-value of 0.05 was used as the significance level for all tests. Boxplots were produced to facilitate visual interpretation of the data again using SPSS (IBM Corporation, US) and descriptive

statistics including median, mean, standard deviation of mean, Q1, Q3, minimum, maximum and variance were obtained for each drug property for the 3 groups. The properties selected were S+logP (Octanol Water Partition Coefficient), HBD (Number of Hydrogen Bond Donors), HBA (Number of Hydrogen Bond Acceptors), T_PSA (Topological Polar Surface Area), Dose (mg), S+logD (Partition Coefficient pH 7.4), S+Sw (Aqueous Solubility), MAD (Maximum Absorbable Dose), D/S (Dose/Solubility Ratio), MWt (Molecular Weight) and RB (Rotatable Bonds).

BCS Classification

The BCS class of the drugs studied were obtained where available from the EPAR, FDA label or from literature. Fleisher et al. (191) previously described the general trend of FE on drug absorption (AUC) based on BCS classification where, BCS Class I compounds are likely to have no FE; BCS Class II compounds are likely to have a positive effect; BCS Class III compounds are likely to exhibit a negative effect, while there is insufficient evidence for any clear identifiable trends for BCS Class IV compounds. Further separation of drugs within BCS classes as previously recognised (201), was not conducted to facilitate comparisons to a previously published analysis (214). Accordingly, the database was classified into these 3 FE categories, while BCS class 4 drugs were disregarded for this portion of the analysis.

Machine Learning/Model Development

FE classifications were predicted using two ML algorithms, ANN and SVM. To facilitate direct comparison of the predictive power of both algorithms the same training:test split was used. Principal component analysis (PCA) using the Unscrambler XI (Camo Analytics, US) was applied for a randomised assignment of

training:test data. Training set criteria was that it covered the chemical space of the test set and ensured an almost equal representation of positive, negative and no FE drugs in the training set to avoid any potential for classification bias. Such imbalance in datasets has proved to be a widely reported obstacle to ML classification problems in the past (173). In accordance with previous reports of classification prediction using ML algorithms (124, 220), initial variable reduction was conducted. Either a one-way ANOVA analysis with Tukey Multiple comparisons test (parametric) or Kruskal-Wallis analysis with Dunn's multiple comparison test (non-parametric) were applied to the training data. Variables with a p-value less than 0.05 for at least one class pair in the respective post tests were highlighted for further investigation. A correlation analysis of these identified variables was carried out with highly correlated variables clustered into the same group and the most significant variable of the group chosen for inclusion in the model development. Final models as well as BCS predictions were compared in terms of various accuracy statistics including, overall accuracy of prediction, as well as sensitivity, precision, specificity and Matthews correlation coefficient (MCC) for each FE class, as previously defined (223). TP, FP, TN and FN refer to true positive, false positive, true negative and false negative results respectively. MCC was previously suggested as a reliable statistical metric for ML performance quality evaluation (224). A high MCC score (close to 1) is only achieved if the model obtained good results in all four confusion matrix metrics (TP, FP, TN, FN).

$$\text{Overall Accuracy: } \frac{\sum_{i=1}^{M} TP_i}{\sum_{i=1}^{M} TP_i + FN_i} \; x\; 100\%$$

$$\text{Sensitivity: } \frac{TP}{TP + FN} \; x\; 100\%$$

$$\text{Precision: } \frac{TP}{TP + FP}\ x\ 100\%$$

$$\text{Specificity: } \frac{TN}{TN + FP}\ x\ 100\%$$

$$\text{MCC: } \frac{(TP\ x\ TN) - (FP\ x\ FN)}{\sqrt{(TP + FN)\ x\ (TP + FP)\ x\ (TN + FN)\ x\ (TN + FP)}}$$

Support Vector Machine Classification

A SVM algorithm was used to build a classification model using Unscrambler XI (Camo Analytics). Resulting variables from the initial variable reduction protocol were mean centred, de-identified and standardized through scaling by standard deviation. Variables were added one-by-one to assess which combinations produced the highest accuracy in both the training and test sets. C-SVC was used, where as part of the optimisation procedure the optimum values of key parameters of the regularization parameter C and gamma were sought from a grid search, performed across 10 orders of magnitude in logarithmic scale. In this grid search these key parameters were varied systematically to monitor which combination provided the highest classification accuracy in training and cross validation. This grid search was repeated for each variable combination using various kernel types (linear, polynomial, radial basis function and sigmoid) until the combination of kernel, C, gamma and input variables resulting in the best classification performance was obtained. The variables with the highest contribution to SVM classification were interpreted from the loadings plot of the PCA analysis which mapped the multidimensional data into a two-dimensional space.

Artificial Neural Networks

Multilayer perceptron artificial neural networks (MLP-ANN) were produced using SPSS Statistics (Version 26, IBM Corporation, US). The output layer consisted of three responses/categories: positive, negative or no FE. After initial variable reduction, as described above, the remaining significant variables were rescaled through standardisation where values were converted to their z-scores. Hyperbolic tangent was chosen as the activation function for the hidden layer, while an identity output function was used as the output layer activation function (225). Supervised learning using the scaled conjugate gradient algorithm was chosen (226). Batch training was selected. Topologies with only one hidden layer were considered. The optimum number of neurons in the hidden layer was identified following a systematic trial-and-error approach were the number of neurons in the hidden layer were manually altered between 2 and 50, with runs being performed in triplicate. The optimal network size was chosen from the solution which resulted in the highest prediction accuracy in the training and test sets. All combinations of the significant variables were tested to ascertain which combination produced the highest prediction accuracy. This procedure was repeated until no improvement in model performance was observed. The relative contributions of each variable in the final network were elucidated from the normalised importance chart.

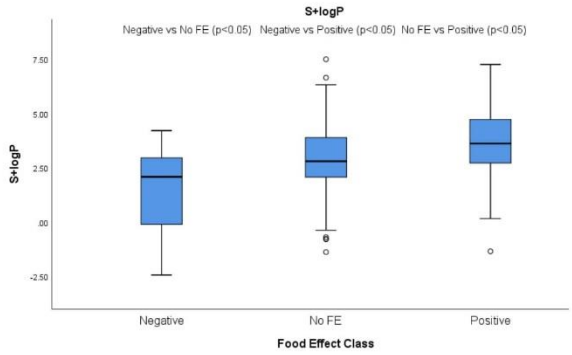
Results

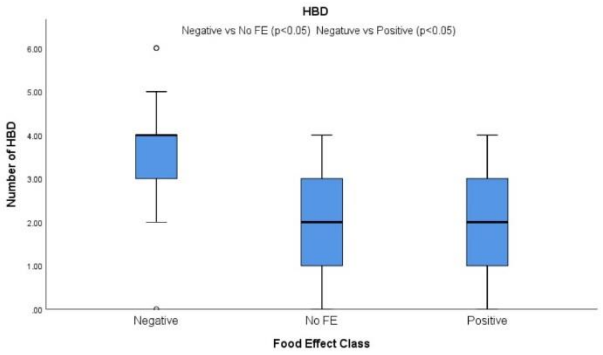
Analysis of Trends in Physicochemical Descriptors between Food Effect Classifications

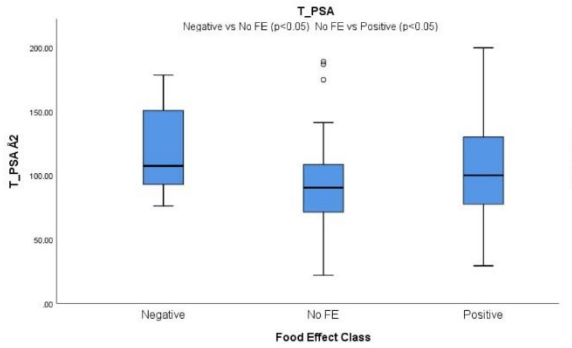
A selection of well-known molecular and physicochemical properties of drugs licensed in the last five years were statistically compared with respect to the three FE classes. It was anticipated that as a result of this preliminary analysis of the entire dataset, any significant differences in properties between classes could inform which properties may be significant for subsequent ML prediction. Tabular results of the statistical analysis are shown in Appendix 2 Table 2-2. A visual representation of properties which demonstrated significant differences is illustrated in Figure 3-1.

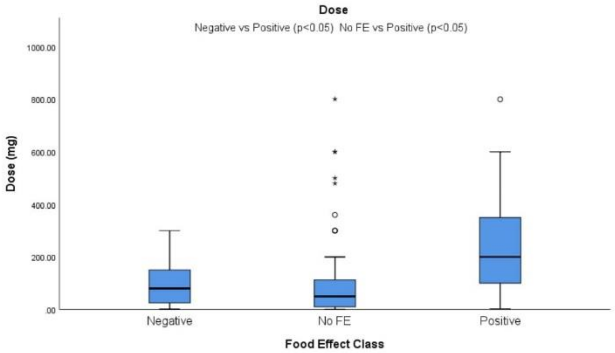
Upon statistical analysis, the properties of S+logP, HBD, T_PSA, Dose, S+Sw, S+logD, MWt, D/S and MAD were all significantly different for at least one pairing i.e., negative versus positive, negative versus no FE or positive versus no FE. Of these, 6 properties, namely S+logP, HBD, Dose, S+Sw, D/S and S+logD were found to be significantly different between drugs classified as having either a positive or negative FE. In particular, the mean S+logP value was observed to be almost 2.5 times greater for positive FE drugs. Conversely, it was observed that drugs with a positive FE could be differentiated from drugs which displayed no FE in terms of the 8 properties of S+logP, T_PSA, Dose, S+Sw, S+logD, MWt, D/S and MAD. Where the median dosage strength for positive FE drugs was four times higher than that of drugs with no FE (200mg vs. 50mg). Finally, drugs classified as either displaying a negative or no FE appeared more difficult to differentiate as only three properties were found to be statistically different for this pairwise comparison, namely S+logP, HBD and T_PSA, where the median number of HBD for negative FE drugs was double that of no FE (4 vs 2). Overall, it was observed that only one property, S+logP, was found to be

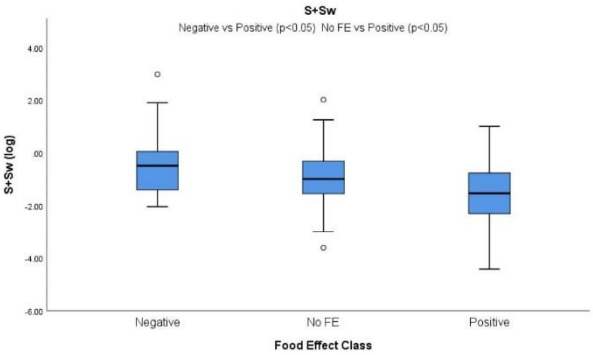
significantly different between all three classification groups (Figure 3-1). Furthermore, the properties of HBA and RB were not found to be significantly different between any of the three pairwise comparisons.

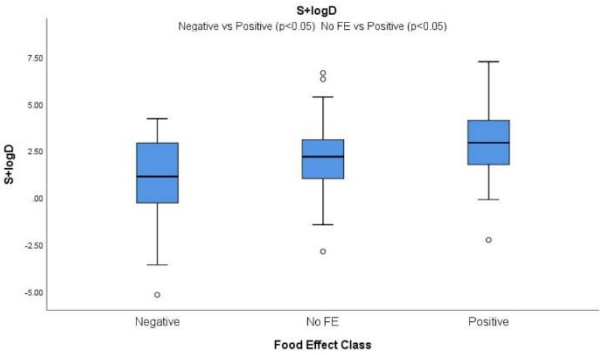

S+logP
Negative vs No FE (p<0.05) Negative vs Positive (p<0.05) No FE vs Positive (p<0.05)
S+logP
Negative
No FE
Positive
Food Effect Class

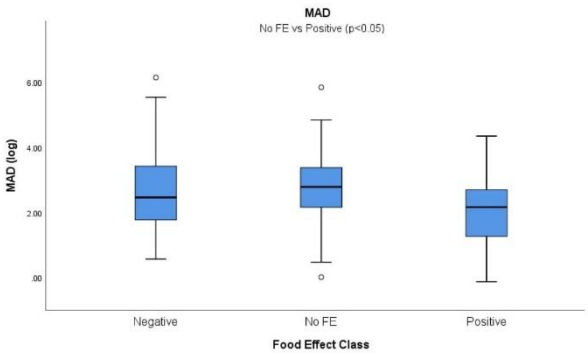

HBD
Negative vs No FE (p<0.05) Negatuve vs Positive (p<0.05)
Number of HBD
Negative
No FE
Positive
Food Effect Class

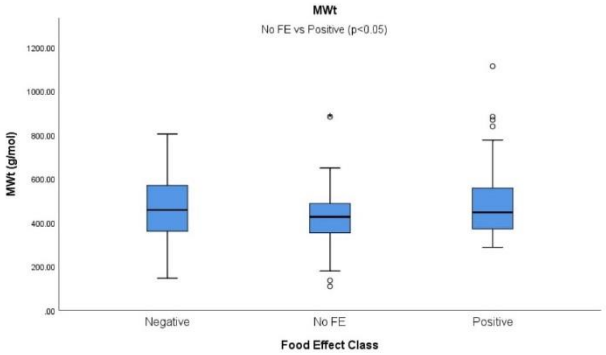

T_PSA
Negative vs No FE (p<0.05) No FE vs Positive (p<0.05)
T_PSA Å2
Negative
No FE
Positive
Food Effect Class


Dose
Negative vs Positive (p<0.05) No FE vs Positive (p<0.05)
Dose (mg)
Negative
No FE
Positive
Food Effect Class


S+Sw
Negative vs Positive (p<0.05) No FE vs Positive (p<0.05)
S+Sw (log)
Negative
No FE
Positive
Food Effect Class


S+logD
Negative vs Positive (p<0.05) No FE vs Positive (p<0.05)
S+logD
Negative
No FE
Positive
Food Effect Class


MAD
No FE vs Positive (p<0.05)
MAD (log)
Negative
No FE
Positive
Food Effect Class


MWt
No FE vs Positive (p<0.05)
MWt (g/mol)
Negative
No FE
Positive
Food Effect Class

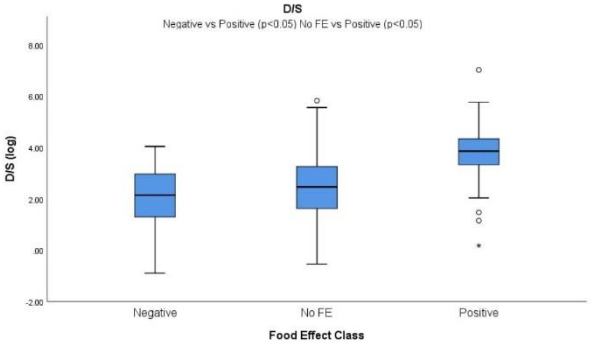

D/S
Negative vs Positive (p<0.05) No FE vs Positive (p<0.05)
D/S (log)
Negative
No FE
Positive
Food Effect Class

Figure 3-1*:* Visual representation of the statistically significant differences found between Positive, Negative and No Food Effect classes as part of the preliminary statistical analysis. p-values for the statistically significant pairwise comparisons are shown. The dark line in the middle of the boxes is the median. The bottom and top of the box indicates the 25th (Q1) and 75th percentile (Q3). The T-bars are inner fences/whiskers which extend to 1.5 times the box height. The circles are outliers that do not fall in the inner fences. The asterisks are extreme outliers which have values greater than three times the height of the boxes.

Food Effect Prediction using BCS Classification

The dataset collated in this study consisted of 61/141 (43%) drugs displaying either a significant positive or negative FE. Previous reports indicate that BCS classifications (BCS Class I, II, III), using dose/solubility ratio and extent of absorption, can aid identification of likely FE, and such an approach was investigated to facilitate comparison of this simple classification rule of thumb versus data-driven ML techniques (191). In this study, using classifications based on the BCS, a poor overall accuracy of prediction (46%) was obtained. While the sensitivity of the method was acceptable for positive (87%) and negative (69%) FE drugs, relating to lower numbers of false negative classifications, the same could not be said for the no FE class. There only 22% of drugs found to display no FE were BCS class I (Figure 3-2). While in terms of precision i.e., positive prediction rate, using the BCS tool, poor results were seen for positive (41%) and negative FE (35%) drugs with a poor rate of specificity also seen for the positive FE group (49%). Finally, in terms of MCC, scores close to zero of 0.3 (positive), 0.2 (negative) and 0.2 (no FE) respectively were calculated (Figure 3-2). Therefore, these results suggest that use of the BCS tool to predict FE classification results in high relative amounts of either false positive or false negative predictions.

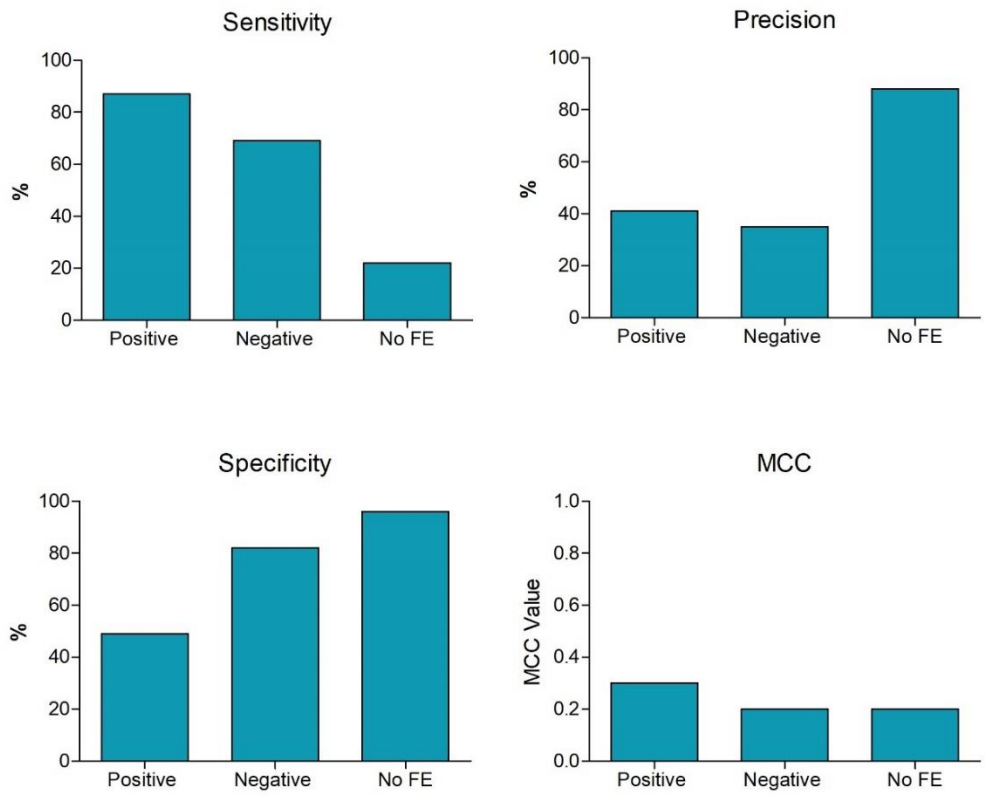


Figure 3-2: Visual representation of the sensitivity, precision, specificity, and Matthew's correlation coefficient (MCC) scores of the classification predictions (positive, negative and no Food Effect (FE)) using the biopharmaceutics classification system (BCS) criteria where an overall accuracy of 46% was achieved for the dataset.

Applying ANN and SVM to Predict Food Effect Classification

Two ML algorithms, ANN and SVM were employed to qualitatively predict if a drug displayed positive, negative or no FE (Table 3-1). Using the SVM algorithm, the optimum model required 6 drug properties for prediction, namely S+logP, HBD, T_PSA, dose, logarithm of the air-water partition coefficient (Henry's Law Constant at 25°C) (logHLC) and population average number of protons available for hydrogen bonding divided by the number of non-hydrogen atoms (F_HBP). Using the radial basis function (RBF) kernel which outperformed the other kernel functions tested, results demonstrated that the SVM model could correctly predict FE classification of

the training and set sets with 72% and 69% overall accuracy respectively (Figure 3-3). No FE drugs were predicted with the highest sensitivity upon both training (86%) and testing (71%) (Appendix 2 Figure 2-1), demonstrating better or comparative performance to ANN in these cases. Conversely, it was observed that the SVM algorithm was less successful in differentiating drugs with a negative FE compared to the ANN model described below (59%). In terms of precision, positive prediction rates of 79%, 91% and 60% were observed for the positive, negative and no FE groups (Figure 3-3). The SVM model showed the highest specificity in negative FE prediction (98%) where the lowest specificity was calculated for no FE drugs (69%). Intermediate MCC scores of 0.6, 0.7 and 0.5 were observed for the positive, negative and no FE classes respectively. Upon interpretation of the PCA correlations loading plot, HBD was the most influential property for classification prediction using this SVM-based model.

ANN model development resulted in an optimum three-layer feed forward network denoted MLP 4-13-3 (Figure 3-4). This network consisted of a single input layer with four descriptors, octanol-water partition coefficient (S+logP), number of hydrogen bond donors (HBD), Topological Polar Surface Area (T_PSA) and Dose (mg), a single hidden layer with 13 nodes and an output layer with three output variables representing the 3 possible FE classifications (positive, negative, no FE). This network achieved high overall prediction accuracy in both the training (82%) and test sets (72%) (Figure 3-3). In terms of the individual classes, in contrast to the SVM model, drugs displaying positive FE could be distinguished the easiest displaying the highest sensitivity rates in both the training (91%) and test sets (73%), while the lowest sensitivity was seen for the no FE group (76% training, 71% test). For the positive, negative and no FE groups respectively, the precision of the predictions were 83%, 93% and 73%,

specificity was 90%, 98% and 85% and higher MCC scores of 0.8, 0.8 and 0.6 closer to 1 were obtained (Figure 3-3). From the normalised importance chart, the most important property for prediction was S+logP, followed by T_PSA, HBD and Dose. However, all properties displayed over 67% importance to prediction.

Overall, relatively similar overall prediction accuracy was achieved using both ML algorithms, as ANN marginally outperformed SVM. The proposed ANN network demonstrated equivalent or higher sensitivity, precision, and specificity statistics for all but one metric for the three FE classes. In terms of MCC, which have been reported as a more reliable overall performance evaluator, comparatively higher scores were observed for the ANN model. Resultantly, considering these superior and more consistent classification results along with the requirement for less input descriptors, it was concluded that the ANN algorithm produced the more robust model for this dataset.

Table 3-1: Overview of the SVM and ANN machine learning models produced in this study, detailing the inputs, model architecture and the comparative overall accuracies upon training and testing.

Model Type	Input Properties Used	Architecture	Overall Accuracy Training Set	Overall Accuracy Test Set
SVM	Dose, HBD, F_HBP, S+logP, T_PSA, logHLC	Kernel: RBF Gamma: 0.01 C value: 16.68	72%	69%
ANN	Dose, HBD, S+logP, T_PSA	1 hidden layer 13 hidden nodes	82%	72%

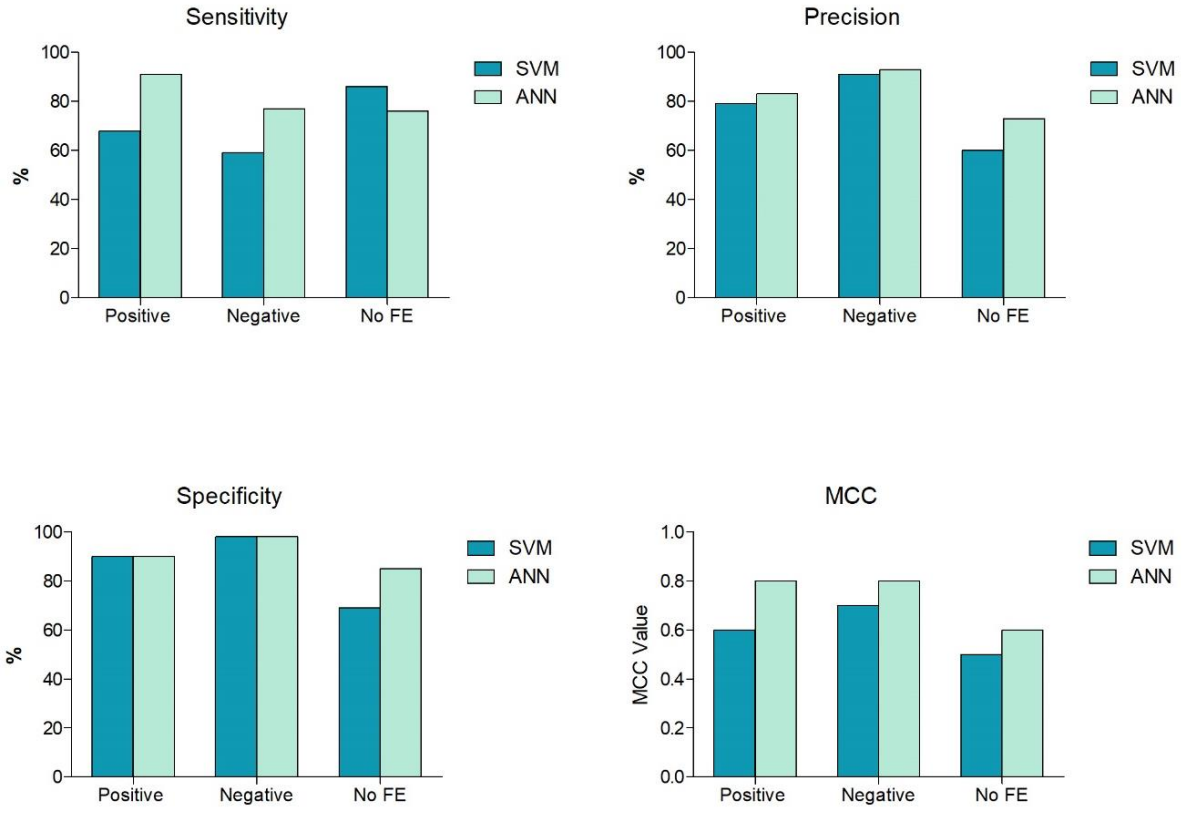


Figure 3-3: Visual comparison of the sensitivity, precision, specificity and Matthew's correlation coefficient (MCC) performance metrics calculated for the optimum support vector machines (SVM) and artificial neural networks (ANN) models produced in this study to predict food effect (FE) classification.

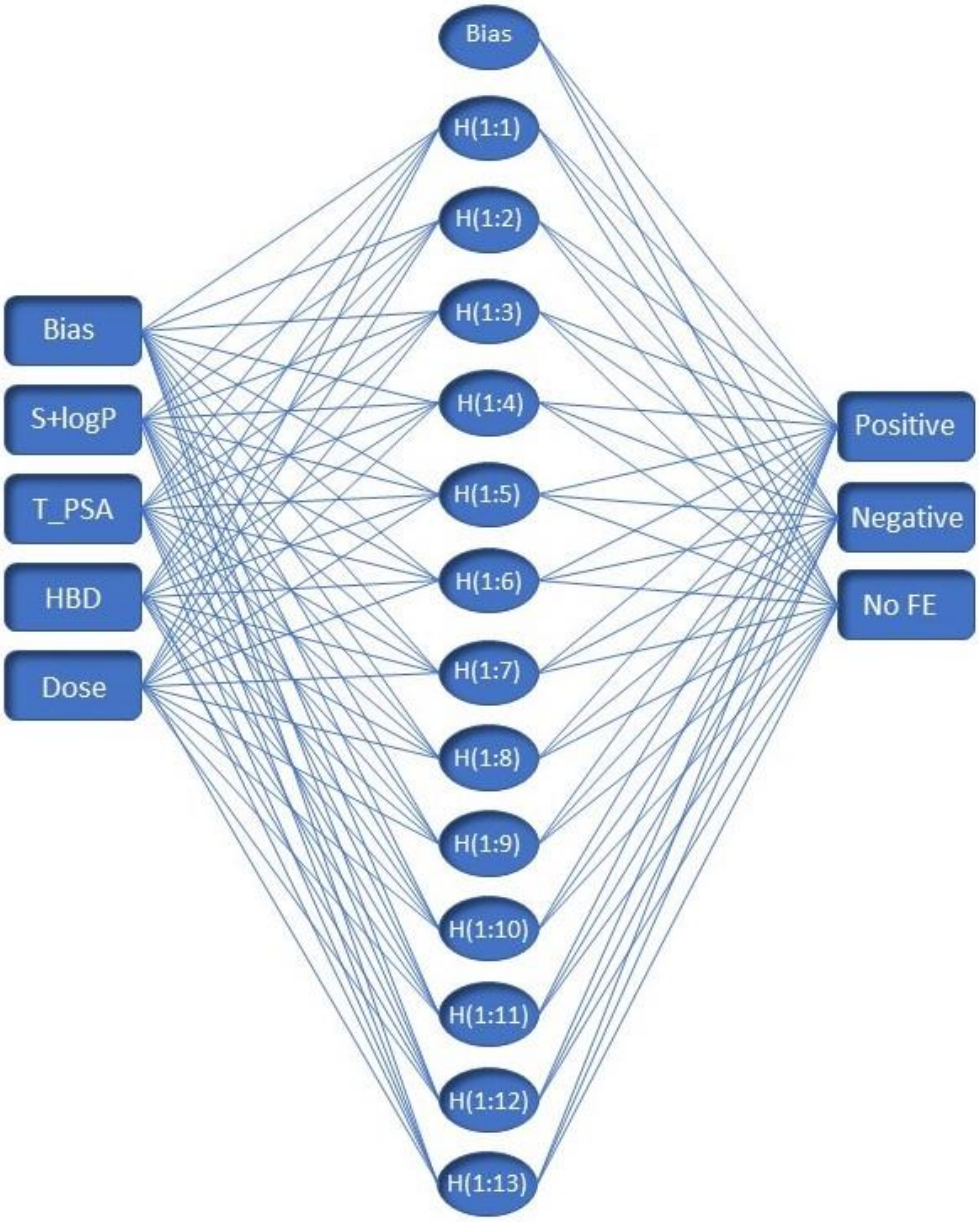


Figure 3-4: Schematic representation of the optimum multilayer perceptron (MLP) 4-13-3 artificial neural network (ANN) produced which outperformed support vector machines (SVM) for the food effect (FE) classification of drugs licensed from 2016-2020. H refers to hidden layer node.

Discussion

The significant effects of concomitant food intake on the pharmacokinetics of many drugs highlights the importance of FE predictions. The capability to predict FE early in the drug development process can potentially expedite progression of drug products through pipelines and streamline the regulatory approval process through identifying where clinical FE studies may or may not be required. As a result, the earliest possible elucidation of important drug properties which indicate a likely FE is vital, as it was seen in this study that 43% of drugs licensed in the last 5 years reported either a significant positive or negative FE. While various efforts, dating as far back as 1996 (227) have been made to correlate FE with drug properties, the purpose of this study was to investigate if modern ML capabilities could predict the FE classification of drugs. As the dataset employed embodied drugs brought to market in the last 5 years, which are largely diverse in terms of their structural, physicochemical, and pharmacokinetic properties, it was analysed if these ML algorithms could yield predictive tools relevant for the contemporary drug development landscape.

Our study demonstrated the capability of ML algorithms to predict FE classification using drug descriptors. Upon comparison of the two ML algorithms, results suggested that while both approaches demonstrated capacity for prediction, ANN outperformed SVM. The optimum ANN, containing 1 hidden layer of 13 nodes and using 4 input properties (S+logP, T_ PSA, HBD, Dose), yielded strong overall accuracy in classification for both training (82%) and test sets (72%). Comparatively, the SVM model which employed 6 input properties (S+logP, T_PSA, HBD, Dose, F_HBP, logHLC) for prediction, did also demonstrate relatively good accuracy upon training (72%) and testing (69%). However, in terms of the performance metrics analysed i.e., precision, sensitivity, specificity and MCC, the ANN model proved either superior or

equivalent to SVM in 11/12 cases, including MCC scores closer to 1 for all three classes. The exact reason for the improvement in performance metrics and requirement for less inputs for ANN, while difficult to pinpoint, may be attributable to the varying mathematical approaches used by both methods to classify this non-linearly separable dataset and map the data to higher dimensional spaces, i.e. kernel tricks (SVM) and non-linear activation functions (ANN) (184, 185, 223, 228). While the comparative accuracy of SVM versus ANN modelling is dependent on the specific dataset involved, the easily interpretable, readily obtainable, and widely recognisable nature of the four properties used in the ANN model to formulation scientists expands its applicability in a preclinical setting, as at this time only limited resources and preliminary information are available regarding a model compound. Accordingly, this study supports the application of ML algorithms, in particular ANN, to provide accessible tools for FE prediction of newly licensed drugs.

Our models identified key physicochemical and molecular properties which contribute to the classification of newly licensed drugs according to the likely effect of food on extent of drug absorption. While the ANN model found four properties of S+logP, Dose, HBD and T_PSA to be noteworthy, using SVM the properties of logHLC and F_HBP were also required. The inclusion of S+logP as the most important property for ANN prediction matches observations from our preliminary statistical analysis. There S+logP, a widely used drug parameter used in PBPK modelling (229), was the only property significantly different between all three pairwise comparisons (Figure 3-1). While another type of partition coefficient, logHLC was also significant for SVM modelling, S+logP or increasing drug lipophilicity in particular, has previously been linked with increased susceptibility to positive FE (3, 213, 230), through the increased dissolution and solubilisation effects of food for lipophilic drugs. This is a partition

property, while a hypothesis for its significance may relate to ingestion of fatty foods resulting in greater partitioning of lipophilic drugs between the digestion phases which could be related to the kinetics of partition between octanol and water. Using a cohort of pre-2005 drugs, Singh et al. found a significant positive correlation between AUC ratio and logP, where the positive effect of food was more pronounced for lipophilic drugs (213). In agreement, in this study, we observed that 41% of drugs with a positive FE had S+logP values >4, with 84% >2. Gu et al. and Omachi et al. have also successfully utilised logP or the closely related logD as a critical parameter to predict FE on AUC using both logistic regression and decision tree analysis (214, 215). Nevertheless, while the significance of S+logP in our model suggests the importance of a lipophilicity indicator to predict FE on AUC, caution has previously been advised in terms of using this blanket generalisation of likely positive FE for all PWSD as exceptions exist (213). In addition to its importance as a partition property, the requirement for S+logP may also reflect its significance to predict other biopharmaceutical properties such as membrane permeability, as it is possible that some dietary foods may affect permeability, however, it is unlikely that this would be unique to certain drugs. Lipophilicity has previously been correlated with permeability measurements for compounds which are passively transported (72, 74, 75, 154, 231).

In this study S+logP alone did not prove sufficient for accurate prediction of FE, as other properties including HBD and T_PSA were found in the final models. While these were not previously significant parameters for FE classification using logistic regression (214), using this contemporary cohort of drugs ML identified their predictive abilities. Both parameters, have been formerly associated with membrane permeability (72), as poor permeation is suggested to be likely with drugs of PSA >140Å (232, 233), where T_PSA is the commonly used PSA descriptor (3), and an

excessive number of HBD groups impairs permeability across a membrane bilayer (76, 155, 234), as specified by a cut-off of 5 in Lipinski's rule-of-5 developability criteria (72). Both parameters were used previously along with logP, to predict human jejunal permeability (P_{eff}) using PLS (154). It could be suggested that the role of both T_PSA and HBD in the ML models was to aid identification of drugs with negative FE as HBD and T_PSA demonstrated the highest mean values for the negative FE class upon preliminarily statistical analysis (Figure 3-1). This finding correlates well to the Fleisher et al. summary of FE prediction based on BCS class where Class III drugs, of poor permeability and good solubility, are likely to exhibit negative FE (191). Such negative FE were previously associated with highly hydrophilic drugs displaying a narrow window of absorption (203, 214).

Dose was also a significant parameter in classify drugs by FE on AUC. Upon early statistical analysis, it was observed that over 18% of drugs displaying a positive FE used a dose >500 mg in their respective pharmacokinetic studies, compared to 0% and 6.25% of negative and no FE drugs respectively, with 59% of this positive class using a dose >200 mg. While the exact reason for this substantial difference in dosage strength between classes is unknown, previously a logistic regression model found dose number and MAD to be significant for FE classification (214). However, when tested neither were found to be significant in our ML models. In addition, the dose/solubility ratio of a drug was previously significantly correlated with AUC ratio (213), however, in our analysis neither this ratio nor S+Sw, despite displaying significant differences between classes in our preliminary statistics, were important for ML prediction. It could be hypothesised that this dose parameter aided the prediction of positive FE drugs due to its ability to differentiate positive FE drugs from both negative and no FE classes in our initial statistical analysis (Figure 3-1). Overall,

the properties most significant for FE classification demonstrate the importance of both solubility and permeability for FE prediction. These properties reflect widely known drug-likeness filters, drug classification systems and properties used in previous models for FE prediction. It is likely that any differences in significant properties compared with previous publications reflect the different datasets of drugs used to build the respective models, as this study reflects the most contemporary drug products licensed in the last 5 years.

As previously stated, in terms of accuracy the ANN model outperformed SVM for FE classification. However, when compared to predictions using BCS class, both ML models performed strongly as the overall BCS accuracy (46%) was substantially weaker than that of the ML algorithms. BCS classifications appeared inadequate compared to ML in terms of consistency in precision, sensitivity, specificity and MCC scores. Owing in part to a large number of false positive predictions (58 drugs), this BCS accuracy of 46% is substantially lower than its previous 67% accuracy in classifying FE for a database of pre-2007 drugs (214). Possible reasons for this inaccuracy compared to pre-2007 drugs are unclear but may reflect that the drugs licensed between 2016-2020 represent a different chemical design space. While direct accuracy comparisons with the ML models are not achievable as no trends in FE have been suggested for BCS class IV drugs, perhaps improved predictions with the BCS tool could be obtained going forward with further subdivision of categories into weak acids, weak bases, and lipophilic compounds (201), however this was not within the scope of this study. Additionally, the accuracy of the ANN model compared favourably to accuracies of 80% and 66% achieved in previous computational approaches to predict FE classification (214, 215). Ultimately, direct comparisons of prediction accuracies are unachievable due to the absence of external test sets in these

previous publications and differences in the datasets used to build the respective models.

This work identifies ANN as an accurate and efficient solution in detecting correlations between FE and drug properties from a cohort of newly licensed drugs using only 4 well-known drug properties. In the grand paradigm of drug development and computational pharmaceutics this model would provide early indication of significant FE, facilitating informed decisions in drug development. Including whether redesign of the drug candidate may be relevant based on drug properties or whether a bio-enabling strategy should be applied, prompting subsequent application of other computational models, such as an example from our own group (Chapter 4) (86, 114, 160, 235), which would give an indication as to whether a lipid-based formulation (LBF) or other formulation strategies which are reported to overcome FE, may offer a high likelihood of success.

The effects of food on drug bioavailability are multifaceted, involving physiological, physicochemical, and biochemical mechanisms, and as a result, similar to all current methods to anticipate FE, limitations of our models are acknowledged. By design, data-driven modelling approaches identify correlations between individual drug properties of the dataset and classifications, thereby facilitating prediction of the general physiological changes exert by food on drug absorption. However, such an approach, limited by the physicochemical parameters employed, cannot capture all contributing factors drug specific FE such as effect on metabolism, bile-micelle binding, specific chelation between food and a drug or activity of a specific transporter or enzyme (92, 200, 201). As understanding of such effects continues to grow across future studies, in line with continued improvements and understanding of other approaches of FE prediction, opportunities will exist to broaden the utility of these

predictions to incorporate such factors. This aside, these current models successfully predict FE category, providing evidence of a computational tool suited for easy integration within current pre-clinical drug development.

Conclusion

In this study, innovative predictive models using two ML algorithms (SVM and ANN) were developed which accurately predicted the FE category of drugs licensed between 2016-2020, of which 43% demonstrated significant FE. These models were found to possess greater prediction accuracy than FE predictions using the BCS criteria and performed strongly upon comparison to previously published tools using older drug datasets. This predictive modelling enabled key physicochemical parameters that contribute to the effect of food on the extent of drug absorption to be identified, namely S+logP, T_PSA, HBD and Dose. Therefore, this exploratory work provides a further mechanistic basis to understand a drugs behaviour in fed and fasted conditions using a contemporary cohort of licensed drugs. Of course, the rationale and requirements for FE determination will differ depending on the stage of drug development, be that preliminarily formulation testing or the investigation of specific pharmacokinetic parameters. Regardless, the ML tools, particularly the ANN model produced in this study, can facilitate screening of drug candidates, with little cost and effort at the early stages of drug development, utilising only easily recognisable drug properties.

Chapter 4: Applying Computational Predictions of Biorelevant Solubility Ratio Upon Self-Emulsifying Lipid-Based Formulations Dispersion to Predict Dose Number

Harriet Bennett-Lenane[1], Niklas J. Koehl[1], Patrick J. O'Dwyer[1,2], Karl J. Box[2], Joseph P. O'Shea[1], Brendan T. Griffin[1].

[1]School of Pharmacy, University College Cork, Cork, Ireland.

[2]Pion Inc. (UK) Ltd., Forest Row, East Sussex, UK.

Published in: *Journal of Pharmaceutical Sciences (2020), 110 (1)*

DOI: 10.1016/j.xphs.2020.10.055

Abstract

Purpose: Computational approaches are increasingly utilised in development of bio-enabling formulations, including self-emulsifying drug delivery systems (SEDDS), facilitating early indicators of success. This study investigated if *in silico* predictions of drug solubility gain i.e., solubility ratios (SR), after dispersion of a SEDDS in biorelevant media could be predicted from drug properties.

Methods: Apparent solubility upon dispersion of two SEDDS in FaSSIF was measured for 30 structurally diverse poorly water-soluble drugs. Molecular descriptors were used as inputs during partial least squares (PLS) and multiple linear regression (MLR) modelling.

Results: Increased drug solubility upon SEDDS dispersion was observed in all cases, with higher SRs observed for cationic and neutral versus anionic drugs at pH 6.5. PLS models for SR_{MC} (r^2 = 0.81) and SR_{LC} (r^2 = 0.77) were developed. MLR facilitated generation of simplified SR equations with high predictivity (SR_{MC} r^2 = 0.74; SR_{LC} r^2 = 0.69), requiring only three drug properties; partition coefficient at pH 6.5 ($logD_{6.5}$), melting point (T_m) and aromatic bonds as a fraction of total bonds (F_AromB).

Conclusion: This study demonstrated that computational predictions can be incorporated within conventional biopharmaceutics related classification systems. By using the equations to inform drug developability classifications (DCS) for drugs that have already been licensed as lipid-based formulations, merits for development with SEDDS was predicted for 2/3 drugs.

Introduction

Increasing numbers of poorly water-soluble drugs (PWSD) in development pipelines has intensified the need for bio-enabling formulations to enhance oral bioavailability (6, 14). One such approach involves administration of drug substances in lipid-based formulations (LBFs), which enhance apparent solubility of PWSD, while potentially also increasing absorption via stimulation of endogenous lipid absorption pathways for lipophilic xenobiotics. Despite numerous commercial examples of LBFs, with previous estimations of up to 4% of orally administered drug products utilising LBFs (99), it was recently observed that relative numbers of new commercial products using LBFs have declined over the last decade (222). Such statistics suggest challenges to more widespread adoption of LBFs among pharmaceutical companies, potentially linked to a lack of clear guidance on appropriate early screening to guide bio-enabling formulation selection (15).

Self-emulsifying drug delivery systems (SEDDS) fall under the umbrella term LBFs, and refer to combinations of oils with surfactants and co-solvents which spontaneously emulsify forming a stable emulsion on dispersion in the gastrointestinal tract (GIT) (25). Ability to self-emulsify and maintain solubilisation on dispersion is a key SEDDS performance determinant. Typically, the drug dose should be soluble in the SEDDS vehicle, and much effort is focused on determining the inherent lipid solubility of the drug, usually involving resource intensive drug solubility screening in a range of lipid excipients (87, 236). More recently, the application of computational models, in conjunction with drug biopharmaceutical profiling, has been explored to support higher throughput formulation selection for LBFs (3, 54). While certain computational approaches aim to determine drug properties which produce favourable oral drug candidates, other tools have instead examined molecular properties that may

signal necessity for use of bio-enabling strategies or alternatively signal greater suitability for a particular type of bio-enabling strategy. Regarding the latter, a number of noteworthy studies have demonstrated utility of computational pharmaceutics approaches to predict lipid solubility to act as a guide for maximal dose loading in LBF pre-concentrates (114, 115, 161, 237). Critically, while predictions of the drug solubility in lipids are useful to guide initial understanding of the maximum dose loading in the SEDDS vehicle, this approach does not represent the sole criterion for LBF suitability.

Modifications in the GIT upon SEDDS ingestion are crucial in determining formulation performance, as solubilisation capacity in luminal media can be altered dramatically following SEDDS emulsification and through interactions of lipid excipients with endogenous solubilising species (17, 238, 239). SEDDS dispersion leads to increased drug solubilisation, transient supersaturation, and potentially precipitation, thereby presenting drugs to intestinal fluids at concentrations exceeding their equilibrium solubility (240). From a biopharmaceutical perspective, apparent drug solubility in intestinal fluid upon SEDDS dispersion appears critical in determining LBF suitability. Accordingly, the lipid formulation performance classification system emphasises formulation capability to retain solubilisation upon dispersion and digestion (88). The use of simulated biorelevant fluids in such *in vitro* assessments is likely to be a more reliable indicator of whether a SEDDS approach can effectively solubilise the dose *in vivo*. Biorelevant testing is an integral part of pharmaceutical characterisation, revealing concentrations likely to be soluble within human intestinal fluids (HIF) (241, 242), while a key tenet of the developability classification system (DCS) is the use of biorelevant solubility in fasted state simulated intestinal fluids (FaSSIF) as an improved guide to *in vivo* performance and drug

developability (13). More recently, a refined DCS (rDCS) extended this developability concept to include customised *in vitro* assessments of supersaturation and precipitation risks involving absorption number (An) and dose number (Do) (67). Such developability guides, along with decision trees utilising biorelevant media instead of buffered aqueous media (5, 13, 243, 244), signify emerging emphasis on developability and biopharmaceutical concepts in early product testing. However, as *in vitro* techniques utilised to predict the dose that is effectively solubilised *in vivo* can be complex and resource heavy, development of models capable of predicting this dose are strongly merited (51).

With regard to both advancing LBF computational pharmaceutics and use of biopharmaceutically relevant conditions, our hypothesis was to apply a computational approach to predict solubility increases upon SEDDS dispersion. Given the inherent complexity of the mixed colloidal species formed upon dispersion of SEDDS with endogenous biliary lipids, approaches to predict apparent solubility are considered complex at this stage. As an alternative, the solubility increase achieved via SEDDS dispersion in FaSSIF, relative to drug solubility in FaSSIF, represents a more realistic modelling parameter. This can be used to inform the maximal dose solubilised within the intestine, assuming experimental drug solubility in FaSSIF is known.

Accordingly, this study attempted to apply a computational approach in relating drug properties to predict solubility increases (i.e., solubility ratios) following SEDDS dispersion. This approach can therefore be used to effectively guide the dose number (Do) produced in intestinal fluids. Subsequently, this study explored suitability of linking the predicted Do to the framework provided by the DCS and hence, providing a tool for guiding developability of a SEDDS formulation strategy in early stage drug development. To achieve this aim, apparent drug solubility of 30 PWSD was

experimentally determined upon dispersion in FaSSIF of two prototype SEDDS. SEDDS were selected based on prior ability to spontaneously emulsifying, forming a stable microemulsion and were composed of either a medium chain ($SEDDS_{Migylol812}$) or long chain ($SEDDS_{OliveOil}$) oil phase, with a common surfactant, co-surfactant blend in order to examine their excipient effects (51). Solubility ratios (SR_{MC} and SR_{LC}) achieved versus FaSSIF solubility were collated with drug descriptors to develop computational models and predictive equations. Through prediction of DCS classifications, this work aimed to advance the concept of computational pharmaceutics to inform drug developability, exemplifying use of predictive tools to expedite formulation options in early development.

Methods

Dataset Selection

A dataset of 30 structurally diverse PWSD was selected (Table 4-1). Drugs were selected based on a range of criteria including availability of published reports of drug properties, utilisation in previous LBF computational modelling publications and drugs commercially licensed as LBFs (114, 115, 119, 243). Light absorbing ability of the compounds' UV-chromophores were also considered, to allow sufficient detectability by the fibre optic UV probes of the µDISS Profiler. A final drug data set was selected to ensure a sufficient representation of drugs categorised as anionic (8), cationic (9) and neutral (13) overall at pH 6.5. The Henderson-Hasselbalch equation was used to determine ionisation at pH 6.5 (Table 4-1). All drug compounds were purchased from Kemprotec Ltd (Cumbria, United Kingdom). The final dataset displayed a wide physicochemical profile of molecular weight (MW) (230-868.44 g/mol), lipophilicity (clogP) (2.1-7.1) and melting point (T_m) (79-296.5ºC).

Table 4-1: Selection of physicochemical and molecular properties of investigated compounds collated from literature or ADMET Predictor 9.5. 0 = no charge at pH 6.5, + = positive charge, - = negative charge. Am = Ampholyte. % refers to the percentage of the drug's ionisable groups ionised at pH 6.5 according to the Henderson-Hasselbalch Equation.

Drug Compound	MW (g/mol)	clogP	$logD_{6.5}$	Acid/Base /Neutral	pKa (% ionised at pH 6.5)	Classification of Charge pH 6.5	T_m (°C)	PSA (Å^2)	HBD	HBA	Ro5	Max Dose (mg)
Albendazole	265.3	2.81	2.80	Ampholyte	10.26 (0%), 2.8 (0%)	0	209	67.01	2	3	0	200
Candesartan Cilexetil	610.7	5.70	2.89	Ampholyte	6 (76%)	-	163	143.3	1	8	1	32
Carbamazepine	236.27	2.40	2.40	Basic	13.9 (0%)	0	190.2	46.3	1	1	0	300
Carvedilol	406.4	3.88	2.36	Basic	7.8 (95%)	+	114.5	75.7	3	5	0	25
Celecoxib	381.37	3.81	3.81	Acidic	11.1 (0%)	0	158	77.98	1	3	0	200
Cinnarizine	368.6	4.92	3.98	Basic	8.4 (99%)	+	119	6.48	0	2	1	25
Clofazimine	473.40	7.11	4.54	Basic	8.51 (99%)	+	211	40	1	4	1	50
Clotrimazole	344.9	5.08	5.06	Basic	6.7 (96%)	+	142	17.8	0	1	1	10
Danazol	337.5	4.26	4.26	Neutral	-	0	227	46.3	1	2	0	200
Dipyridamole	504.64	3.10	3.02	Basic	6.59 (55%)	+	163	145	4	12	2	200
Felodipine	384.3	5.03	5.03	Basic	5.07 (3%)	0	143	64.4	1	3	0	10
Fenofibrate	360.9	5.20	5.20	Neutral	-	0	79	52.6	0	4	1	150
Glipizide	445.5	2.12	1.48	Acidic	5.9 (80%)	-	201.5	130.15	3	6	0	10
Griseofulvin	352.77	2.51	2.51	Neutral	-	0	220	71.1	0	6	0	500
Haloperidol	375.9	3.82	2.06	Basic	8.3 (98%)	+	151	40.5	1	3	0	20
Indomethacin	357.8	4.03	1.45	Acidic	4.5 (99%)	-	160	68.5	1	4	0	50
Irbesartan	428.53	3.68	2.84	Ampholyte	4.12 (0%), 7.4 (11%)	0	180.5	87.13	1	5	0	300
Isotretinoin	300.44	6.07	3.99	Acidic	4 (99%)	-	174	37.3	1	2	1	40
Itraconazole	705.7	4.89	4.89	Basic	3.7 (0%)	0	166	104.7	0	9	2	100
Ketoconazole	531.43	3.67	3.51	Basic	6.75 (64%), 4.22 (0%)	+	146	69.06	0	6	1	200
Mefenamic Acid	241.29	4.91	2.36	Acidic	3.89 (99%)	-	230.5	49.3	2	3	1	500
Naproxen	230.26	3.21	1.10	Acidic	4.15 (99%)	-	153	46.5	1	3	0	500
Nifedipine	346.34	3.10	3.10	Acidic	3.93 (99%)	-	173	110.45	1	5	0	90
Phenytoin	252.27	2.09	2.07	Acidic	8.3 (2%)	0	296.5	58.2	2	2	0	300
Progesterone	314.5	3.94	3.94	Neutral	-	0	128	34.1	0	2	0	200
Spironolactone	416.57	3.28	3.28	Neutral	-	0	134.5	60.44	0	3	0	100
Tamoxifen	371.52	6.59	4.61	Basic	8.5 (99%)	+	97	12.5	0	2	1	20
Terfenadine	471.67	5.60	3.61	Basic	10 (99%)	+	147	47.3	2	3	1	60
Tolfenamic Acid	261.7	5.13	2.44	Acidic	5.11(96%)	-	213	49.3	2	3	0	200
Venetoclax	868.44	6.76	6.54	Ampholyte	3.4 (99%), 10.3 (99%)	0	138	172.03	3	10	2	100

Formulations

Two SEDDS previously utilised for oral delivery of a model PWSD in preclinical studies were chosen (51). $SEDDS_{Migylol812}$ contained 40% w/w medium chain triglycerides (Miglyol 812) with 20% w/w surfactant (Kolliphor RH 40 - polyoxyl-40-hydrogenated castor oil) and 40% w/w co-surfactant (Tween 85 - polyoxyethylene-(20)–polysorbitan trioleate). $SEDDS_{OliveOil}$ contained 40% w/w long chain triglycerides (olive oil), while quantities of surfactant and co-surfactant remained similar to $SEDDS_{Migylol812}$, with 20% Kolliphor RH 40 and 40% Tween 85. Migylol 812N is primarily composed of C8 and C10 fatty acids (approx. 60:40%). Olive oil contains saturated and unsaturated fatty acids of primarily C16-C18 chain length. Miglyol 812N was kindly gifted from IOI Oleo GmbH (Hamburg, Germany), while Olive Oil, Tween 85 and Kolliphor RH 40 were purchased from Sigma-Aldrich (Ireland). SEDDS were prepared by weighing exact excipient quantities into a screw cap glass tube and incubated at 37 °C, overnight on a stirring plate 200 rpm (Mixdrive 15, 2MAG, Germany).

Media Preparation

Phosphate buffer ($PhB_{pH6.5}$) and FaSSIF-V1 were prepared according to biorelevant.com (Croydon, UK) protocol and adjusted to pH 6.5 using a Model 3510 pH/mV/Temperature Meter (Jenway, UK). FaSSIF-V1 was chosen due to high correlation with HIF and availability of drug solubility datasets (245, 246). Water was obtained from a MilliQ water system. All chemicals and solvents were of analytical or high-performance liquid chromatography (HPLC) grade and purchased from Sigma-Aldrich (Ireland). Conditions for simulating dispersion of the SEDDS in intestinal fluids were produced by dispersing SEDDS (1:200 dilution) in $PhB_{pH6.5}$ (i.e.,

$PhB_{pH6.5}$-$SEDDS_{Migylol812}$ and $PhB_{pH6.5}$-$SEDDS_{OliveOil}$) and FaSSIF (i.e., FaSSIF-$SEDDS_{Migylol812}$ and FaSSIF-$SEDDS_{OliveOil}$). This lipid dilution was chosen to be typical of reasonable lipid concentrations found in a biorelevant volume.

Media Characterisation: Media Droplet Size and Zeta Potential

Droplet size (nm) and polydispersity index (PDI) of FaSSIF, FaSSIF-$SEDDS_{Migylol812}$, FaSSIF-$SEDDS_{OliveOil}$, $PhB_{pH6.5}$-$SEDDS_{Migylol812}$ and $PhB_{pH6.5}$-$SEDDS_{OliveOil}$ were measured using Dynamic Light Scattering (DLS) with a Malvern Zetasizer Nano ZS (Malvern Analytical, US) with a 4mW 633nm He-He laser at 37°C with a backscattering angle of 173° using the Stokes-Einstein equation. Measurements were performed with unfiltered samples in disposable UV-cuvettes from Sarstedt AG & Co. KG (Numbrecht, Germany) (10 x 4 x 45 mm). Refractive indices used were 1.1333 ($PhB_{pH6.5}$) and 1.334 (FaSSIF) (247). The electrophoretic mobility i.e. ζ-potential, of colloidal structures in the media was measured using the Zetasizer in disposable folded capillary cells (DTS1070) using the Helmholtz-Smoluchowski equation (248). Each analysis was conducted in triplicate, presented as mean ± standard deviation.

Experimental Solubility Determination

Apparent drug solubility studies in FaSSIF, FaSSIF-$SEDDS_{Migylol812}$ and FaSSIF-$SEDDS_{OliveOil}$ were experimentally determined over 24 hours as the 24-hour time point was used for solubility ratios. Solubility was determined via either shake flask with RP-HPLC/UV analysis (6 drugs) or µDISS Profiler (Pion INC, Woburn, MA) (24 drugs), where preliminary studies verified method comparability (Appendix 3 Table 3-3).

Shake Flask Method

Drug was added in excess to triplicate glass vials containing either FaSSIF, FaSSIF-SEDDS$_{Migylol812}$ or FaSSIF-SEDDS$_{OliveOil}$ (n=3). pH was maintained at 6.5 prior to experiments. Vials were placed on a stirring plate (Mixdrive 15, 2MAG, Germany) in a 37°C incubator at 300 rpm. 300 µl samples were removed at 2, 4, 6 and 24 hours. Excess solid was separated using a centrifuge for 15 minutes at 21,380 x g (Mikro 200 R, Andreas Hettich GmbH & Co. KG, Germany). Samples were diluted in acetonitrile for analysis via RP-HPLC/UV. Drug Detection was conducted using an Agilent 1200 series HPLC system. The columns and mobile phases used for each drug analysed along with injection volume, flowrate and detection wavelength can be found in Appendix 3 Table 3-2.

µDISS Profiler

Apparent drug solubility (n = 3) was determined at a stirring rate of 300 rpm over 24 hours (37°C). Path length of the *in situ* UV probes was varied (1 - 5 mm) depending on anticipated concentration range and the UV absorbance properties of the drug molecule. Standard spectra were collected for each compound at pH 6.5 and a linear relationship ($r^2 > 0.99$) was established between absorbance and concentration in each case. The experimental run was performed in six vials where a large excess of API was added (10-20 times more than the anticipated FaSSIF solubility) to account for the potentially large solubility enhancement. These vials contained 15 ml FaSSIF-SEDDS$_{Migylol812}$ or FaSSIF-SEDDS$_{OliveOil}$ and a cross-bar magnetic stirrer. Two additional channels were used as blanks to consolidate for potential issues with background changing FaSSIF UV absorbance over time. The *in-situ* UV probes scanned the samples at predefined time intervals (30 minutes). Concentrations were

determined by considering area-under-the-curve (AUC) in second derivative spectra, to lessen interference from background turbidity. A range of wavelengths were utilised to quantify drug. Data was interpreted using the Au Pro software (Version 5, Pion INC, MA, USA).

Drug Physicochemical and Molecular Properties

In excess of 250 descriptors including physicochemical and modelling descriptors were obtained from ADMET Predictor 9.5 (Simulations Plus, USA). T_m was obtained from literature (5, 114, 119). Biorelevant solubility values in FaSSIF, FeSSIF and $PhB_{pH6.5}$ were obtained from literature sources where available (119, 249). In absence of published data, predicted solubility values were generated (ADMET Predictor, Ver. 9.5, Simulations Plus Inc., US). Highest licensed drug dosage strengths were obtained from the European Medicines Agency (EMA) or Food and Drug Administration (FDA) databases.

Biopharmaceutical Data Analysis

Apparent drug solubility values in all media are presented as mean ± standard deviation (n=3) (Appendix 3 Table 3-1). Solubility ratios (SR) for the 30 drugs with either FaSSIF-SEDDS$_{Migylol812}$ (referred to as SR_{MC}) or FaSSIF-SEDDS$_{OliveOil}$ (referred to as SR_{LC}) versus FaSSIF were calculated via Equation 1:

$$(1)\ Solubility\ Ratio\ (SR) = \frac{Apparent\ Solubility\ FaSSIF{-}SEDDS}{Apparent\ Solubility\ FaSSIF}$$

SR standard error (SE) was calculated from Equation 2 as previously reported (5):

$$(2)\ \mathrm{SE} = SR\ x\ \sqrt{\frac{SA^2}{A^2} + \frac{SB^2}{B^2}}$$

Where A, B, SA and SB refer to the mean measured solubility values (24hrs) and standard errors for A (FaSSIF) and B (FaSSIF-SEDDS) respectively. In order to assess capacity for SEDDS to bridge the fasted-fed state solubility gap, SR_{MC} and SR_{LC} were related to comparative SRs for each drug using FeSSIF solubility in place of drug solubility upon SEDDS dispersion i.e., FeSSIF/FaSSIF. Graphs illustrating SRs were obtained using Prism (Version 5, Graphpad, USA). Linear regression was performed using Excel (Microsoft Office, 2016) to assess correlations between SR and individual drug properties or solubility in various media. To test significance between paired solubility values in FaSSIF-SEDDS$_{MC}$ versus FaSSIF-SEDDS$_{LC}$ the distribution of the difference was used to determine normality. A two-sided bootstrap-paired test (5000 samples) was used to determine significance ($p < 0.05$). A simple scatter plot was produced for FaSSIF-SEDDS$_{Migylol812}$ versus FaSSIF-SEDDS$_{OliveOil}$ and regression coefficients fitted for interpretation and a bootstrap test for the coefficients conducted. A two-sided independent samples t-test was used to analyse media droplet sizes and Levene's test was used to check for equality of variances. A p-value < 0.05 indicated a violation of equal variance. All statistical analysis was conducted using SPSS Statistics (Version 26, IBM Corporation, US).

Multivariate Data Analysis and Modelling Parameters

Multivariate data analysis (MDA) was conducted using Unscrambler (Version 11, Camo Analytics, US). Molecular structures were acquired as smiles from PubChem and used as inputs for the ADMET Predictor software (Version 9.5, Simulations Plus, California, USA) to calculate >250 molecular descriptors. These were added to PSA

and T_m and used as variable inputs for principal component analysis (PCA) and partial least squares (PLS) modelling. Modelling responses were the logarithm of SR in both SEDDS ($logSR_{MC}$ and $logSR_{LC}$). PCA was first applied randomly to aid training/test set identification. A split of 70:30 (21:9 drugs) of training:test set was used to increase model robustness. Training set criteria was that it covered the test set chemical space along with a relatively even spread of SRs. Influential outliers were placed in the test set if they displayed both large residual and high leverage in the influence plot. A Hotelling's T^2 ellipse was also applied for outlier detection (95% confidence interval).

PLS was used to establish important descriptors for predicting SR_{MC} and SR_{LC}. The nonlinear iterative partial least squares (NIPALs) algorithm was utilised and all 250+ variables were mean centred, de-identified and standardized through scaling by standard deviation. Descriptors displaying the same value for all drugs were removed, along with skewed descriptors. To limit overfitting potential, a limit of two principal components was used. Variable reduction was performed to decrease complexity and noise. A Martens' uncertainty test was applied to help identify important variables and assess stability. This involved a "jackknifing" procedure and production of sub-models to identify non-significant variables (250). Variable weighted beta coefficient rankings from the Important Variables plot and their p-values were also used to remove unimportant variables. Variables in the same area of the correlation loadings plot were removed leaving a singular variable. Variables near the centre of this plot were removed. Any change in r^2 calibration and r^2 validation was monitored. Model accuracy was validated by the Root Mean Square Error (RMSE) of validation and calibration. Models were validated by a full cross validation (leave-one-out) to improve power and by test sets of drugs not used in model development to strengthen general applicability.

Solubility Equation for Predicting Biopharmaceutical Dose Number and DCS Class.

It was then investigated if easily interpretable equations based on drug properties could predict SR_{MC} and SR_{LC}. Multiple linear regression (MLR) was performed using Excel (Microsoft Office, 2016) to investigate correlations between selected significant PLS model variables versus $logSR_{MC}$ and $logSR_{LC}$. Equation development was monitored by descriptor p-values, the f-value, r^2 and adjusted r^2. The same training and test sets as PLS were used.

DCS classification of each drug was obtained using solubility and permeability parameters outlined previously (13, 67). While drug permeability was predicted from the ADMET Predictor (Version 9.5, Simulations Plus Inc., US), solubility criteria was obtained using a dose/solubility ratio in 500mls of media using equation (3):

$$(3)\ Do = \frac{Dose}{(Ssi\ )(Vsi)}$$

Where, Dose is the highest dose, S_{si} apparent solubility in biorelevant media i.e., FaSSIF, V_{si} is the available fluid volume for dissolution in the small intestine (500 mL).

Solubility criteria for DCS classification using experimentally determined solubility's upon SEDDS dispersion was calculated using equation (4):

$$(4)\ Do_{(SEDDS)} = \frac{Dose}{(Cs)(\ Vsi)}$$

Here, C_s is apparent drug solubility upon SEDDS dispersion in biorelevant media i.e., FaSSIF-SEDDS.

For DCS classifications using the predicted solubility ratios (SR) from the MLR equations, $Cs_{(Predicted)}$ was calculated using equation 5:

$$(5)\ Cs_{(Predicted)} = SR * Ssi$$

Where SR is the predicted solubility ratio upon SEDDS dispersion from the MLR equations and S_{si} is apparent solubility of the compound in biorelevant media i.e., FaSSIF. Incorporating equation 5, solubility criteria for DCS classifications upon SEDDS dispersion was predicted using equation 6:

$$(6)\ Do_{(Predicted)} = \frac{Dose}{(Cs_{(Predicted)})(Vsi)}$$

Predicting DCS Classifications of Commercial LBF Drugs

To assess the equations' general applicability to make predictions for drugs outside equation development and validation, $Do_{(Predicted)}$ was applied to a list of drugs that have been successfully commercialised as LBF products (222). DCS classifications were produced using $Do_{(Predicted)}$ values (Equation 6), to predict if dose solubility limitations for the commercial drugs would be overcome upon SEDDS dispersion. FaSSIF solubility was obtained from literature or from the ADMET Predictor 9.5 (Simulations Plus, USA) (119, 249). Predicted classifications were compared to classifications using FaSSIF solubility alone (Equation 3).

Results

SEDDS Characterisation on Biorelevant Dispersion

SEDDS$_{Migylol812}$ and SEDDS$_{OliveOil}$ both dispersed in FaSSIF and PhB$_{pH6.5}$ to form uniform stable microemulsions with droplet sizes between 36-70 nm. PhB$_{pH6.5}$-SEDDS$_{Migylol812}$ and PhB$_{pH6.5}$-SEDDS$_{OliveOil}$ displayed significantly different mean droplet sizes (* $p < 0.05$) (Table 4.2). Droplet sizes of FaSSIF-SEDDS$_{Migylol812}$ and FaSSIF-SEDDS$_{OliveOil}$ also differed (* $p < 0.05$) and were smaller than PhB$_{pH6.5}$-SEDDS$_{Migylol812}$ and PhB$_{pH6.5}$-SEDDS$_{OliveOil}$. All PDI's obtained were below 0.26 indicating droplet sizes on dispersion were moderately homogenous. In terms of charge, values close to zero mV were observed for PhB$_{pH6.5}$-SEDDS$_{Migylol812}$ and PhB $_{pH6.5}$-SEDDS$_{OliveOil}$ as all SEDDS excipients were non-ionic and neutral. FaSSIF displayed an overall net negative charge (-14.67 mV), which remained, though reduced in magnitude, through dispersion of SEDDS$_{OliveOil}$ (-5.73 mV) and SEDDS$_{Migylol812}$ (-5.35 mV) (Table 4-2).

Table 4-2: Size determination and ζ-potential of the media used during the course of the analysis demonstrating that both SEDDS dispersed uniformly to form stable microemulsions.

Media	Size nm (SD)	PDI (SD)	ζ-potential mV (SD)
PhB$_{pH6.5}$-SEDDS $_{Migylol812}$	47.71 (1.587)	0.147 (0.01)	-0.76 (0.49)
PhB$_{pH6.5}$-SEDDS$_{OliveOil}$	70.18 (3.003)	0.259 (0.009)	-1.27 (0.32)
FaSSIF-SEDDS $_{Migylol812}$	36.41 (1.096)	0.081 (0.014)	-5.35 (0.31)
FaSSIF-SEDDS$_{OliveOil}$	44.76 (0.303)	0.197 (0.002)	-5.73 (0.16)
FaSSIF	62.52 (5.867)	0.234 (0.073)	-14.67 (0.42)

Solubility in Biorelevant SEDDS Dispersions – Comparison of SEDDS $_{Migylol812}$ and SEDDS $_{OliveOil}$

For the 30 drugs, solubility in FaSSIF-SEDDS$_{Migylol812}$ was higher than FaSSIF-SEDDS$_{OliveOil}$ as a paired bootstrap test revealed a significant difference in drug solubility between these medium chain and long chain lipid dispersions (* $p < 0.05$). Comparatively, the beta coefficient of the regression line for FaSSIF-SEDDS$_{Migylol812}$ versus FaSSIF-SEDDS$_{OliveOil}$ was significant according to a bootstrap for coefficients test (* $p < 0.05$). A strong correlation was established between drug solubility in FaSSIF-SEDDS$_{Migylol812}$ and FaSSIF-SEDDS$_{OliveOil}$ (r^2 0.97) (Figure 4-1), suggesting that for every 100 unit increase in FaSSIF-SEDDS$_{OliveOil}$ solubility units, FaSSIF-SEDDS$_{Migylol812}$ increases on average by 105.6 solubility units. Consequentially, this indicates that solubility determined in one lipid dispersion may be used to estimate solubility in the other.

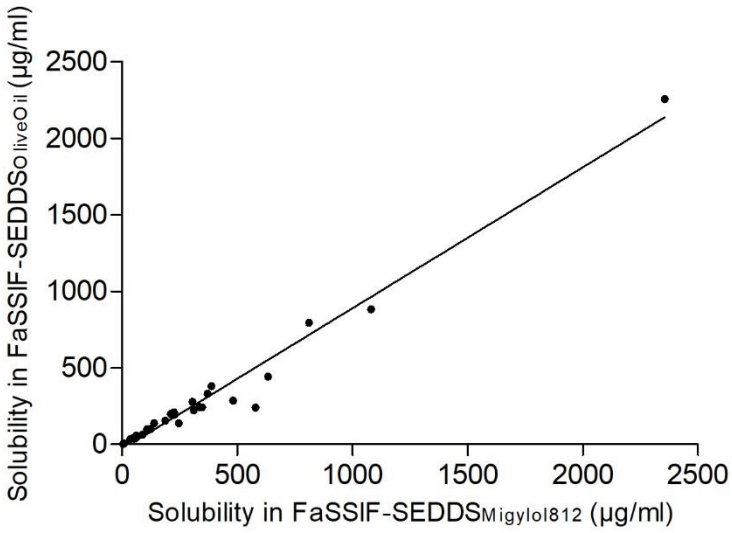


Figure 4-1: Scatter plot of drug solubility in FaSSIF-SEDDS$_{Migylol812}$ versus Solubility in FaSSIF-SEDDS$_{OliveOil}$ displaying a high correlation (r^2 = 0.9722). Linear regression line: FaSSIF-SEDDS$_{Migylol812}$ = 42.47 + 1.056(FaSSIF-SEDDS$_{OliveOil}$)

Solubility Ratio Trends

Solubility ratios (SR) for 30 PWSDs upon dispersion of two SEDDS was experimentally determined (Figure 4-2), where SR >1 was seen in all cases, indicative of increased drug solubility on SEDDS dispersion in intestinal media. SRs ranged from 1.13 - 64.4 fold for $SEDDS_{Migylol812}$ and from 1.04 - 59.7 fold for $SEDDS_{OliveOil}$. In presence of both SEDDS, Clotrimazole and Fenofibrate displayed the highest SRs. Trends in ionisable drugs were analysed. Cationic drugs appeared to consistently display high SR, with all such compounds displaying solubility gains of >2, with 3 and 2 drugs respectively displaying SR >10 fold in presence of $SEDDS_{Migylol812}$ and $SEDDS_{OliveOil}$. In contrast, solubility gains for anionic compounds appeared less pronounced, with 8/9 anionic drugs displayed SR <5. However, for Candesartan Cilexetil, a SR >16 was observed with both SEDDS. Candesartan Cilexetil is an ampholyte where the hydrogen attached to the O-CH(CH_3)-O group in the cilexetil side chain is moderately acidic, being between the oxygen rich ester moieties, while also possessing a basic functional group. This ampholytic nature may have contributed to its deviation from the general trends observed for other anionic drugs.

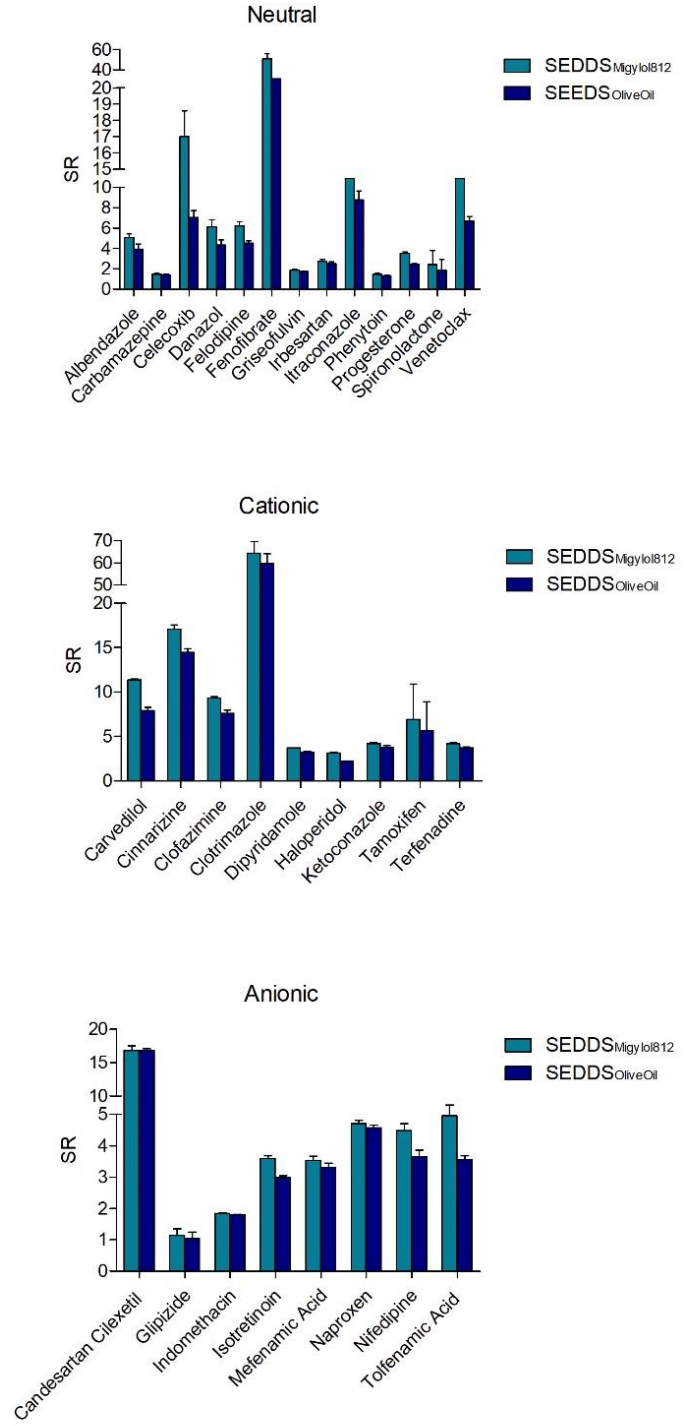


Figure 4-2: SR (drug solubility in dispersed SEDDS media/FaSSIF) achieved for neutral, cationic and anionic drugs (pH 6.5). Higher SRs are seen in general for cationic and neutral drugs versus anionic drugs where every anionic drug except candesartan cilexetil achieved a SR < 5 in both SEDDS.

Neutral drugs displayed a wide range of SRs, while Celecoxib and Venetoclax deviated strongly from the trend of similar SRs in $SEDDS_{Migylol812}$ and $SEDDS_{OliveOil}$, with Celecoxib displaying a SR_{MC} of 17 compared to a SR_{LC} of 7, while Venetoclax also displayed a difference between SR_{MC} and SR_{LC} i.e., 12 versus 7. To assess SEDDS ability to mirror solubility increases in fed-state versus and fasted-state media, SRs obtained were compared to FeSSIF/FaSSIF solubility ratios. SR_{MC} and SR_{LC} exceeded $SR_{FeSSIF/FaSSIF}$ for 24 and 23 of the 30 drugs respectively (Figure 4-3). This observation confirms the utility of SEDDS as effective bio-enabling systems to bridge the fasted-fed state solubility gap (92).

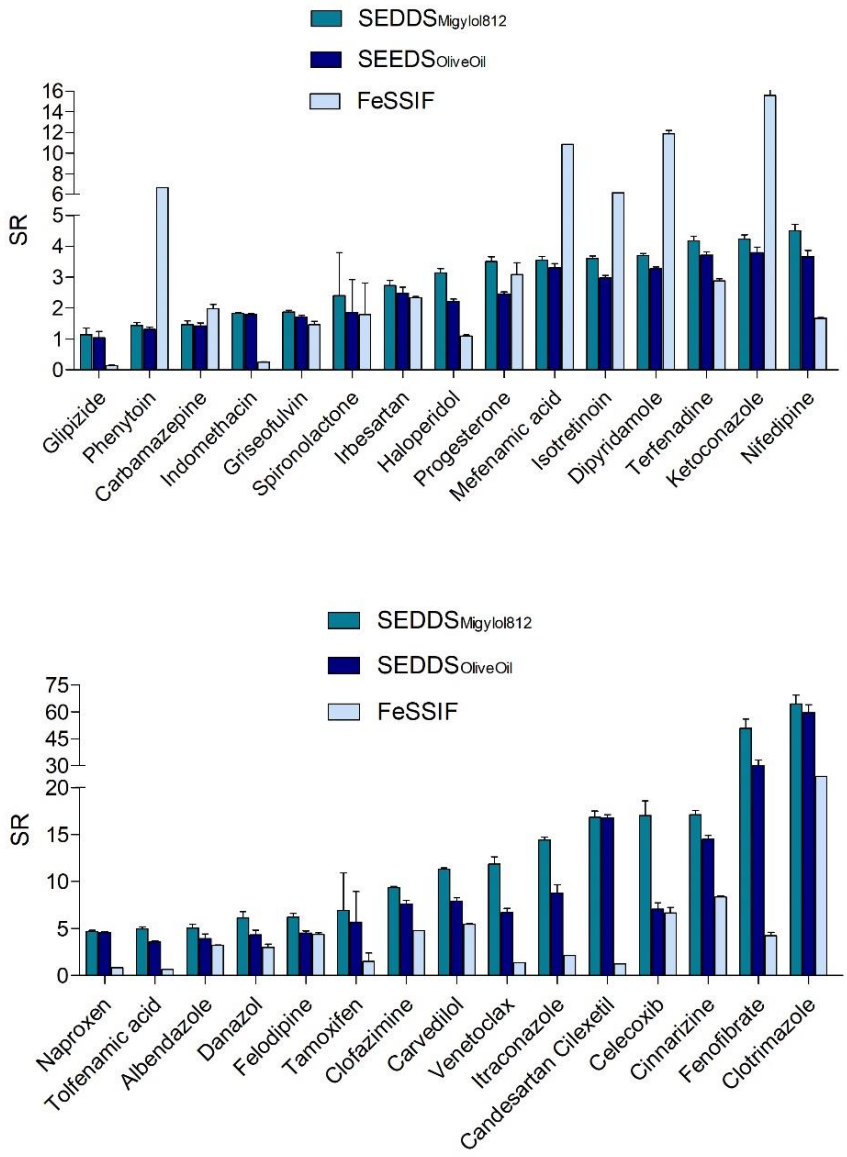


Figure 4-3: SR (drug solubility in both dispersed SEDDS media/FaSSIF and FeSSIF/FaSSIF). FeSSIF/FaSSIF SR is overcome with $SEDDS_{MC}$ for 24 drugs and with $SEDDS_{LC}$ for 23 drugs, demonstrating ability of the SEDDS to bridge the FeSSIF-FaSSIF solubility gap.

Computational Prediction of Biorelevant Solubility Gain with SEDDS.

Linear regression revealed weak correlations between both SR_{LC} and SR_{MC} versus individual drug properties. Lipophilicity and T_m, commonly utilised as guides towards LBF suitability, displayed poor quantitative relationships e.g., logP (r^2 0.33, 0.32), $logD_{6.5}$ (r^2 0.43, 0.35) and T_m (r^2 0.23, 0.25). Therefore, a combination of variables was required to improve quantitative prediction accuracy. Firstly, PCA verified the structural diversity of the dataset (Appendix 3 Figure 3-1). PLS model development resulted in predictive PLS models for both SRs ($logSR_{MC}$ and $logSR_{LC}$). The PLS models used 1-2 principal components (PC) and 5-6 variables. The $logSR_{MC}$ 1 PC model produced predictions of r^2 calibration 0.81, r^2 validation 0.73 requiring 5 variables; $logD_{6.5}$, melting point (T_m), molecular weight (MW), aromatic bonds as fraction of total bonds (F_AromB) and Atom-Type Cumulative Electrotopological State (E-state) index for methylene carbons (SssCH2). While the $logSR_{LC}$ 2 PC model required 6 variables; $LogD_{6.5}$, MW, T_m, F_AromB, SssCH2 and number of aliphatic rings (N_AliphR) to produce predictions of r^2 calibration 0.77, r^2 validation 0.67. These models demonstrated good predictions of test sets, summarized in Table 4-3.

Table 4-3: Overview of the PLS models and MLR equations produced for SR_{MC} and SR_{LC}**.** RMSEC = root mean square error of calibration, CV = cross validation, RMSEP = root mean square error of prediction.

PLS

Y-Variable	**$logSR_{MC}$**	**Y-Variable**	**$logSR_{LC}$**
X-Variables	$logD_{6.5}$ MW T_m F_AromB SssCH2	**X- Variables**	$logD_{6.5}$ MW T_m F_AromB SssCH2 NAlip_R
Explained Y-Variance (%)	**81%**	**Explained Y-Variance (%)**	**77%**
No. of PC's	**1**	**No. of PC's**	**2**
RMSEC	**0.19**	**RMSEC**	**0.20**
RMSEP CV	**0.24**	**RMSEP CV**	**0.26**
RMSEP Test Set (n=7)	**0.36**	**RMSEP Test Set (n=7)**	**0.37**
r^2 (Calibration)	**0.81**	**r^2 (Calibration)**	**0.77**
r^2 (Validation)	**0.73**	**r^2 (Validation)**	**0.67**

MLR

Y Variable	**r^2**	**$RMSE_{Tr}$**	**$RMSE_{Te}$**	**F-value**	**p-value**	**Equation**
$logSR_{MC}$	0.74	0.23	0.39	16.02	$3.34x10^{-5.}$	$\boldsymbol{logSR_{MC}} = 0.6 + 0.2(logD_{6.5}) + 1.02(F_AromB) - 0.01(T_m)$
$logSR_{LC}$	0.69	0.43	0.37	12.64	$1.37x10^{-4}$	$\boldsymbol{logSR_{LC}} = 0.54 + 0.17(logD_{6.5}) + 1.04(F_AromB) - 0.01(T_m)$

Enhanced Biorelevant Solubility Ratio Equation

As 5-6 descriptors could predict SR_{MC} and SR_{LC}, MLR was performed to produce easily interpretable predictive equations. All significant variables from PLS modelling were initially included in MLR. Insignificant variables ($p > 0.05$) from these initial equations were subsequently removed, resulting in final equations with higher F-values and significant variables. Two equations were produced (Table 4-3), both utilising 3 properties: $logD_{6.5}$, T_m and F_AromB. Similarities between equations was expected due to the high correlation between dispersed SEDDS (Figure 4-1).

Use of Predicted Solubility Ratios to Predict Drug DCS Class with SEDDS.

Application of the equations to predict drug DCS class with SEDDS was assessed and accuracy compared to comparative DCS classifications using experimentally determined solubility's upon SEDDS dispersion. DCS permeability classifications were estimated using drug permeability predictions from the ADMET Predictor 9.5 (Simulations Plus, USA). While use of a computationally derived permeability estimate has been applied in other studies (251), it must be acknowledged that drug specific effects may not be adequately captured in these predicted permeability estimates. In total, using experimental solubility's, 10 drugs overcame a solubility limitation i.e., transitioned to DCS Class I/III. Using the $Do_{(Predicted)}$ approach (Equation 6), this transition was correctly predicted for 8/10 drugs (Table 4-4) i.e. Clotrimazole, Cinnarizine, Fenofibrate, Isotretinoin, Naproxen, Terfenadine, Glipizide and Venetoclax. DCS Classification using $Do_{(SEDDS)}$ (Equation 4) also resulted in transitions to "good solubility" for Candesartan Cilexetil and Celecoxib ($SEDDS_{Migylol812}$ only), however, as previously discussed experimental results for both

drugs differed significantly from general trends observed, which may suggest a drug specific effects in these cases that was not captured in the MLR equations.

Table 4-4*:* DCS classification and transitions of the 30 drugs using both experimental and predicted solubility values.

Drug **Do Equation Used:**	**FaSSIF** **Do**	**SEDDS*Migylol812*** **Do(SEDDS)**	**SEDDS*Migylol812*** **Do(Predicted)**	**SEDDS*OliveOil*** **Do(SEDDS)**	**SEDDS*OliveOil*** **Do(Predicted)**	**Transition**
Albendazole	IIb	IIb	IIb	IIb	IIb	
Candesartan Cilexetil	IV	III	IV	III	IV	IV → III
Carbamezapine	IIa	IIa	IIa	IIa	I	
Carvedilol	I	I	I	I	I	
Celecoxib	IIb	I	IIb	IIa	IIb	IIb → IIa/I
Cinnarizine	IIa	I	I	I	I	IIa → I
Clofazimine	IIb	IIa	IIa	IIa	IIa	IIb → IIa
Clotrimazole	IIa	I	I	I	I	IIa → I
Danazol	IIa/IIb	IIa	IIa	IIb	IIa	
Dipyridamole	IV	IV	IV	IV	IV	
Felodipine	I	I	I	I	I	
Fenofibrate	IIb	I	I	I	I	IIb → I
Glipizide	IV	III	III	III	III	IV → III
Griseofulvin	IIb	IIb	IIb	IIb	IIb	
Haloperidol	I	I	I	I	I	
Indomethacin	I	I	I	I	I	
Irbesartan	IV	IV	IV	IV	IV	
Isotretinoin	IIa	I	I	I	I	IIa → I
Itraconazole	IIb	IIb	IIb	IIb	IIb	
Ketoconazole	IIb	IIb	IIa/IIb	IIb	IIb	
Mefenamic Acid	IIb	IIa	IIa/IIb	IIa	IIb	
Naproxen	IIa	I	I	I	I	IIa → I

Drug Do Equation Used:	FaSSIF Do	SEDDS$_{Migylol812}$ Do$_{(SEDDS)}$	SEDDS$_{Migylol812}$ Do$_{(Predicted)}$	SEDDS$_{OliveOil}$ Do$_{(SEDDS)}$	SEDDS$_{OliveOil}$ Do$_{(Predicted)}$	Transition
Nifedipine	IIb	IIa	IIa/IIb	IIa	IIa/IIb	IIb → IIa
Phenytoin	IIb	IIb	IIb	IIb	IIb	
Progesterone	IIb	IIa	IIa	IIa	IIa	IIb → IIa
Spironolactone	IIa/IIb	IIa	IIa	IIa	IIa	
Tamoxifen	I	I	I	I	I	
Terfenadine	IIa	I	I	I	I	IIa → I
Tolfenamic Acid	IIa	IIa	IIa	IIa	IIa	
Venetoclax	IV	III	III	IV	III	IV → III

Predicted DCS Classifications of Commercial LBF Drugs.

The utility of the $Do_{(Predicted)}$ approach to guide a LBF formulation strategy was subsequently assessed by applying the MLR equations to a range of drugs that have been successfully licensed as LBFs. A total of 49 drugs were selected initially, and the DCS classification using FaSSIF solubility alone was employed to determine DCS class. In total, 23 drugs were initially classified as DCS class I/III, and therefore, did not display solubility limitations. These compounds were therefore excluded from further analysis as a bio-enabling strategy was not considered necessary. Applying the $Do_{(Predicted)}$ approach to the remaining 26 drugs, 10 drugs were predicted to transition from poor to good solubility, and a further 7 drugs were found to transition from DCS Class IIb to Class IIa i.e "dissolution rate limited" which can offer delivery opportunities, where the compensatory influence of high permeability has been stated to be significant for acceptable oral absorption during the transit time in the intestine (13). Therefore, this approach predicted that in 65.4% (i.e., 17/26) of drugs, a SEDDS approach was likely to overcome solubility limited absorption. Of the 9 drugs that remained in poor solubility classification after applying the $Do_{(Predicted)}$ approach, 8 were DCS Class IV, which may indicate that decisions to employ a SEDDS approach were not solely influenced by solubility considerations and that other factors, such as increased permeability, may have been a consideration in the choice to develop as a LBF (Table 4-5).

Table 4-5: DCS Class of commercial LBF drugs which displayed dose solubility limitations in FaSSIF.

Drug **Do Equation Used**	**FaSSIF** **Do**	**SEDDS*Migylol812*** **Do(Predicted)**	**SEDDS*OliveOil*** **Do(Predicted)**	**Transition**
Clomethiazole Edisilate	IIa	I	I	IIa → I
Dronabinol	IIa	I	I	IIa → I
Ergocalciferol	IIa	I	I	IIa → I
Isotretinoin	IIa	I	I	IIa → I
Cholecalciferol	IIb	I	I	IIb → I
Clofazimine	IIb	IIa	IIa	IIb → IIa
Efavirenz	IIb	I	IIa	IIb → I/IIa
Enzalutamide	IIb	IIa	IIa	IIb → IIa
Ethyl Eicosapentaenoate	IIb	IIa	IIb	IIb → IIa
Fenofibrate	IIb	I	I	IIb → I
Loratidine	IIb	IIa	IIa	IIb → IIa
Menatetrenone	IIb	IIa	IIa	IIb → IIa
Nimodipine	IIb	I	IIa	IIb → I/IIa
Progesterone	IIb	IIa	IIa	IIb → IIa
Teprenone	IIb	IIa	IIa	IIb → IIa
Tocopherol Nicotinate	IIb	I	IIa	IIb → I/IIa
Amprenavir	IV	III	III	IV → III
Nintedanib	IIb	IIb	IIb	
Azithromycin	IV	IV	IV	
Ciprofloxacin	IV	IV	IV	
Cyclosporin A	IV	IV	IV	
Lopinavir	IV	IV	IV	
Ritonavir	IV	IV	IV	
Saquinavir	IV	IV	IV	
Tipranavir	IV	IV	IV	
Vinorelbine Tatrate	IV	IV	IV	

Discussion

Over the last two decades, significant strides have been made in applying computational approaches across the full spectrum of drug development (252). In their many forms, computational tools can include discovering new lead candidates with optimal drug-receptor binding affinity (e.g., Quantitative Structural Activity Relationships (QSAR)), to guiding on optimal physicochemical profiles (e.g., Quantitative Structural Property Relationships (QSPR)) or predictions of biopharmaceutical properties including solubility and permeability (3). While the major advances in the use of computational tools to-date have been focused on chemical structural design to assist the selection of new drug substances with optimal pharmacodynamic and/or pharmacokinetic properties, commonly referred to as "druggability", more recently, the use of computational tools to guide on formulation design, or computational pharmaceutics, have been reported (3, 131, 157). These include approaches such as computational biopharmaceutical drug profiling, recently reported as an approach to predict physicochemical and molecular properties of drug candidates that render them more or less suitable for formulation via a specific bio-enabling formulation approach (3). Accordingly, there exists an increasing focus on development of reliable computational pharmaceutics tools, capable of guiding selection of appropriate bio-enabling formulation strategies, in particular for drug candidates which display either solubility and/or permeability limitations.

LBFs are one such bio-enabling formulation technology that exploit the benefit of lipid excipients to harness the absorption pathways of dietary fats, leading to increased intestinal drug solubility and improving intestinal absorption. The benefits of lipid excipients to increase drug solubility were clearly prevalent in this study, where increased solubility was observed for all 30 PWSDs following dispersion of the

SEDDS in biorelevant media. Indeed, the solubility increases observed were on average higher than the fed/fasted biorelevant solubility ratio, as SR_{MC} and SR_{LC} exceeded $SR_{FeSSIF/FaSSIF}$ for 24 and 23 of the 30 drugs respectively (Figure 4-3). However, despite clear benefits as a bio-enabling technology, it is generally considered that LBFs have an unfulfilled potential in a commercial sense. Over the last decade, prevalence of commercial LBFs appears to be decreasing relative to solid dispersions (SD) (222), reflecting improved scientific knowledge on the pharmaceutical benefits of SDs in terms of bio-enabling effects (e.g. increased drug solubility), but also an improved understanding of factors influencing industrial scalability and regulatory approval (e.g. long term stability). On the other hand, the prevalence of commercial LBFs has tended to be relatively few, reflecting gaps in understanding both in terms of bio-enabling benefits and from an industrial perspective, as recently reviewed (15). With this in mind, significant strides have been made in the use of *in silico* approaches to reliably predict dose loading capacity in LBFs (114, 115). This current study sought to advance the application of computational pharmaceutics tools to consider the impact of *in vivo* dispersion of SEDDS on drug solubility in GI fluids. In recognition of the importance of *in vivo* dispersion on SEDDS performance (159) we hypothesised that computational prediction of drug solubility increases seen upon dispersion of SEDDS in simulated biorelevant fluids is likely to be a key performance indicator of whether a SEDDS approach can effectively solubilise the dose *in vivo*. As such, a computationally predicted solubility ratio (SR), based on drug properties in combination with experimentally determined solubility in FaSSIF, would support more informed decisions on formulation options in early development, by allowing estimation of a biopharmaceutically relevant Do.

Resultantly, our hypothesis that a relationship could be elucidated between a biorelevant SR for a SEDDS formulation and drug properties was demonstrated and shown to be robust. We observed, on a dataset of 30 PWSDs using PLS computational modelling, that 5-6 drug properties were sufficient to reliably predict SR upon dispersion of two prototype SEDDS ($logSR_{MC}$ r^2 0.81, $logSR_{LC}$ r^2 0.77). Subsequently, employment of MLR facilitated simplified equations for SR to be generated, requiring only 3 drug properties namely, partition coefficient pH 6.5 ($logD_{6.5}$), melting point (T_m) and aromatic bonds as fraction of total bonds (F_AromB). These represent common drug properties typically identified and integrated into an early-stage pharmaceutical drug profiling environment (54), forgoing requirements for molecular fragment profiling or specialised chemometric software.

Inclusion of drug properties in this computational model, implies their importance to SR upon SEDDS dispersion at a mechanistic level. For the $logSR_{MC}$ model, important descriptors were $logD_{6.5}$, T_m, MW, F_AromB and SssCH2. Additionally, the $logSR_{LC}$ model also included N_AlipR. In terms of, $logD_{6.5}$, T_m and MW, these are widely recognised drug properties from a pharmaceutical profiling context. In both PLS models, $logD_{6.5}$ and MW were positively correlated with SR while T_m was negatively correlated. Inclusion of a partition coefficient descriptor was not unexpected due to addition of lipophilic SEDDS to the media, while $logD_{6.5}$ was previously observed to be strongly correlated with PWSD solubilisation in biorelevant media (243), and an influential descriptor in modelling the $FaSSIF/PhB_{pH6.5}$ ratio (5). Additionally, distribution coefficient has been used to characterise drug release from SEDDS, or more specifically the drug diffusion process from the SEDDS pre-concentrate into aqueous media has been related to $logD_{SEDDS/RM}$ i.e. the distribution coefficient of solubility in SEDDS pre-concentrate and the release medium (253). Conversely, the

negative correlation between T_m and SR is most likely attributable to high T_m molecules exhibiting solid state limited solubility or 'brick dust' drugs, which results in poor solubility in lipid excipients, translating to more modest SR values upon SEDDS dispersion. While the importance of MW as a descriptor is not unexpected given the influence of MW on both crystalline structure characteristics and solvation properties, in contrast to trends observed between MW and aqueous solubility (254), MW and SR in this case were positively correlated. Accordingly, as increasing size negatively influences aqueous solubility, MW may be indirectly conveying information regarding relative drug affinity for lipophilic formulation excipients to that for the comparatively more aqueous environment within the biorelevant medium. Finally, a recent retrospective analysis of selected physicochemical and molecular properties of drugs produced commercially as LBF products versus commercial SD drugs and a database of drugs not produced via either bio-enabling approach, found logD, T_m and MW to be significant descriptors signally commercial success with LBFs (Chapter 1) (222). Similar to this study, increasing logD and MW were found to be significant for LBF commercial success, while a lower relative drug T_m was found to be significant to reach commercialisation (222). The fact that these descriptors were significant in both a retrospective analysis of successfully commercialised LBFs and in this prospective SR prediction upon SEDDS dispersion, re-emphasises their importance as contributing factors to drug-LBF technology success and suitability.

Additionally, SssCH2, F_AromB and N_AlipR were significant in PLS modelling. F_AromB was positively correlated to SR. While this positive correlation is in contrast to previous predictions of HIF solubility (119), it is likely that as increasing aromatic ring count decreases aqueous solubility (81, 255), and an increase in affinity for lipid excipients is seen. In this case, compounds with larger aromatic structures are likely

to have a negative influence on aqueous solubility. Upon SEDDS dispersion, such compounds will associate with greater affinity to the lipid rich microemulsions droplets formed, resulting in a higher SR. However, contributions of aromaticity are likely complex, reliant on numerous factors including attached substituents and their polarity, existence of 'through resonance' with attached substituents, as well as ion-dipole and dipole-dipole interactions with other moieties. Number of aromatic bonds was previously significant for *in silico* prediction of FeSSIF/FeSSIF blank buffer, further highlighting the significance of aromaticity for solubility in media with increasing lipids (243). N_AlipR also influences drug shape and size and is also affected by adjacent moieties. Meanwhile, SssCH2 examines the topological and electronic features of a structure (256) and was previously significant in an *in silico* prediction of solubility in FaSSIF buffer (243).

This work also investigated other factors influencing SR in order to understand of how drugs associate with biorelevant SEDDS dispersions. In terms of drug ionisation, general trends of higher SR for cationic (charged basic) versus anionic (charged acidic) drugs were observed. These observations are in line with previous research where solubility increases in biorelevant media versus corresponding blank buffers for bases and neutral drugs were higher than acids (243). Such increases for cationic drugs, have previously been suggested to stem from favourable electrostatic interactions between negatively charged polar head groups of taurocholate bile salts (257, 258) and positively charged drugs. In this case, such bile salt related electrostatic interactions are likely to occur in both FaSSIF and FaSSIF-SEDDS, with net negative charges observed for all three media. The general trend for increased SR for cationic compounds occurred despite an overall reduction in net negative charge in both FaSSIF-SEDDS media relative to FaSSIF, demonstrating the possibility that

additional electrostatic interactions may exist. As both alterations to droplet sizes and to overall charge of the media upon SEDDS dispersion in FaSSIF versus $PhB_{pH6.5}$ were observed, interactions between the SEDDS and biorelevant solubilising components of FaSSIF are probable. In particular, the negative charges of FaSSIF-$SEDDS_{Migylol812}$ (-5.35 mV) and FaSSIF-$SEDDS_{OliveOil}$ (-5.73 mV), were intermediate of the overall charges of FaSSIF (-14.67 mV) and the values close to zero observed upon SEDDS dispersion in $PhB_{pH6.5}$ (-0.76 mV, -1.27 mV), suggesting surface association of charged bile salts to the oil droplets formed upon SEDDS dispersion. Such an association was previously proposed upon initial *in vitro* dispersion of a SEDDS in a biorelevant media (259). It therefore could be suggested upon SEDDS dispersion, favourable interactions between cationic drugs and these charged bile salts found at the oil droplet surface may help explain the increased SRs observed. Previously, electrostatic interactions between cationic drugs and free fatty acids in post digestive media have also been suggested as a potential mechanism for increased drug solubilisation (159). However, presently such interactions are poorly understood, and electrostatic interactions appear to not be the sole solubilising mechanism involved, given that both neutral and cationic drugs also displayed SRs between 1.1 and 51, hence indicating that there are a number of additional factors governing drug associated with mixed colloidal dispersion.

In terms of excipient effects, $SEDDS_{Migylol812}$ and $SEDDS_{OliveOil}$ were compared. A strong correlation was observed between solubility in FaSSIF-$SEDDS_{Migylol812}$ versus FaSSIF-$SEDDS_{OliveOil}$, suggesting that strong correlations previously observed between drug solubility in MCT versus LCT preconcentrates, and C8 versus C10 triglycerides are also observed upon SEDDS dispersion of these exemplary MCTs and LCTs (114, 238). In all cases solubility in FaSSIF-$SEDDS_{Migylol812}$ was higher than

FaSSIF-SEDDS$_{OliveOil}$, with an overall significant difference observed (* $p < 0.05$). However, the extent of solubility difference between both was relatively small i.e., for 20 out of the 30 drugs the difference was < 20 %. Therefore, in terms of the choice of these exemplary MCT or LCT containing SEDDS, the practical implications in terms of solubility difference on dispersion in biorelevant buffer appear relatively minor. The merits of MCT versus LCT have been widely discussed (22, 260). While in general, drug solubility in most examples of MCTs is higher (14, 261), following formulation digestion, the digestion products of LCT may confer additional advantages (262), while it must also be acknowledged that these trends may not be observed for all MCT and LCTs comparisons. This study also identified two specific drug examples, namely Celecoxib and Venetoclax, where large differences in SR were observed, relating to large solubility percentage differences (58% and 43%) being observed between both SEDDS dispersions. The possible reason for these higher associations with dispersed SEDDS$_{Migylol812}$ for these two neutral drugs is unclear, however this highlights a potential limitation of computational predictions to capture specific drug-excipient solubility effects. Therefore, future work with a wider range of drugs could help to increase robustness of the predictions achieved.

Overall, this work endeavoured to advance the field of computational pharmaceutics by demonstrating the capacity for such predictive tools to inform developability, and specifically to guide formulation decisions regarding SEDDS by assessing their ability to improve the biopharmaceutical dose number. Do$_{(Predicted)}$ (Equation 6) can be easily applied as a computational pharmaceutics tool to guide formulation suitability, requiring only 3 readily obtainable drug properties, in addition to an experimentally determined drug solubility in FaSSIF. The suitability of Do$_{(Predicted)}$ to forecast developability was validated by comparing predicted to experimental Do values,

showing that 8/10 drugs were correctly predicted to transition to a "good solubility" DCS class (I/III). The two drugs, Candesartan Cilexetil and Celecoxib that were not predicted to transition most likely reflect the limitation of the model to capture drug specific solubility increases, as discussed previously. Subsequently, to demonstrate the real-time applicability of such predictions in a pharmaceutical developability context, $Do_{(Predicted)}$ was applied to a drug dataset outside the training and test sets, namely drugs previously successfully produced as commercial LBF products. The $Do_{(Predicted)}$ approach predicted that two out of three (65.4%) of these drugs would offer benefits for development as a LBF. Furthermore, when DCS classes using FaSSIF solubility versus DCS class using predicted solubility with SEDDS were compared, 8 of the 9 commercial drugs which demonstrated no class transition were DCS Class IV. Therefore, as these predictions are based upon drug solubility gains with SEDDS it is likely that permeability considerations, not only solubility benefits, were influential in the development of these poorly soluble and poorly permeable drugs with LBFs.

Comparable to the stated limitation of the original DCS (13), potential for supersaturation was not explored in these predictions. This would have particular relevance for ionisable drugs displaying pH dependent solubility, while weakly basic drugs in particular exhibit higher solubility in gastric media, along with potential for intestinal supersaturation and precipitation. Further limitations of the predictions are also acknowledged in terms of the deliberate omission of exploration of the effect of SEDDS digestion on drug solubility. We therefore acknowledge that this tool is conservative in its approach to solubility predictions and the solubility gains are likely to be under predictive of the kinetic solubility's achieved in the gastrointestinal tract. However, from an industry perspective, where conservative risk:benefit approaches are often applied to formulation development, this low risk approach may be in line

with current industrial preferences. To overcome any conservative nature in the application of a predicted Do, we suggest incorporation of this tool into the refined DCS (rDCS) as part of the initial "standardised investigations" (67). For a weakly basic drug, customised investigations such as the small-scale supersaturation/precipitation experiments as specified in the rDCS could be then triggered to test the potential effects of supersaturation.

Conclusion

Through combinations of *in silico* predictions based on drug properties, and drug solubility screening in FaSSIF, this work demonstrated capacity for computational pharmaceutics to inform drug developability. By applying a computational pharmaceutics approach this study identified drug properties that can be used to predict SR for SEDDS dispersions. The results demonstrated that integration of biorelevant experimentally determined FaSSIF solubility into computationally predicted dose numbers (i.e., combining molecular, physicochemical and biopharmaceutical properties), allows more reliable biopharmaceutically relevant and data-driven decisions to be made on drug-SEDDS developability. While it is acknowledged that *in silico* predictions are not intended to completely circumvent experimental solubility screening, when used in conjunction with appropriate screening assays, such tools can guide likely successful bio-enabling approaches in a biopharmaceutically informed manner. In order to advance this growing field of computational pharmaceutics for LBFs, renewed emphasis should be placed upon creating validated and increasingly robust computational predictions of drug developability with bio-enabling formulations.

Chapter 5: Application of Artificial Neural Networks to Predict the Apparent Degree of Supersaturation in Supersaturated Lipid-Based Formulations: A Pilot Study

Harriet Bennett-Lenane[1], Joseph P. O'Shea[1], Jack D. Murray[1], Alexandra-Roxana Ilie[1,2], René Holm[2,3], Martin Kuentz[4], Brendan T. Griffin[1].

[1]School of Pharmacy, University College Cork, Cork, Ireland.

[2]Drug Product Development, Janssen Research and Development, Johnson & Johnson, Turnhoutseweg 30, 2340 Beerse, Belgium.

[3]Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense, Denmark.

[4]School of Life Sciences, University of Applied Sciences and Arts Northwestern Switzerland, CH 4132 Muttenz, Switzerland

Published in: *Pharmaceutics MDPI, (2021), 13*

DOI: https://doi.org/10.3390/pharmaceutics13091398

Note: *The reader is also directed at this point to Appendix 4-6 where an additional pilot study is detailed using Partial Least Squares Discriminant Analysis (PLS-DA) to classify the drug dataset utilised in this study according to an apparent degree of supersaturation of greater or less than two. This was removed from the main text based on reviewer feedback during the journal peer-review process.*

Abstract

Purpose: In response to the increasing application of machine learning (ML) across many facets of pharmaceutical development, this pilot study investigated if ML, using artificial neural networks (ANN), could predict the apparent degree of supersaturation (aDS) from two supersaturated LBF (sLBF).

Methods: Equilibrium solubility in Capmul MCM and Maisine CC was obtained for 21 poorly water-soluble drugs at ambient temperature and 60ºC to calculate the aDS ratio. These aDS ratios and drug descriptors were used to train the ML models. Accuracy was compared to partial least square (PLS) regression models.

Results: ANN outperformed PLS for both $sLBF_{Capmul}^{MC}$ (r^2 0.90 vs. 0.56) and $sLBF_{Maisine}^{LC}$ (r^2 0.83 vs. 0.62), displaying smaller root mean square error (RMSE) and residuals upon training and testing. Across all models, descriptors involving reactivity and electron density were most important for prediction.

Conclusion: This pilot study showed that ML can be employed to predict the propensity for supersaturation in LBF, but even larger datasets need to be evaluated to draw final conclusions.

Introduction

In the face of increasing pressures for accelerated development, the work of formulation scientists could be advanced through miniaturized screening tools, computational methods, and a structured approach in preclinical testing (16, 263). Currently, more conservative "tried-and-tested" approaches to formulation design are typically employed, often leading to suboptimal formulations that may disregard influential molecular and physicochemical drug properties or compound interactions with formulation excipients. However, such classical formulation development is likely to change as different computational tools are already widely used in drug discovery and are gaining momentum in pharmaceutical development. Quantity structure-activity relationships (QSAR) have streamlined selection of candidates with optimal binding profiles (3), physiologically-based pharmacokinetic (PBPK) models have aided the simulation of pharmacokinetic parameters (197), while theory or data-driven modelling applications have improved formulation development (68, 124, 157, 166, 167, 264-266). Using data-driven machine learning (ML) approaches, improved success rates are achievable by ascertaining statistical relationships between molecular descriptors and the intended response.

The main goal of predicting an outcome using input variables is the same for both partial least squares (PLS) and artificial neural networks (ANN) ML algorithms. However, the mathematical approaches used differ, in terms of dimensionality reduction of data versus potential for non-linear data fitting. PLS is a well-established multivariate regression dimensionality reduction method. The model calculates the X- and Y-matrices to find the principal components in X (independent variables) that capture most of the variance in Y (dependent variable). This initial data is projected into a latent variable space, thereby maximising the covariance between X and Y

(176). While PLS aims to find a linear (or polynomial) relationship between X and Y, ANN represents an emerging ML algorithm. ANN differs in its capability to detect complex non-linear X-Y relationships, while detecting possible interactions between X variables (267). ANN mimic basic human biological information processing methods, as the structure of the multilayer perceptron (MLP) algorithm contains some main elements: input layer, hidden layer, output layer, activation functions and connection weights. Each neuron receives signals/inputs from other neurons in the preceding layers or directly from the independent variables. This signal has an associated weighted value which determines the strength of this interconnection. A weighted sum of these inputs is computed and transformed using an activation function to produce an output signal which is sent to the next neurons in subsequent layers. During training samples are passed through the network and synaptic weights are continuously adjusted until a minimum prediction error is achieved. While an in-depth analysis of ANN can be found in the literature (172, 185), current research suggests that ANN may provide a promising alternative tool to decode complex pharmaceutical datasets.

Over the last decade, interest regarding use of ML algorithms across diverse disciplines in pharmaceutical design and development has grown (68, 158, 163, 169, 216, 217, 268-272). While ML models have been produced to optimise lipid-based formulation (LBF) development (3, 114, 115, 159-163, 235) the application of more novel ML approaches for bio-enabling formulations currently focuses on solid dispersions (SD) (168-170). However, their application to LBFs, in particular for supersaturated LBF (sLBF), remains unexplored. LBFs, in their most utilised form of lipid solutions, aim to solubilise poorly water-soluble drugs (PWSD), and improve biopharmaceutical properties by simulating endogenous lipid absorption pathways

(25). However, commercial utilisation has been declining (222), likely partly attributable to the dose loading limitations given by the inherent drug solubility in the lipid vehicle (33, 34). One delivery solution has involved the development of sLBFs. These are kinetically stable solutions containing a drug concentration above the thermodynamic solubility where increased drug loads and exposure are achieved through thermally inducing supersaturation (33, 273, 274). Previously, supersaturated solutions such as sLBFs have been characterised by the apparent degree of supersaturation (aDS) ratio (32, 275, 276), calculated to determine the propensity of drugs to supersaturate in specific lipid systems (i.e., fold-increase in drug solubility with elevation of temperature). This has been used as an indicator of the likelihood of designing sLBFs and is critical regarding the ability to maintain drug supersaturation upon storage (32). Therefore, we hypothesise that an *in silico* ML model predicting aDS from molecular properties would support streamlined screening of sLBFs.

Consequently, this pilot study sought to investigate if ANN modelling could be used to predict the aDS in sLBFs using a dataset generated for 21 PWSD. PLS regression models produced from the same dataset facilitated a comparison of the two computational techniques for this dataset. Two medium-chain (MC) and long-chain (LC) based mono/di-glycerides formulations were chosen as mono-/di-glycerides systems that previously facilitated improved supersaturation propensity and streamlined drug-excipient screenings (15, 33, 277). PLS has been previously employed in computational modelling for LBF (114, 235). However, this study provided, to the best of our knowledge, the first investigation into the application of ANN to predict maximum dose loading in LBFs.

Materials and Methods

Chemical and Materials

Celecoxib was purchased from Astatech Inc. (Bristol, PA, USA), cinnarizine, JNJ-2A, ibuprofen and itraconazole were obtained from Janssen Pharmaceutica (Beerse, Belgium). Fenofibrate and indomethacine were purchased from Sigma-Aldrich (Ireland). Progesterone, felodipine, sulfalazine, haloperidol, danazol, naproxen, venetoclax, carvedilol, dipyridamole, niclosamide, griseofulvin, fenofibric acid, ketoconazole and clotrimazole, were purchased from Kemprotec (UK), Capmul MCM C8 was kindly donated by Abitec (Columbus, OH, USA). Maisine CC was a kind gift from Gattefossé (Lyon, France). All other chemicals and solvents were of analytical or high-performance liquid chromatography (HPLC) grade, purchased from Sigma-Aldrich (Wicklow, Ireland).

Formulations

Two prototype single component LBF were chosen based on their previous successful applications as sLBF (32). The MC system contained Capmul MCM, a blend of MC mono- and di-glycerides where caprylic acid (C8) is considered the predominant fatty acid. The LC system contained Maisine CC, a blend of LC mono- and di-glycerides where linoleic acid, C18:2 is considered the predominant fatty acid. These formulations are termed $sLBF_{Capmul}^{MC}$ and $sLBF_{Masine}^{LC}$ when referring to solubility testing at 60°C.

Dataset Selection/Drug Physicochemical and Molecular Properties

Twenty-one structurally diverse PWSD were selected (Table 5-1), where criteria included availability of physicochemical properties and potential utilisation as part of a commercial LBFs, or a sLBF. The compounds were classified according to Glass Forming Ability (GFA) (46), where 8 drugs were Class 1, 3 drugs Class 2 and 10 drugs Class 3. Greater than 250 molecular descriptors were predicted from ADMET Predictor 9.5 (Simulations Plus, USA) and added to experimental drug properties of Melting Point (T_m), Glass Transition Temperature (T_g), Entropy of Fusion (ΔS_{fus}), Enthalpy of Fusion (ΔH_{fus}) , T_m/T_g, and Reduced Glass Transition Temperature (T_{rg}), obtained from literature (33, 46, 124, 278, 279). As the molecular properties can be obtained for any drug once the structure is known, they were used as input data.

Equilibrium Solubility Determination

Equilibrium drug solubility studies were conducted in both LBF at ambient temperature (AT) (22°C) and an elevated temperature (60ºC). Solubility at both temperatures for cinnarizine, celecoxib and JNJ-2A were obtained previously (32). Solubilities for the remaining drugs were conducted using an equivalent protocol as follows. An excess amount of drug was added to 2 mL of either Capmul MCM or Maisine CC in screw cap glass vials containing a magnetic stirrer. Resulting suspensions were stirred on a stirring plate (Mixdrive 15, 2MAG, Germany) at 200 rpm and incubated in temperature-controlled ovens (APT.line^{TM} BD (E2), Binder, GmbH, Germany) at AT and 60°C. Aliquots were sampled at 24 h, 48 h and 72 h etc. (or further if required) and centrifuged at 21,380 g (i.e. relative centrifugal force) (Mikro 200 R, Hettich GmbH, Germany) at 22ºC and 40ºC respectively for 15 min. Daily sampling was continued until equilibrium solubility was reached, i.e. solubility

between two consecutive samples differed by less than 10%. The supernatant was centrifuged under identical conditions. To solubilise the oily excipient, the supernatant was diluted 1:10 (v/v) in acetonitrile:ethyl acetate (1:3, v/v), followed by further 1:10 (v/v) dilution with acetonitrile:ethyl acetate (3:1, v/v) and a final dilution with mobile phase. The efficiency of extraction recovery was >94%, tested using a known amount of each compound. All samples were run in triplicate and drug concentrations were determined using an Agilent 1200 series HPLC system. Columns and HPLC testing conditions for each drug can be found in the Appendix (Appendix 4 Table 4-2).

Subsequently, to assess the short-term stability on storage at AT, following the second centrifugation step, an aliquot of supernatant from the 60°C samples was allowed to cool at AT for 2 h. Then sampling and analysis was conducted as outlined above, with values obtained presented as aDS_{2h}. These short-term stability studies were conducted for the majority of the compounds.

Apparent Degree of Supersaturation (aDS)

The apparent degree of supersaturation (aDS) as previously defined (275), was determined as the ratio of the concentration of the drug in the supersaturated solution according to this experimental methodology and the concentration in the saturated solution. This theoretical aDS was calculated according to equation (1) for both sLBF loaded with drug at 60°C:

$$aDS = C_{supersaturation} / S_{equilibrium} \quad (1)$$

where $C_{supersaturation}$ is the concentration of the drug determined after heating the sLBF (to 60°C) and $S_{equilibrium}$ is the equilibrium solubility at AT.

Subsequently, to facilitate comparisons of the short-term stability of the sLBF after 2 hours, a second aDS (aDS_{2h}) was calculated according to equation (2):

$$aDS_{2h} = C_{supersaturation(2h)} / S_{equilibrium} \quad (2)$$

where in this case $C_{supersaturation(2h)}$ is the drug concentration in the lipid system that was heated to 60ºC followed by cooling to AT for 2 hours. The values are reported as aDS (± standard error (SE)) with the SE calculated from equation (3):

$$\mathrm{SE} = aDS\ x \sqrt{\frac{SA^2}{A^2} + \frac{SB^2}{B^2}} \quad (3)$$

where A, B, SA and SB refer to the mean measured solubility values and standard errors for the equilibrium solubility at AT (A) and the concentration of the drug in the lipid system at 60ºC with/ without 2 hours cooling (B). Graphs were obtained using Prism (Version 5, Graphpad, USA).

Differential Scanning Calorimetry

The majority of GFA classifications and T_g values were obtained from literature. However, for fenofibric acid, progesterone and sulfasalazine, this information was obtained experimentally using differential scanning calorimetry (DSC) equipped with a TA Q1000 with a TA Refrigerated Cooling System 90 (TA Instruments, New Castle, DE, USA). The cell was purged with nitrogen at 50 mL/min. After the midpoint glass transition temperature ($T_{g,mid}$) had been determined, crystallization screening experiments were conducted using the protocol by Baird et al. (46). In brief, 2 mg of

drug weighed into a T-zero pan and heated at 10º C min^{-1} to 10ºC above the T_m of each drug (as per Table 1), held isothermally for 3 min, cooled at a rate of 20ºC min^{-1} to -75ºC and reheated to at 10ºC min^{-1} to 10ºC above the T_m of each drug. Sample weights for each repeat sample were within 1 mg and experiments were run in triplicates. GFA was categorised according to Baird *et al.* into Class I (in case of crystallisation during cooling prior to the T_g), Class II (for no crystallisation during cooling, but crystallisation was observed upon reheating above T_g) and Class III (for no crystallisation observed during cooling nor reheating to T_m) (46).

Table 5-1: Selection of the physicochemical and molecular properties of the investigated compounds collated from the literature, predicted from ADMET Predictor 9.5 or obtained experimentally using DSC. AMPH refers to ampholyte.

Drug Compound	MW (g/mol)	clogP	$logD_{6.5}$	Acid/ Base/ Neutral	GFA Class	T_m (°C)	T_g (°C)	ΔH_{fus} (kJ/mol)	ΔS_{fus}*0.01 (kJ/mol/K)	T_m/T_g	T_{rg}	HBA	HBD	RB
Carvedilol	406.49	3.88	2.36	B	III	114.5	41.9	53.00	13.67	1.23	0.81	5	3	10
Celecoxib	381.38	3.81	3.81	A	II	163	58	34.10	7.80	1.32	0.76	4	1	2
Cinnarizine	368.53	4.92	3.98	B	II	121	8.5	37.50	9.50	1.39	0.72	2	0	5
Clotrimazole	344.85	5.08	5.06	B	III	148	30	33.34	7.97	1.39	0.72	1	0	4
Danazol	337.47	4.26	4.26	N	II	225.5	88.3	35.50	7.12	1.38	0.73	3	1	1
Dipyridamole	504.64	3.11	3.02	B	I	163	40.4	72.00	16.51	1.39	0.72	12	4	12
Felodipine	384.26	5.03	5.03	B	III	145	45	30.98	7.38	1.31	0.76	5	1	4
Fenofibrate	360.84	5.20	5.20	N	III	79	–19	33.00	9.32	1.39	0.72	4	0	5
Fenofibric acid	318.76	3.98	1.25	A	I	184	35.4	99.00	21.66	1.48	0.68	4	1	3
Griseofulvin	352.77	2.51	2.51	N	I	245	89	39.12	7.96	1.36	0.73	6	0	3
Haloperidol	375.87	3.82	2.06	B	I	148	33	54.26	12.80	1.38	0.73	3	1	5
Ibuprofen	206.29	3.64	1.69	A	III	77	-45	26.50	7.56	1.54	0.65	2	1	4
Indomethacin	357.80	4.03	1.45	A	III	161	45	37.60	8.64	1.37	0.73	4	1	3
Itraconazole	705.65	4.89	4.89	B	III	168	58	57.60	13.00	1.33	0.75	9	0	10
JNJ-2A	498.90	5.40	5.40	N	III	142	91.2	22.90	5.50	1.14	0.88	4	3	7
Ketoconazole	531.44	3.67	3.51	B	III	146	45	52.85	12.50	1.32	0.76	7	0	8
Naproxen	230.27	3.21	1.10	A	I	152	5.9	25.65	6.03	1.52	0.66	3	1	3
Niclosamide	327.13	4.03	4.02	A	I	230	86	40.70	8.01	1.40	0.71	5	2	2
Progesterone	314.47	3.94	3.94	N	I	130	55.2	23.67	5.87	1.23	0.81	2	0	1
Sulfalazine	398.40	3.15	–0.35	A	I	245	54.6	99.00	20.08	1.58	0.63	9	3	3
Venetoclax	868.46	6.68	6.54	AMPH	III	138	64	18.40	4.50	1.22	0.82	12	3	11

Statistical Analysis

To test the significance between paired solubility values in Capmul MCM versus Maisine CC and $sLBF_{Capmul}^{MC}$ versus $sLBF_{Maisine}^{LC}$, the distribution of the differences was used to determine normality, or lack thereof. A two-sided bootstrap-paired test (5000 samples) determined the significance ($p < 0.05$). Simple scatter plots were produced for Capmul MCM versus Maisine CC and $sLBF_{Capmul}^{MC}$ versus $sLBF_{Maisine}^{LC}$, regression coefficients fitted for interpretation and a bootstrap test for the coefficients conducted. Statistical analysis was conducted using SPSS Statistics (Version 26, IBM Corporation, USA).

Partial Least Squares Regression (PLS)

Quantitative prediction of aDS using PLS regression was conducted using Unscrambler (Version 11, Camo Analytics, US). PLS model development followed standard steps described previously (235). Molecular structures were acquired as smiles from PubChem and used as inputs for the ADMET Predictor (Version 9.5, Simulations Plus, California, USA) to calculate >250 molecular descriptors which were added to T_m,T_g, ΔH_{fus}, ΔS_{fus}, T_m/T_g, and T_{rg} and used as variable inputs. The individual modelling responses were aDS ratios from both $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$. Principal component analysis (PCA) was applied for a randomised assignment of training:test data. Training set criteria were that it covered the chemical space of the test set along with a relatively even spread of aDS ratios. A Hotelling's T^2 ellipse was applied for outlier detection (95% confidence interval). The nonlinear iterative partial least squares (NIPALs) algorithm was utilised, and all variables were mean centred, de-identified and standardized through scaling by standard deviation. To limit overfitting potential, a limit of two principal components was used. Variable reduction was performed as previously described (235) using a Martens' uncertainty

test (250), an important variables plot and correlation loadings plot. Model accuracy was validated by the root mean square error (RMSE) of the training set and test set.

Artificial Neural Networks (ANN)

Multilayer perceptron artificial neural networks (MLP-ANN) were produced using SPSS Statistics (Version 26, IBM Corporation, US) to predict aDS. A partition variable using the same training:test set split was utilised to compare PLS versus ANN. Input properties were obtained as described above and were rescaled through standardisation where values were converted to their z-scores. Hyperbolic tangent was chosen as the activation function for the hidden layer, while an identity output function was used in the output layer (225). Supervised learning using the scaled conjugate gradient (SCG) algorithm was chosen for its speed, and lack of user-critical parameters (226). Batch training was selected due to the relatively small dataset size and the learning algorithm employed. Variable reduction was initially conducted using an independent variable importance analysis. As an arbitrary criterion, only variables with a relative importance of >70% were included in the architecture going forward. Topologies with only one hidden layer were considered, to avoid overfitting. The optimum number of neurons in the hidden layer was identified following a systematic trial-and-error approach were the number of neurons in the hidden layer were manually altered between 2 and 20, with runs performed in triplicate. The optimal network size was chosen thorough minimum RMSE in the training and test sets. The most important variables in each network were elucidated from the normalised importance chart. PLS and ANN models produced were directly compared in terms of different performance evaluation functions including correlation coefficient (r^2), training set RMSE, test set RMSE and residual by predicted charts.

Results

Comparing the Solubility of MC and LC-based LBF and sLBF

Initially solubility in both LBF (Capmul MCM and Maisine CC) at AT and both sLBF ($sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$) at 60°C was compared. Significant differences were seen at AT (* $p< 0.05$) and at 60ºC (* $p<0.05$). The beta coefficients of the regression lines of both Maisine CC versus Capmul MCM and $sLBF_{Maisine}^{LC}$ versus $sLBF_{Capmul}^{MC}$ were also significant (* $p < 0.05$, * $p < 0.05$). A relatively strong correlation was established between solubility (logS) in both blends at AT ($r^2 = 0.84$). This was stronger at 60°C ($r^2 = 0.9$) (Figure 5-1). Fourteen of the twenty-one (66%) drugs demonstrated a higher aDS ratio in $sLBF_{Maisine}^{LC}$ versus $sLBF_{Capmul}^{MC}$ (Figure 5-2). All 21 drugs showed higher solubility in Capmul MCM when compared to Maisine CC at AT. In general, this trend was repeated at 60°C, except for fenofibrate and cinnarizine where the order of solubility was switched, albeit not significantly so.

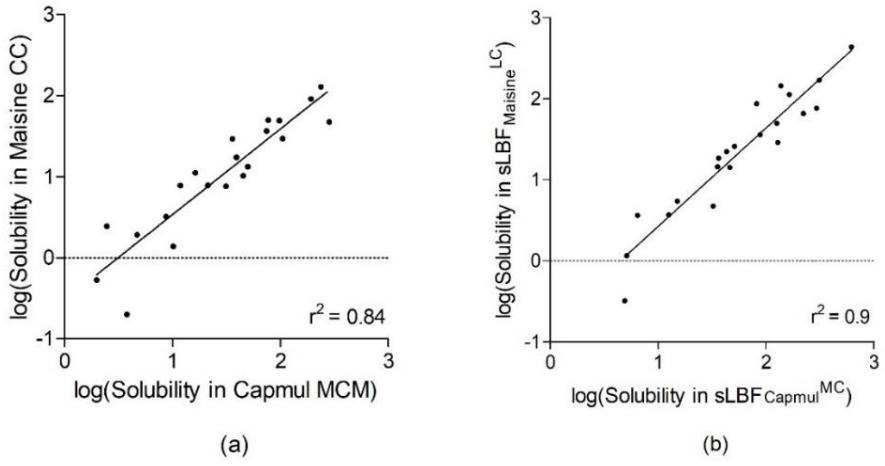


Figure 5-1. Scatter plots of the solubility in Capmul MCM versus Maisine CC (a) and $sLBF_{Capmul}^{MC}$ versus $sLBF_{Maisine}^{LC}$ (b). Formulation abbreviations can be inferred from the main text.

Apparent Degree of Supersaturation

Thermally induced solubility increases were seen for all drugs in both the MC and LC sLBF (aDS ratio >1), reflecting increased dose loading relative to conventional LBFs. Drug solubility in Capmul MCM, Maisine CC, $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$ are presented as mean ± SD (n=3) in the Appendix (Appendix 4 Table 4-1). Extent of aDS ranged from 1.04 to 3.17 in $sLBF_{Campul}^{MC}$ and between 1.06 and 3.4 in $sLBF_{Maisine}^{LC}$ (Figure 5-2). In the rank order of supersaturation propensity, the investigational drug candidate JNJ-2A and felodipine produced the lowest aDS in $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$ respectively. Dipyridamole demonstrated the highest aDS using both sLBF.

While correlations between GFA class and aDS ratios have been previously observed using solvent shift mediated supersaturation (275), our data revealed no clear trend between aDS and GFA (Figure 5-2). The mean aDS for $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$ in each GFA class was class 1 (2.04, 2.08) class 2 (2.22, 2.56), class 3 (2.05, 2.16), indicating that between GFA classes, no significant differences were seen. Mean aDS for the three GFA classes also did not significantly differ according to sLBF fatty acid chain length.

Upon comparison of the aDS values obtained upon cooling of the 60°C samples for 2h at AT (aDS_{2h}) average differences in aDS ratio units of 0.17 ($sLBF_{Capmul}^{MC}$) and 0.16 ($sLBF_{Maisine}^{LC}$) were observed (Appendix 4 Table 4-3). Corresponding to average drug solubility losses of 7.9% and 7.7% upon cooling, respectively. For this dataset which comprised of drugs of a variety of chemical structures, the range of precipitation upon removal of heating was moderate i.e., less than 20%, with 71% of drugs displaying less than 10% loss after 2h.

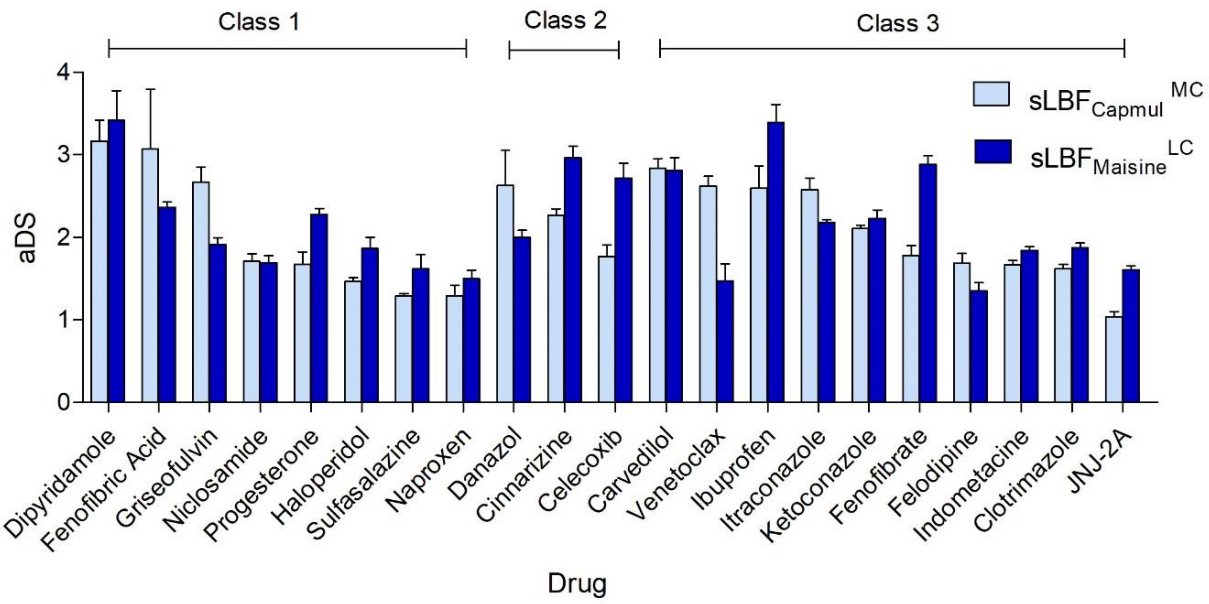


Figure 5-2. Apparent degree of supersaturation (aDS) ratios achieved for the dataset in both $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$. No clear aDS trend was elucidated in terms of the glass-forming ability (GFA) classification (as grouped). Details and definitions of the abbreviations are given in the text.

Quantitatively Predicting aDS using PLS and ANN

Quantitative models predicting aDS were produced using PLS and ANN. Unabridged versions of all the drug descriptor abbreviations in this section are found in Appendix 4 Figure 4-1. PLS models for both aDS $sLBF_{Capmul}^{MC}$ and aDS $sLBF_{Maisine}^{LC}$ of 2 PCs and 8 and 9 input variables respectively were developed (Table 5-2). The aDS $sLBF_{Capmul}^{MC}$ model produced relatively weak predictions of $r^2 = 0.56$, and in the training and test sets the RMSE was 0.4 and 0.79 using 8 variables: VMcGowans, N_Hydrgn, EEM_Afc, EEM_AFnp, SHCH_321, SHaaCH, EEM_NFc and Pi_FMi4 (Figure 5-3). The Martens' uncertainty test designated SHCH_321 and EEM_NFC as the most important variables. Comparatively, the 2 PC aDS $sLBF_{Maisine}^{LC}$ PLS model displayed a correlation coefficient of $r^2 = 0.62$ and RMSE in the training and test sets of 0.4 and 0.45 using 9 input variables: HIVI-TC, N_FrRotB, NPA_Q2, EEM_Nfc,

EEM_NFnp, Pi_Aqo, Pi_AQc, Pi_FPI3 and Pi_FMi6. In this case, N_FrRotB and Pi_FMi6 were the most important variables.

Table 5-2. Overview of the ANNs produced to predict aDS for $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$ from their drug properties, including their architecture and various performance indicators. Tr and Te refer to training and test sets.

Y Variable	Model Type	Architecture	Input Variables	r^2	RMSE Tr	RMSE Te
aDS $sLBF_{Capmul}^{MC}$	PLS	2 PCs	VMcGowan, N_Hydrogn, SHCH_321, SHaaCH, EEM_Afc, EEM_Afnp, EEM_NFc, and Pi_FMi4	0.56	0.40	0.79
aDS $sLBF_{Capmul}^{MC}$	ANN	1 hidden layer, 5 nodes	Pi_FPl5, NPA_Q6, ΔH_{fus}, EEM_F4, EqualEta, M_CX, MlogP, MolVol, N_CYPAtoms, N_Electr, NPA_Q1, Pi_FPl3, Pi_MinQ, S+S_Intrins, and SolFactor	0.90	0.19	0.36
aDS $sLBF_{Maisine}^{LC}$	PLS	2 PCs	HIVI-TC, N_FrRotB, NPA_Q2, EEM_Nfc, EEM_NFnp, Pi_AQo, Pi_AQc, Pi_FPI3, and Pi_FMi6	0.62	0.40	0.45
aDS $sLBF_{Maisine}^{LC}$	ANN	1 hidden layer, 8 nodes	F_AromB, HBDch, MaxQ, N_Atoms, N_Bonds, NPA_Q2, NPA_Q5, Pi_FMi1, Pi_FPl1, SsssCH, and T_Rads	0.83	0.28	0.25

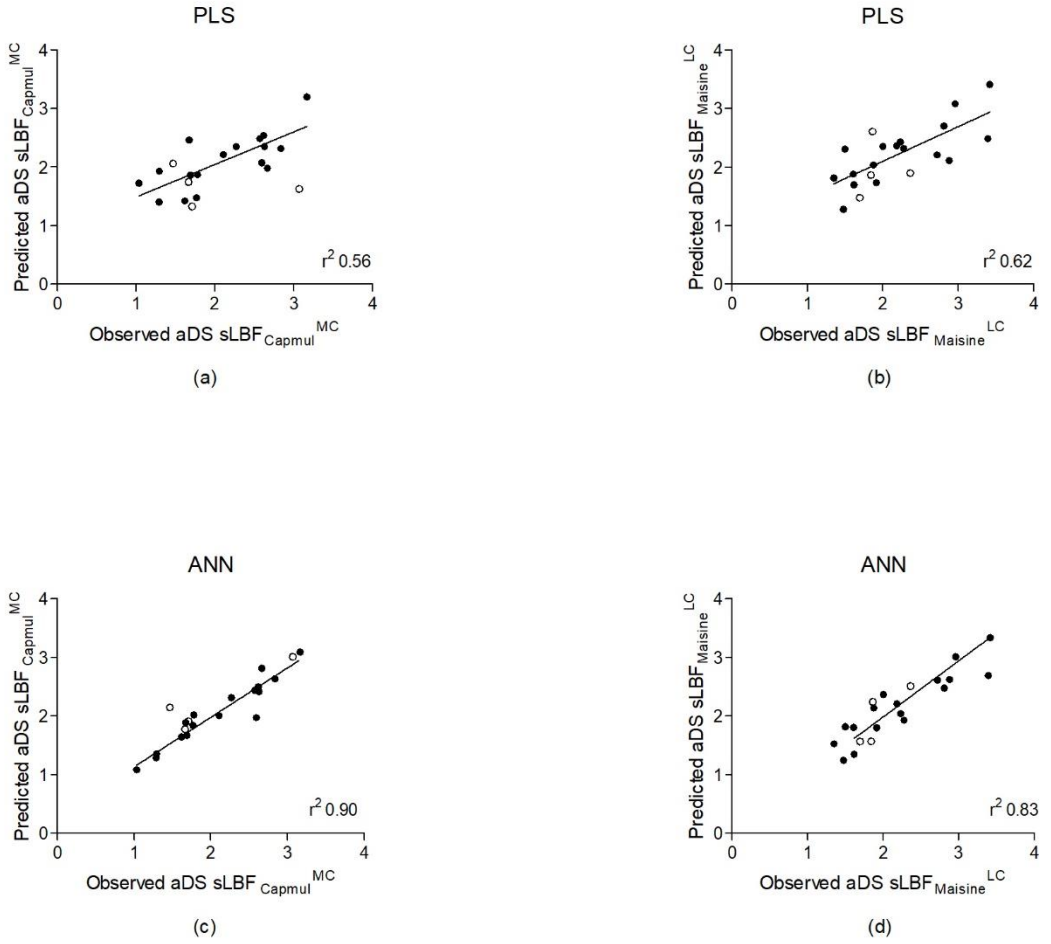


Figure 5-3. Scatter plots illustrating the predicted versus observed aDS values obtained for aDS $sLBF_{Capmul}^{MC}$ using PLS (r^2 = 0.56) and ANN (r^2 = 0.90) (a,c). Scatter plots illustrating the predicted versus observed aDS values obtained for aDS $sLBF_{Maisine}^{LC}$ using PLS (r^2 = 0.62) and ANN (r^2 = 0.83) (b,d).

Using ANN, MLP 15-5-1 for $sLBF_{Capmul}^{MC}$ and MLP 11-8-1 for $sLBF_{Maisine}^{LC}$ networks were produced (Table 5-2). These equated to input layers with 15 and 11 drug properties, one hidden layer with 5 and 8 nodes and singular output layers i.e., predicted aDS. A strong correlation between predicted and observed aDS values was observed for the $sLBF_{Capmul}^{MC}$ network (r^2 = 0.90) (Figure 5-3). This demonstrated low RMSE upon training (0.19) and testing (0.36) (Table 5-2). The properties included in the network were; Pi_FPl5, SolFactor, N_CYPAtoms EEM_F4, Pi_FPl3, NPA_Q6,

MlogP, MolVol, NPA_Q1, S+S_Intrins, EqualEta, ΔH_{fus}, M_CX, Pi_MinQ and N_Electr. The normalised importance chart signified ΔH_{fus}, EEM_F4 and N_Electr as the three most significant variables (Figure 5-4). Predicted and observed aDS values for aDS $sLBF_{Maisine}^{LC}$ were strongly correlated (r^2 0.83), as training and testing RMSE of 0.28 and 0.25 were observed (Figure 5-3). Drug properties in the final network were N_Bonds, Pi_FPl1, T_Rads, MaxQ, N_Atoms, Pi_FMi1, HBDch, F_AromB, NPA_Q2, SsssCH and NPA_Q5. MaxQ, NPA_Q5 and NPA_Q2 were the most important variables (Figure 5-4).

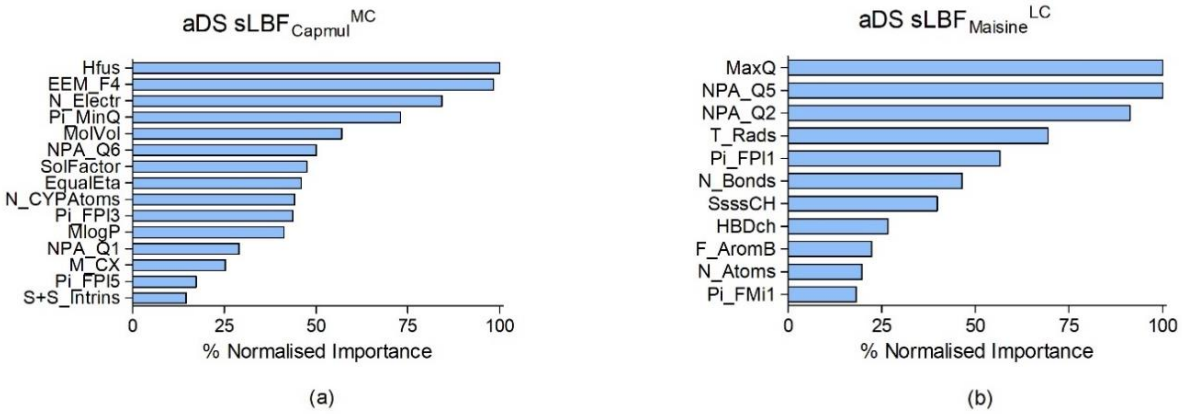


Figure 5-4. Normalised importance charts of the ANNs for $sLBF_{Capmul}^{MC}$ (a) and $sLBF_{Maisine}^{LC}$ (b) detailing the percentage importance of the input variables in predicting aDS. Details and explained abbreviations are given in the main text and Appendix 4 Figure 4-1.

Upon model comparison, the ANN produced improved aDS predictions for both sLBF as both ANN models displayed substantially stronger correlation coefficients, lower training and testing RMSE and smaller residuals. The residuals for both ANN models demonstrated almost complete independence and random distribution in residual by predicted charts (Appendix 4 Figure 4-2). The relatively poor performance of the PLS models indicates that their inclusion was primarily for the purpose of comparison with the ANN.

Discussion

The increasing adoption of model-based approaches across drug design and development has aided increased efficiency in pharmaceutical research. Computational tools exist across the pharmaceutical industry in many forms. However, for LBFs, thus far, drug property-based aspects of computational pharmaceutics have focused on solubility predictions for traditional solution or self-emulsifying drug delivery system (SEDDS) formulations (114, 115, 159, 235). The exploration of ANN to support LBF development remains relatively unexplored. As a result, the main purpose of this research was to investigate if an ANN model could be developed to predict the aDS in sLBF using drug physicochemical or molecular properties. These predictions could be used to guide whether the degree of supersaturation in lipids is sufficient to enable dosing in early development.

Accordingly, as part of this pilot study two ANN were developed which predicted aDS in sLBF from drug properties. These ANN produced superior predictions compared to PLS models developed using the same available dataset. These ANN predicting aDS ($sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$), containing 1 hidden layer of 5 and 8 nodes, and using 15 and 11 drug properties respectively, yielded strong prediction accuracy performance (r^2 = 0.90, 0.83) and low RMSE upon both training (0.19, 0.28) and testing (0.36, 0.25). In comparison, using PLS a lower accuracy of prediction (r^2 = 0.56, 0.62), higher residuals and RMSE upon training (0.4, 0.4) and testing (0.79, 0.45) were observed using 8 and 9 drug properties. Accordingly, this study demonstrates that ANN can be applied to link molecular drug properties to a predicted maximum dose loading capacity i.e., aDS upon thermal induced supersaturation.

These modelling results suggest that aDS prediction is a complex and multifaceted phenomenon, as for this dataset numerous drug descriptors and non-linear

mathematical algorithms were required for higher accuracy. One explanation for the improved performance of ANN for this dataset may be attributed to its capability in decoding multidimensional highly non-linear relationships in datasets in the hidden layer, as opposed to linear relationships of the latent variables obtained through PLS. Consequentially, this work highlights the capability of ANN to provide an industrially applicable alternative to the more established computational pharmaceutics modelling methods such as PLS. While PLS regression has advantages versus ANN in terms of model transparency and decreased complexity in interpretation, in situations of interrelationships or substantial non-linearity as seen here, ANN may improve the accuracy of prediction. Therefore, it is hoped that this pilot study can initiate future larger scale studies to strengthen these predictions.

Modelling indicated that drug properties hold key information about aDS. Overall, a wide range of drug descriptors, reflecting topology, reactivity, structure and size, electrostatics and thermodynamics, were significant. Trends in important properties were revealed. The three most important properties predicting aDS for $sLBF_{Capmul}^{MC}$ were ΔH_{fus} (enthalpy of fusion), EEM_F4 (Fourth component of the autocorrelation vector of sigma Fukui indices) and N_Electr (total number of electrons in a molecule) (Figure 5-4). ΔH_{fus} is a thermodynamic property, involving the amount of thermal energy which must be absorbed or evolved to change 1 mole of a solid to a liquid with no temperature change (280)**.** ΔH_{fus} was shown to previously inversely correlate with potential of a drug to supersaturate from solvent shift induced supersaturation (275). Fukui indices are frontier orbital indices, indicating atomic electron affinity and a molecules ability to become polarized upon changes to electron density (281, 282). Similar Fukui indices were previously important properties governing intrinsic dissolution rate of PWSD in biorelevant media (283) and in support vector machine

modelling to predict GFA for compounds between 200-300 g/mol. In that case a high value, which denoted increased electron reactivity, suggested a non-glass former (124). The number of electrons in a molecule is related to reactivity as electrons in the outermost atom shell determine reactivity. Generally, polarizability increases as the volume occupied by electrons increases. To predict aDS in $sLBF_{Maisine}^{LC}$, MaxQ (maximal PEOE partial atomic charge), NPA_Q5 and NPA_Q2 (fifth and second component of the autocorrelation vector of estimated NPA partial atomic charges) were the most important properties (Figure 5-4). Both natural population analysis (NPA) and partial equalization of orbital electronegativity (PEOE) are methods to calculate partial atomic charges. They describe charge and electron density distributions within molecules, providing clues about chemical behaviour (284, 285). Comparatively, PLS performance was poor in terms of correlation and residual error and therefore more suited here as qualitative models. The fact that PLS and ANN use different mathematical approaches to obtain correlations, and that ANN can incorporate interrelationships between descriptor variables likely explains differences in final model variables. Despite the observed differences, Fukui indices, partial atomic charges and atom type E-state indices were significant for PLS and ANN prediction, supporting their importance for aDS.

As a lack of thermodynamic stability is a fundamental limitation of sLBFs, it is imperative that supersaturation is maintained over a sufficient period to facilitate adequate absorption. In this study, after 2 hours of cooling, the sLBFs maintained relatively high levels of supersaturation across a variety of drugs. aDS was previously suggested as a guide for likelihood of precipitation from sLBF (32), where drugs that generated higher aDS coupled with high T_m/T_g ratios (higher crystallisation tendency), demonstrated quick precipitation on storage at 25ºC, while drugs with low aDS and

low T_m/T_g ratios resulted in good storage stability. Similarly, in this study, dipyridamole a Class 1 GFA drug with a high T_m/T_g and ΔS_{fus} produced the highest aDS in both sLBFs, while Class 3 GFA JNJ-2A and felodipine, both possessing low crystallisation tendencies, produced the lowest aDS. Therefore, this could provide an extended application of these models to anticipate precipitation potential, with reference to indicators of crystallisation tendency (T_m/T_g, ΔS_{fus}) (46). However, investigations regarding the overall accuracy of this combination were not within the scope of this current pilot study.

The influence of fatty acid chain length in terms of both aDS and drug solubility between the MC and LC-based mono/di-glyceride blends was also observed. Like previous work involving MC and LC triglycerides (114, 235), a relatively strong correlation was found between solubility in both blends at AT. Interestingly, it appeared the common effect of heating became more influential for solubility rather than properties of the lipids as heating increased the strength of the correlation. While solubility was higher in $sLBF_{Capmul}^{MC}$ for the majority of drugs, approximately 60% demonstrated higher aDS in $sLBF_{Maisine}^{LC}$. This was potentially aided by the generally lower drug solvation in the long-chain formulation at AT, thereby permitting higher aDS gains upon heating.

Finally, as recent expert commentary has emphasized various shortcomings of data-driven modelling (68), we acknowledge the dataset used in model development here is limited in size. As such, this work was essentially a pilot study seeking to investigate potential for ANN to improve the accuracy of predictive models. Accordingly, the authors support strategies for further research using a larger dataset to confirm the correlations obtained and have produced the ANN models as predictive model markup language (PMML). Increases in the dataset will further clarify which drug properties

are significant for aDS, extending the applicability of the models. Notwithstanding this limitation, this pilot study successfully achieved the intended goal of demonstrating the robust predictive power of ANN to LBF datasets.

Conclusion

This pilot study explored the application of ANN as a computational technique to predict aDS in sLBF. The ANN models demonstrated accuracy in quantitative prediction of aDS ratios versus PLS models from the same dataset. These models, while demonstrating ANNs ability to capture complex data relationships, also facilitated greater insight into the relationship between drug properties and supersaturation propensity. It was revealed that this complex phenomenon is related to molecular descriptors of electron density and chemical reactivity. The study impacts support the application of ML-based computational pharmaceutics in early LBF development testing. Future research with larger datasets will be needed to confirm this pilot study findings. Moving forward, integration and dissemination of computational expertise and *in silico* tools will be vital for efficient decision-making in the development of lipid-based drug delivery systems of the future.

Chapter 6: Exploring Porcine Gastric and Intestinal Fluids using Microscopic and Solubility Estimates: Impact of Placebo Self-Emulsifying Drug Delivery System Administration to Inform Bio-Predictive in vitro Tools.

Harriet Bennett-Lenane[1], Jacob R. Jørgensen[2], Niklas J. Koehl[1], Laura J. Henze[1], Joseph P. O'Shea[1], Anette Müllertz[2], Brendan T. Griffin[1]

[1]School of Pharmacy, University College Cork, Cork, Ireland.

[2]Department of Pharmacy, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen Ø, Denmark

Published in: *European Journal of Pharmaceutical Sciences (2021), 161.*

DOI: 10.1016/j.ejps.2021.105778

Abstract

Purpose: To investigate if solubility and microscopic characterisation of the landrace pig pre-clinical animal models can aid development of bio-predictive *in vitro* screening tools.

Methods: In this work, *post-mortem* gastric and small intestinal fluids were collected in the fasted, fed state and at five sample-points post administration of a placebo self-emulsifying drug delivery system (SEDDS) in the fasted state to pigs. Cryo-TEM and Negative Stain-TEM were used for ultrastructure characterisation. *Ex vivo* solubility of fenofibrate was determined in the fasted-state, fed-state and post-SEDDS administration. Highest observed *ex vivo* drug solubility in intestinal fluids after SEDDS administration was used for optimising the biorelevant *in vitro* conditions to determine maximum solubility.

Results: Under microscopic evaluation, fasted, fed and SEDDS fluids resulted in different colloidal structures. Drug solubility appeared highest 1 hour post SEDDS administration, corresponding with presence of SEDDS lipid droplets. A 1:200 dispersion of SEDDS in biorelevant media matched the highest observed *ex vivo* solubility upon SEDDS administration.

Conclusion: Overall, impacts of this study include increasing evidence for the pig preclinical model to mimic drug solubility in humans, observations that SEDDS administration may poorly mimic colloidal structures observed under fed state, while microscopic and solubility porcine assessments provided a framework for increasingly bio-predictive *in vitro* tools.

Introduction

Trends in physicochemical properties of molecules in the drug development pipelines continuously display an increasing prevalence of poorly water-soluble drugs (PWSD) (2, 3). Resultantly, the pharmaceutical industry must adapt to ensure developability of such candidates. Solubility in the gastrointestinal tract (GIT) is an important parameter in guiding on the oral developability classification, as previous estimates suggest approximately 40% of new chemical entities are rejected in early development owing to insufficient solubility (286). Prevalence of such challenging properties provokes a multifaceted response; including development of bio-enabling formulations, in addition to both development and validation of *in vitro* tools and pre-clinical animal models to accurately forecast *in vivo* behaviour, together counteracting increasing product attrition.

One common bio-enabling formulation strategy involves the use of self-emulsifying drug delivery systems (SEDDS) (222). SEDDS are a type of lipid-based formulation (LBF) composed of an isotropic mixture of oils, surfactants and co-solvents, designed to self-emulsify following dispersion within the GIT. SEDDS include various mixtures of lipophilic and/or hydrophilic surfactants, helping to emulate positive food effects experienced by many PWSD, as concentrations of bile salts and phospholipids are increased in the fed state (257). These endogenous surfactants, in combination with lipid digestion products, may increase the solubilisation capacity in the GIT fluids for PWSD through creation of a range of colloidal structures. In this heterogeneous environment, the solubility deficit between the fasted and fed state can be bridged (287, 288).

Successful application of such bio-enabling oral drug delivery systems is often dependent on existence of efficacious screening processes and predictive *in vitro*

models simulating the GIT. These investigations provide vital tools for progression of bio-enabling drug delivery systems, ideally simulating the likely *in vivo* human response in an efficient and cost-effective manner. However, lack of accurate *in vitro* predictions can result in a reluctance by the pharmaceutical industry to utilise such non-conventional formulation approaches. The past two decades have witnessed a surge in development of *in vitro* models for SEDDS including biorelevant dispersion, digestion and permeability testing, where increasingly detailed simulations of the GIT are seen (52, 243, 289, 290). Additionally, prevalence of computational modelling, as well as *in silico* simulations based on physiologically-based pharmacokinetic (PBPK) modelling platforms, such as Gastroplus, Simcyp and PK-Sim are steadily increasing (61, 114). While collaborative efforts are being made to improve *in vitro* and *in silico* tools (53, 55), the complexity of endogenous formulation processing results in gaps in development of accurate *in vivo* predictions. Resultantly, one method to increase prediction accuracy involves validation and optimisation of *in vitro* simulation conditions, via introduction of increasingly physiologically and biopharmaceutically relevant input parameters.

In addition, pre-clinical animal models provide invaluable early performance indicators for oral bioavailability, formulation performance and impact of dosing conditions (e.g., food effects) (61, 202, 207, 291). Usually this involves collection of plasma samples, but an additional opportunity exists for collation of animal gastrointestinal (GI) luminal aspirates and fluids for solubility screening (292). The utility of the pig model to reliably predict human *in vivo* behaviour has been previously reviewed (291), demonstrating high similarity with human GI conditions and physiology of commonly used breeds such as the domestic miniature-sized pig. However, while similar in anatomy and physiology, the chief principle of utilising

animal models is their ability to provide a reliable estimate of *in vivo* performance of drug delivery systems in humans. It is, therefore, crucial that all biopharmaceutical processes are adequately simulated, including the ability of the intestinal fluids in the target species to provide a comparable solubilisation capacity to their human equivalents (286, 293). Resultantly, previous quantitative assessment of the composition of porcine GI fluids revealed differences in both the concentrations of solubilising components and in the relative quantities of the major bile acids when compared to human intestinal fluids (294). These observations led to the development of a porcine biorelevant medium, fasted state simulated intestinal fluid of pigs (FaSSIFp) based on the composition of porcine GI fluids with respect to pH, buffer capacity, osmolality, surface tension, as well as the bile salt, phospholipid and free fatty acid content in fasted state pigs (294). As these endogenous compounds and their interactions with LBF excipients have been suggested to be vital for solubilisation of PWSD in the GIT (292), further characterisation evaluating similarities and differences in the fluid ultrastructure's formed in both human and porcine fluids through microscopic evaluation is warranted. Furthermore, additional characterisation of fluid structures observed following SEDDS administration may provide insights regarding the capability of SEDDS to mimic post-prandial enhanced solubilisation and improve understanding of the mechanisms by which this enhanced solubilisation is generated.

In response to the necessity for validated *in vitro* models, this research sought to assess if a qualitative evaluation of porcine fluid ultrastructure, as discussed above, in combination with quantitative assessments of drug solubility in these fluids, could inform increasingly bio-predictive *in vitro* tools. In order to achieve this aim, morphological characterisations of porcine luminal media in the fasted and fed state,

as well as at five time points post SEDDS ingestion were conducted at the ultrastructure level. While similar microscopic analysis of both human and simulated fluids have been conducted (257, 295), this research provides the first comparative analysis of porcine fluids using two complementary techniques of Cryogenic Transmission Electron Microscopy (Cryo-TEM) and Negative Stain TEM, previously demonstrated as excellent tools for such analyses (296). These qualitative observations were then compared to *ex vivo* solubility values to investigate any time dependent change in drug solubility post SEDDS ingestion. Fenofibrate was chosen as a BCS Class II neutral drug, and has previously displayed food effect in landrace pigs (207), where the lack of a pH effect between gastric and intestinal samples allowed direct comparisons. Using combined knowledge from the microscopic images and quantitative solubility studies, it was analysed if the maximum observed *ex vivo* solubility with SEDDS could be used to inform more physiologically relevant input parameters, supporting refinement of *in vitro* models for SEDDS.

Materials and Methods

Materials

Fenofibrate was purchased from Kemprotec Ltd. (UK). Hard gelatine capsules (00EL Licaps®) were obtained from Capsugel®. All food components used in preparing the FDA recommended breakfast were purchased commercially. Fasted State Simulated Intestinal Fluid (FaSSIF) and Fasted State Simulated Gastric Fluids (FaSSGF) were produced from FaSSIF/FeSSIF/FaSSGF powder obtained from Biorelevant.com (Croyden, UK). For the fasted state simulated porcine media (FaSSIFp); Lipoid E PC S was obtained from Lipoid GmbH (Germany), Sodium taurodeoxycholate; Sodium hydroxide (NaOH) pellets; Chloroform; Sodium chloride (NaCl); Sodium dihydrogen phosphate monohydrate; Sodium oleate were purchased from Sigma Aldrich (Ireland) and sodium taurocholate was ordered from Thermo Scientific Ltd., Alfa Aesar (UK). Olive Oil, Tween 85 and Kolliphor RH 40 were all purchased from Sigma-Aldrich (Ireland). All other chemicals and solvents were of analytical grade or HPLC grade, respectively, and were purchased from Sigma–Aldrich (Ireland) and used as received. Water of HPLC grade was produced using a MilliQ system (Merck KGaA, Germany).

Gastric and Intestinal Fluid Collection

This study was carried out under a licence issued by the Health Products Regulatory Authority (HPRA), Ireland and EU Statutory Instruments. Local University Ethical Committee approval was obtained. 11 male landrace pigs were sourced locally and housed individually at the Universities Biological Services Unit (17-20 kg and mean 18.3 kg). Pigs were fed approximately 175 g of standard weanling pig pellet feed twice daily and given free access to water. The final feed of 175 g was given 24 h prior to

dosing. Pigs were grouped into either fasted state (3 pigs), fed state (3 pigs) or SEDDS group (5 pigs) for the *post mortem* assessment.

The following *post mortem* fluid collection protocol was repeated for the 3 groups. Firstly, the fasted state group (following a 24 hour fast) received 50 mL of water via a syringe 30 min prior to euthanasia and *post-mortem* sampling. The study was designed to mimic dosage conditions under a fasted leg of a pre-clinical study, therefore 50 mL of water was provided to mimic administration of a dose with water in pigs and access to water was thereafter restricted until sampling, in accordance with the standard protocol applied by Henze et al. (207). The fed state group of 3 pigs, were fed a half portion of a standard high-caloric, high-fat FDA breakfast, the mass which equated to approximately 18–20 g/kg of body weight. The fed group were given this FDA breakfast two hours prior to euthanasia and *post-mortem* luminal fluid sampling, where water was again restricted until sampling. The SEDDS group was orally administered with 1 g of a Type IIIa SEDDS via a dosing device in a gelatine capsule (00EL Licaps®, Capsugel®) followed by 50 mL of water via syringe. 1 g SEDDS was chosen to be representative of a commonly administered amount in comparative animal studies using the landrace pig. SEDDS consisted of 40% olive oil (long chain triglyceride), 40% Tween 85 (co-surfactant) and 20% Kolliphor RH 40 (surfactant). Access to water was restricted up to 3 h post dosing. All pigs were euthanized humanely by intravenous injection of pentobarbital sodium followed by potassium chloride. The peritoneal cavity was exposed by midline incision and the stomach and small intestine were isolated. Occluding ligatures were applied proximal to the cardiac sphincter and distal to the pyloric sphincter as well as at the proximal and distal ends of the small intestine. Once both ends were secured, the stomach and small intestine were removed from the peritoneal cavity. The small intestine was subdivided into three

sections approximating to the duodenum (USI), jejunum (MSI) and ileum (LSI). Gastric, USI, MSI and LSI luminal fluids were then collected and transferred to sterile 50 mL sample tubes at time intervals of 0.5, 1, 2, 3 and 4 h post dosing respectively (n = 1). Further digestion in the samples post-sampling was inhibited with 1 µM orlistat (297). All samples were first immediately frozen at -20 °C, then stored at -80 °C until further analysis.

Cryogenic and Negative Stain Transmission Electron Microscopy Studies

All GI samples (fasted, fed and SEDDS) were centrifuged at 30,000 × *g* for 15 min at room temperature in an Optima MAX-XP Ultracentrifuge from Beckman Coulter (Brea, CA, USA) and the supernatant collected for ultrastructure characterisation. Cryo-TEM samples were prepared by depositing 3 µL of the supernatant (some diluted in ultrapure water to ensure proper vitrification) on glow-discharged 300 mesh lacey carbon grids from Ted Pella Inc. (Redding, CA, USA). Sample vitrification was then carried out in liquid ethane using a Vitrobot Mark IV from FEI (Hillsboro, OR, USA) under controlled (4 °C, 100% relative humidity) and automated conditions (blot time 3 s, blot force '0'). The vitrified samples were then kept in liquid nitrogen and images obtained with an accelerating voltage of 200 kV using a Tecnai G2 20 TWIN Transmission Electron Microscope equipped with a 4K CCD Eagle digital camera from FEI. Negative Stain TEM samples were prepared by depositing 4 µL on glow-discharged 200 mesh carbon grids from Ted Pella Inc. After 60 s, 10 µL of water was added and the grids carefully aspirated using the edge of a filter paper. Gastric samples were then stained with 10 µL of a uranyl acetate solution (pH 2) for 30 s, while intestinal samples were stained with a phosphotungstic acid solution (pH 7). Lastly, the grids were washed twice with 10 µL water and aspirated. Images were recorded

using a CM100 TWIN Transmission Electron Microscope (Philips, Amsterdam, The Netherlands) with an accelerating voltage of 100 kV and equipped with a side-mounted Veleta Camera (Olympus). Use of both techniques allowed for cross-referencing of samples to verify presence of different colloidal structures and increased the robustness of the analysis.

Solubility Studies

Ex vivo apparent solubility studies of fenofibrate were conducted on gastric and USI samples obtained from the SEDDS group at 0.5, 1, 2, 3 and 4 h post sampling, as well as fasted and fed gastric and USI samples. pH of the SEDDS samples was measured using a using a Model 3510 pH/mV/Temperature Meter (Jenway, UK). Fenofibrate was added in excess to triplicate glass vials containing a specified volume of each fluid preheated to 37 °C and a magnetic stirrer. Vials were placed on a stirring plate at 300 rpm (Mixdrive 15, 2MAG, Germany) in a 37 °C incubator for the period of the study. 150 µL samples were removed at 2, 4, 6 and 24 h, with the mean of the 24 h samples used for data analysis. Samples were centrifuged at 11,400 × g for 10 min at 37 °C (Mikro 200 R, Andreas Hettich GmbH & Co. KG, Germany), followed by a 10-fold dilution in acetonitrile. Next the samples were centrifuged a second time to remove precipitated proteins (11,400 × g, 10 min, 4 °C). Supernatant was then transferred to a separate centrifuge tube and suitably diluted in mobile phase in preparation for analysis via RP-HPLC/UV.

In vitro solubility studies were carried out using commercial FaSSGF, FaSSIF along with FaSSIFp previously described (294). FaSSGF and FaSSIF were prepared using the Biorelevant.com protocol (Croyden, UK) and FaSSIFp was prepared using the

previously published protocol (294). Fenofibrate solubility was obtained in FaSSGF, FaSSIF and FaSSIFp, as well as 1:50, 1:100, 1:200, 1:500 and 1:1000 i.e., 1 g:50 mL, 1 g:100 mL, 1 g:200 mL, 1 g:500 mL and 1 g:1000 mL, dilutions of SEDDS dispersed through mixing in the biorelevant medias (n = 3). The same method as above was followed, however, only one centrifugation at 37 °C, 11,400 × g for 15 min was used. The solubility result for 1:200 dispersion of SEDDS in FaSSIF was obtained from a previous publication (Chapter 4) (235).

RP-HPLC/UV Analysis

Detection of fenofibrate was conducted using an Agilent 1200 series HPLC system comprising a binary pump, degasser, autosampler and variable wavelength detector. Data analysis was conducted with EZChrom Elite version 3.2. A Waters Symmetry® C18 column (4.6 × 150 mm, 5 µm) maintained at 25 °C was used during separation. The mobile phase used consisted of 80:20 (v/v) acetonitrile and sodium acetate 25 mM buffer adjusted to pH 5. The flow rate was 1 mL/min and the detection wavelength was 287 nm. The drug concentration in each vial was calculated from a calibration curve run on the same day. The solubility value presented was the mean value of the 24 h triplicate samples. The analysis displayed linearity over the range 0.01-25 µg/mL ($r^2 > 0.999$). The precision of the method at 1 and 10 µg/mL, expressed as the coefficient of variation, was 0.442% and 0.327% within days and 2.796% and 2.92% between days respectively.

Statistical Analysis

Prior to statistical analysis, drug solubility data in the different *ex vivo* and *in vitro* media were compared to Levene's Test for Equality of Variances where a p-value <0.05 indicated a violation of equal variance. Solubility comparisons were conducted using a one-way ANOVA with pairwise comparisons of the groups completed using Tukey's multiple comparison test. All statistical analysis was conducted using SPSS (IBM, California) as a p-value <0.05 indicated a significant result. Graphs of solubility and pH were obtained using Prism 5 (GraphPad Software, CA, USA).

***In Silico* Prediction**

A multiple linear regression (MLR) equation previously developed to predict the solubility ratio (SR) of drugs upon SEDDS dispersion in FaSSIF relative to FaSSIF solubility (235) was employed (Chapter 4). This equation was previously produced using Excel (Microsoft Office, 2016), where correlations were investigated between a selection of drug properties and SR for a database of 30 PWSD, resulting in equation 1:

$$(1)\ logSR = 0.54 + 0.17(logD6.5) + 1.04(F_AromB) - 0.01(T_m)$$

Where logD6.5 is the partition coefficient at pH 6.5, F_AromB is aromatic bonds as a fraction of total bonds and T_m is the melting point of the drug (fenofibrate). The antilog of the result was then multiplied by the solubility of fenofibrate in FaSSIF obtained from literature (249), in order to obtain the *in silico* prediction of fenofibrate solubility upon SEDDS dispersion in FaSSIF and compared to the *ex vivo* and *in vitro* results obtained in this study.

Results

pH Characterisation of Gastric and Intestinal Porcine Media with a SEDDS

In order to assess if administration of a SEDDS would alter the pH profile in the GI fluids of pigs, *post mortem* samples of both the gastric and USI samples were collected at various time points post SEDDS administration in the fasted state (Figure 6-1). In total five pigs were administered a SEDDS formulation, and *post mortem* samples of the gastric and USI samples were collected at 0.5, 1, 2, 3 and 4 hours (n = 1). Due to the limited sample availability, no pH could be obtained for USI at 0.5 h. In terms of the gastric samples, after 30 min the low pH value observed was in line with previously reported values of fasted state pH in landrace pigs (range reported 1.2 - 4.0) (291). However, at 1 h the pH observed was higher (5.38), while values appeared to subsequently fall back to low levels thereafter. However, given that only one sample was available, limited statistical relevance can be derived. The pH of the USI samples ranged from 5.06 - 7.67, consistent with previous fasted-state observations in the landrace pig (291). Overall, it would appear that administration of SEDDS has a limited effect overall on gastric and USI pH over the 4-hour period, and while further studies would be required in a larger number pigs, to assess a transient increase in gastric pH, such studies were not considered justified given the findings of this initial pilot study.

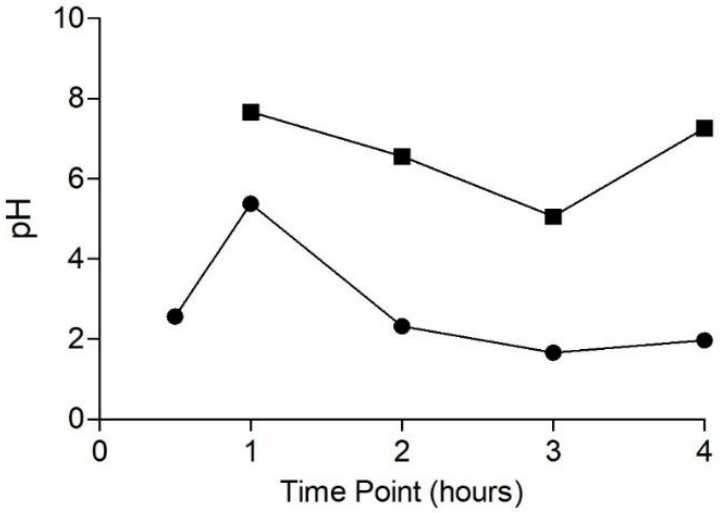


Figure 6-1. pH values obtained as a function of time from gastric (black circles) and USI samples (black squares) following administration of a placebo SEDDS to fasted pigs (n=1)

Microscopic Evaluation of Fasted and Fed State Gastric and Intestinal Porcine Fluids

Two complementary techniques of Cryo-TEM and Negative Stain TEM were used (Figure 6-2). For the fasted-state, Cryo-TEM and Negative Stain gastric images revealed the presence of small structures, which may represent micelles and a vesicle/lipid droplet (~100-150 nm). In the USI fasted state samples, bilamellar vesicles (200 nm), a ruptured vesicle (~400 nm) and small micelle-like structures (10-30 nm) were the predominant features. A MSI fasted sample also revealed an abundance of fiber-like structures. In the fed-state gastric images, while micelles (10-50 nm) were seen, overall, these images displayed evidence of a heterogeneous population with clustering of structures and higher concentrations of colloidal structures of different sizes ranging from approximately 50-500 nm. Structures observed included unilamellar vesicles, multi-compartmental vesicles and multilamellar vesicles and lipid droplets. In terms of the USI images, once again clusters of multivesicular structures including unilamellar, bilamellar and multi-compartmental vesicles were dominant (approximately 100-600 nm). Furthermore, either a mixed micelle or lipid droplet was seen (100 nm), as well as a mixture of micelles and vesicles in the Negative Stain TEM USI and MSI samples.

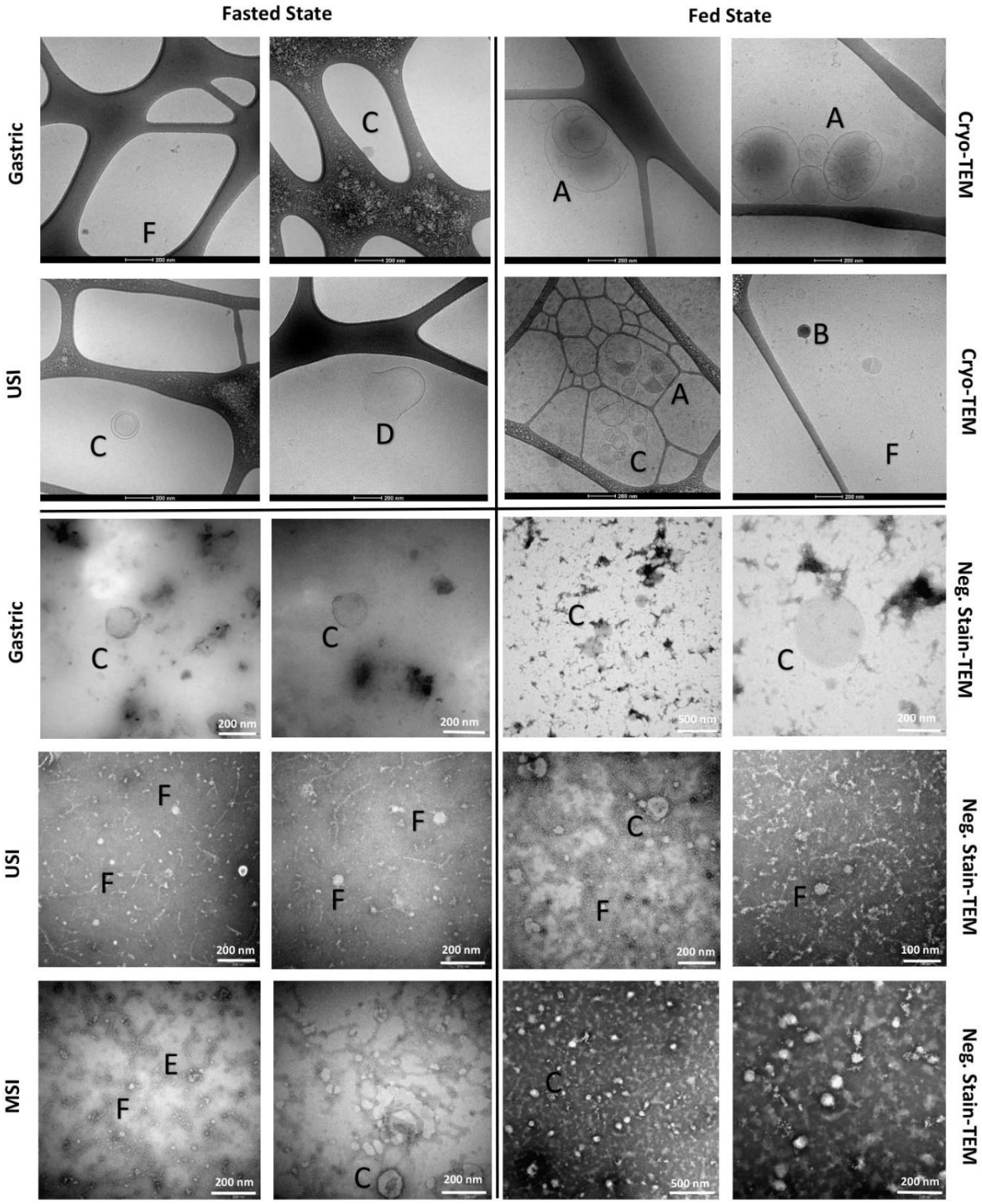


Figure 6-2. Cryo-TEM and Negative Stain TEM images of fasted and fed state (2 h post-prandial) gastric and intestinal samples. Letters indicate representative colloidal structures. A (Multi-Compartmental Vesicles 200-800 nm), B (Lipid droplet), C (Unilamellar and Bilamellar Vesicles or Lipid Droplets 150-400 nm), D (Ruptured Vesicle 400 nm), E (Fiber-like structures), F (Micelles/Small structures 10-50 nm). Measurement scales are shown. To ensure proper vitrification the fed state USI samples were diluted in ultrapure water.

Microscopic Evaluation of Porcine Gastric and Intestinal Fluids Post Administration of a Placebo SEDDS

Changes in luminal fluid ultrastructure were observed at different time points (0.5, 1, 2, 3, 4 h) post SEDDS oral administration in the fasted state. Cryo-TEM images were obtained for the 0.5, 1 and 2 h SEDDS samples only (due to sample unsuitability) and Negative Stain TEM images for each sampling point, except 0.5 h USI and MSI and 3 h MSI (no sample collection) (Figure 6-3 and 6-4). When compared to the fasted state gastric composition, high concentrations of small micelles 10-40 nm are seen in the 0.5, 1 and 2 h Cryo-TEM images, in addition to small lipid structures (20-60 nm) after SEDDS administration. In terms of the USI Cryo-TEM images, the 1 h samples demonstrated examples of what resembled lipid droplet clusters, similar in appearance to structures seen in the 2 h gastric and fed state USI samples (Figure 6-3). SEDDS administration did not appear to produce similar multivesicular structures to the fed state, while a clear difference in composition was observed from the fasted state in terms of higher concentrations of micelles and clusters of lipids droplets, particularly at the 1 h sampling point.

Similarly, Negative Stain TEM revealed differences between SEDDS administration and the fasted and fed state (Figure 6-4). Firstly, similar to Cryo-TEM a high concentration of small lipid structures 10-40 nm was visualised in the 0.5 h gastric image. The 1 h gastric image revealed a heterogeneous mix of structures with a higher concentration of structures versus the fasted state. A large lipid droplet or vesicle, approximately 200-250 nm in diameter, could be observed and was similar in appearance to structures observed in the fed state gastric sample. In the 2 h gastric, USI and MSI samples, vesicles/lipid droplets (100-200 nm) were observed, while the compositional characteristics of the 3 h samples depicted a great mix of micelles and

vesicles of 10-200 nm. Interestingly, the 4 h samples, appeared similar in overall composition to the previous fasted state images, in terms of appearance of small structures resembling micelles (10-50 nm), and fiber-like structures.

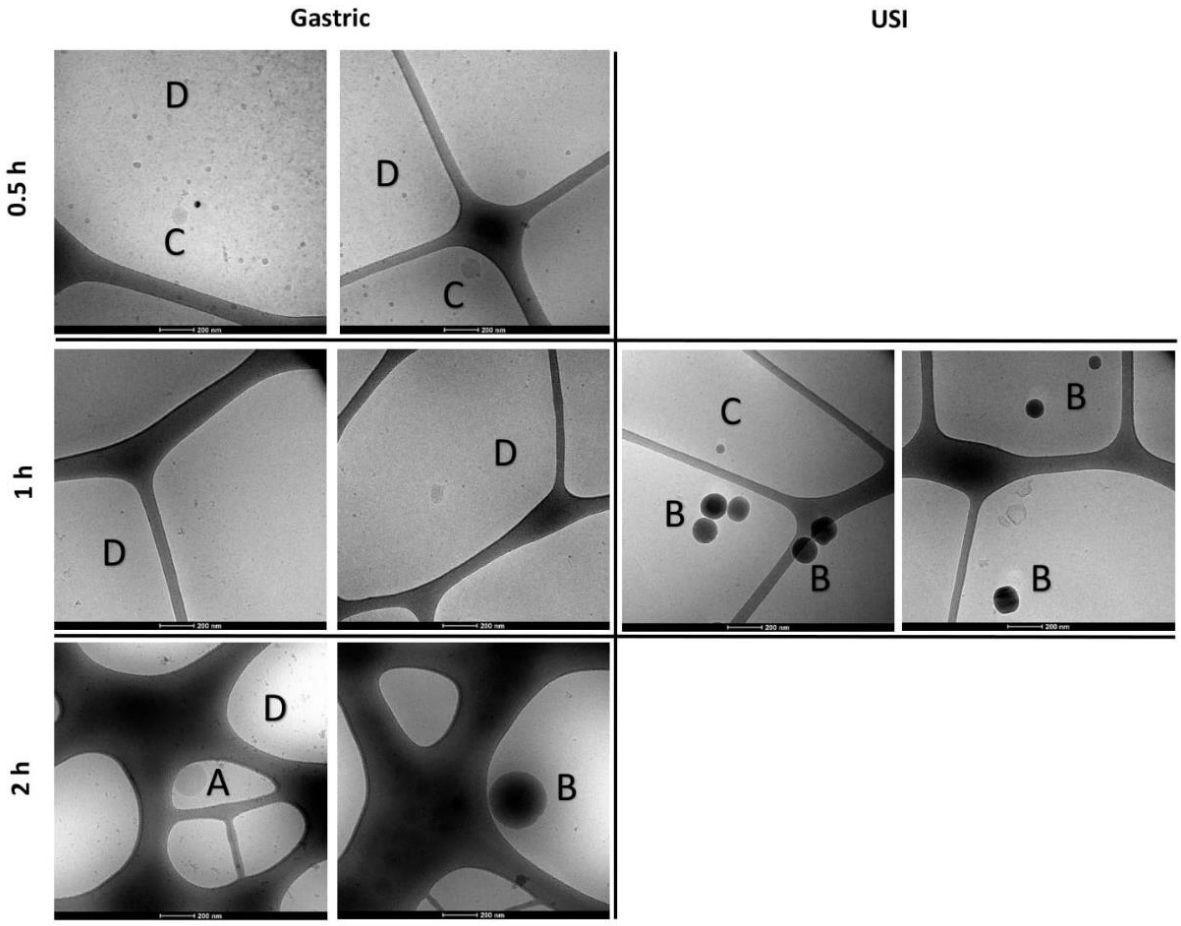


Figure 6-3. Cryo-TEM images of SEDDS Gastric and USI fluids at 0.5, 1 and 2 h post SEDDS administration. Letters indicate representative colloidal structures. A (Unilamellar Vesicles 100-600 nm), B (Lipid Droplet), C (Small Lipid Structures 20-60 nm), D (Small Micelles 10-40 nm). A 200 nm scale is shown for all images. To aid proper vitrification the 1h USI samples were diluted in ultrapure water

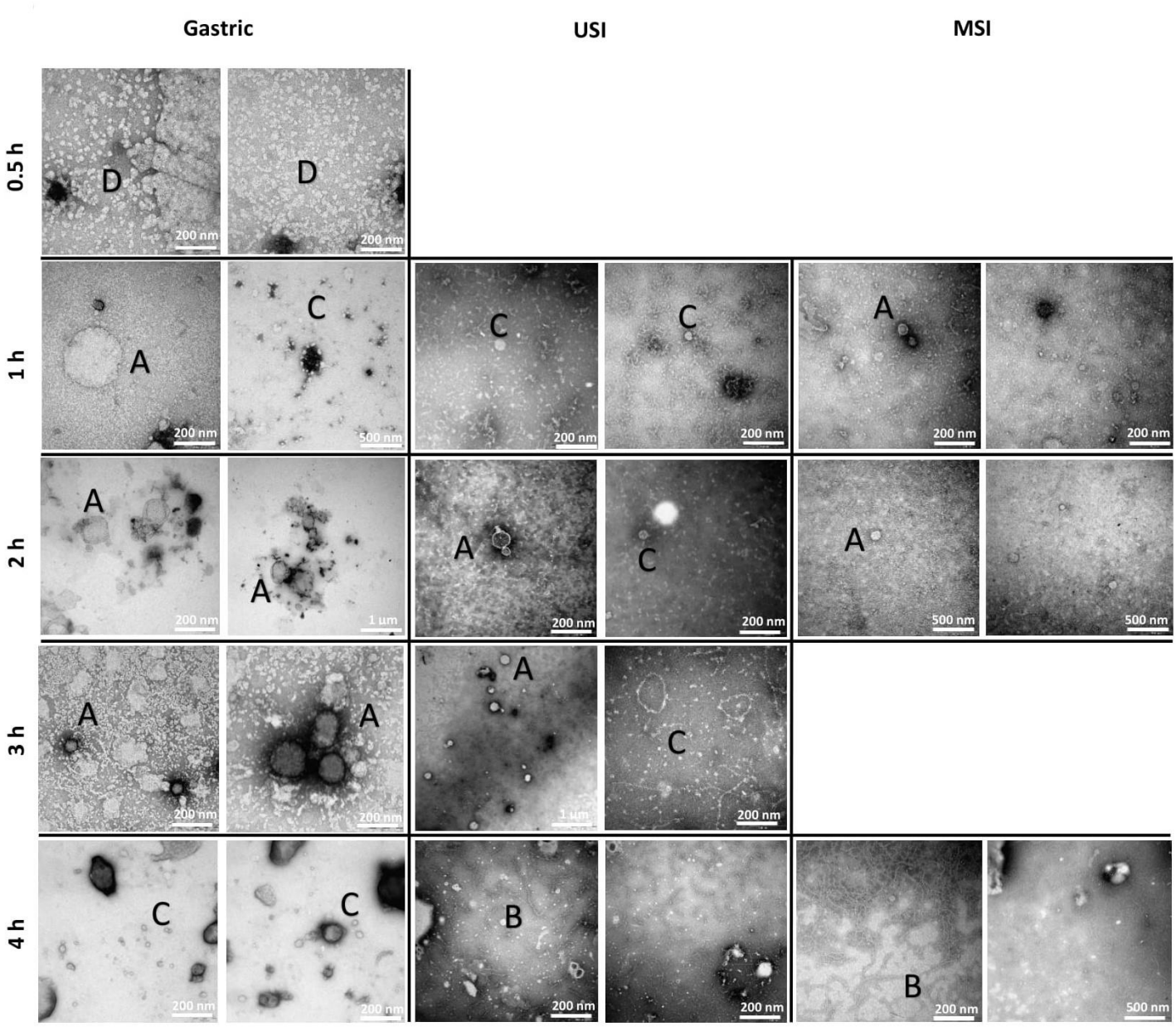


Figure 6-4. Negative Stain TEM images of SEDDS gastric and intestinal fluids at 0.5, 1, 2, 3 and 4 h post placebo SEDDS administration. Letters indicate representative colloidal structures. A (Vesicles 200-400 nm), B (Fiber-like Structures), C (Micelles/Small Structures 10-60 nm), D (Small Lipid Structures 10-40 nm). Measurement scales are shown for each image.

Fenofibrate Solubility in Pig Gastric and Intestinal Media Post Ingestion of a Placebo SEDDS

To investigate potential correlations with the colloidal species visualised upon microscopic evaluation of the SEDDS samples, *ex vivo* solubility studies using gastric and USI fluids were conducted. These results were also compared to the *ex vivo* fed, fasted and biorelevant media solubility results (Figure 6-5). As seen in Figure 6-5, in both the gastric and USI samples, the highest *ex vivo* solubility was observed 1 h post SEDDS ingestion where both samples displayed similar values, while the lowest solubility was at 4 h. In terms of gastric solubility at each time point, fenofibrate solubility increased from 0.5 h (189 ± 10 µg/mL) to 1 h (285 ± 31 µg/mL) post ingestion (Figure 6-5). After the 1 h sample, a decrease in drug solubility was seen, where 2 h (61 ± 8 µg/mL) and 3 h (68 ± 4 µg/mL) samples displayed similar solubility, both below the value obtained from the gastric fed state sample (86 ± 6 µg/mL). Finally, the 4 h sample displayed a low drug solubility (9 ± 4 µg/mL), similar to the value obtained from fasted gastric media (6 ± 2 µg/mL). USI samples displayed a similar solubility trend, though no sample fluid could be collected for the 0.5 h sample thus, was not available for comparisons. Once again, the 1 h time point demonstrated the highest drug solubility (271 ± 36 µg/mL). After this time, drug solubility decreased substantially as solubility at 2 h (117 ± 12 µg/mL) was higher than the 3 h (85 ± 9 µg/mL) sample in contrast to the gastric samples where 2 h and 3 h displayed similar values. In this case, the 2 h sample did exceed the fed state USI solubility obtained (104 ± 19 µg/mL), however not significantly ($p > 0.05$). Again, the lowest drug solubility was observed at 4 h (4 ± 1 µg/mL), also similar to the fasted state USI sample (7 ± 1 µg/mL).

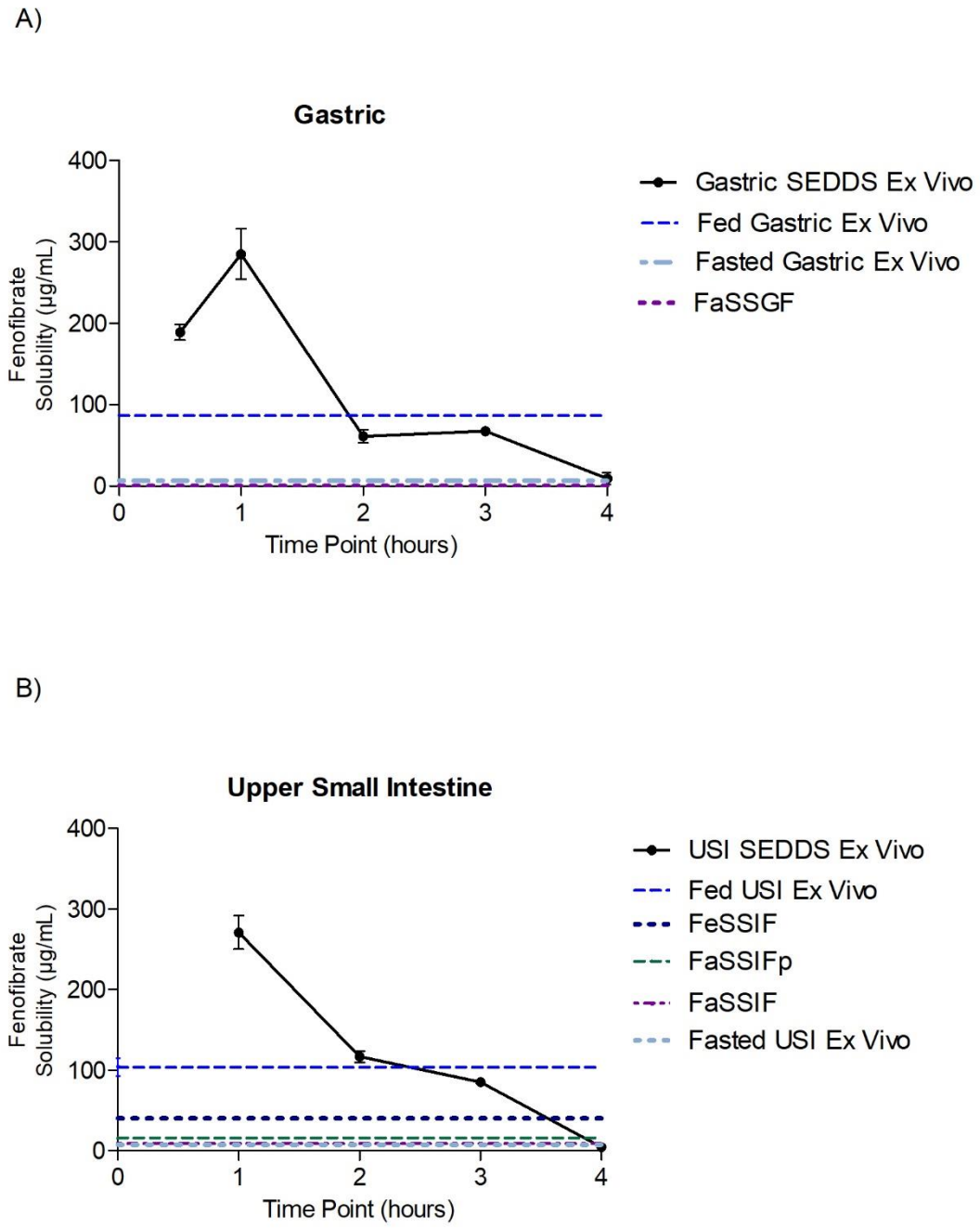


Figure 6-5 A) Fenofibrate solubility in gastric porcine luminal fluids 0.5, 1, 2, 3 and 4 h post placebo SEDDS ingestion for five pigs (Gastric SEDDS Ex Vivo) compared to fasted and fed gastric porcine (2 h post-prandial) and FaSSGF apparent solubility (n =3). B) Fenofibrate solubility in USI porcine luminal fluids 1, 2, 3 and 4 h post placebo SEDDS ingestion for four pigs (USI SEDDS Ex Vivo) compared to fasted and fed USI, FaSSIF, FeSSIF and FaSSIFp solubility (n =3).

Exploration of *in vitro* SEDDS Screening Tool using Enhanced Biorelevant Media and Investigation of Appropriate Dilution Conditions.

Apparent solubility was determined in SEDDS dispersions in simulated human or simulated porcine fluids and compared to the apparent solubility from the *ex vivo* studies where 1 g of SEDDS was administered. The gastric and intestinal *ex vivo* 1 h samples were taken as an approximation of the maximum *in vivo* solubility with SEDDS, and consequently, as the value which the *in vitro* conditions should replicate. Apparent solubility was determined in 1:50, 1:100, 1:200, 1:500 or 1:1000 dispersions of SEDDS in the various biorelevant medias (FaSSGF, FaSSIF, FaSSIFp). All medias containing dispersed SEDDS were significantly different from the solubility of fenofibrate in the respective medias alone i.e., FaSSGF (0.25 ± 0.01 µg/mL), FaSSIF (9.6 ± 1.4 µg/mL) and FaSSIFp (15.69 ± 0.9 µg/mL). One-way ANOVA analysis and a Tukey post-test found that in the three medias, no significant difference was found between the *ex vivo* results and *in vitro* solubility in the 1:200 SEDDS media (Figure 6-6). Conversely, in all three media, the 1:50, 1:100, 1:500 and 1:1000 dispersions significantly differed ($p < 0.05$) from the *ex vivo* 1 h gastric and 1 h USI results respectively. The 1 h USI *ex vivo* solubility and 1:200 *in vitro* solubility results in FaSSIF were also not significantly different from the solubility predicted from a previously published equation which predicts drug solubility gain upon SEDDS dispersion (Chapter 4) (235).

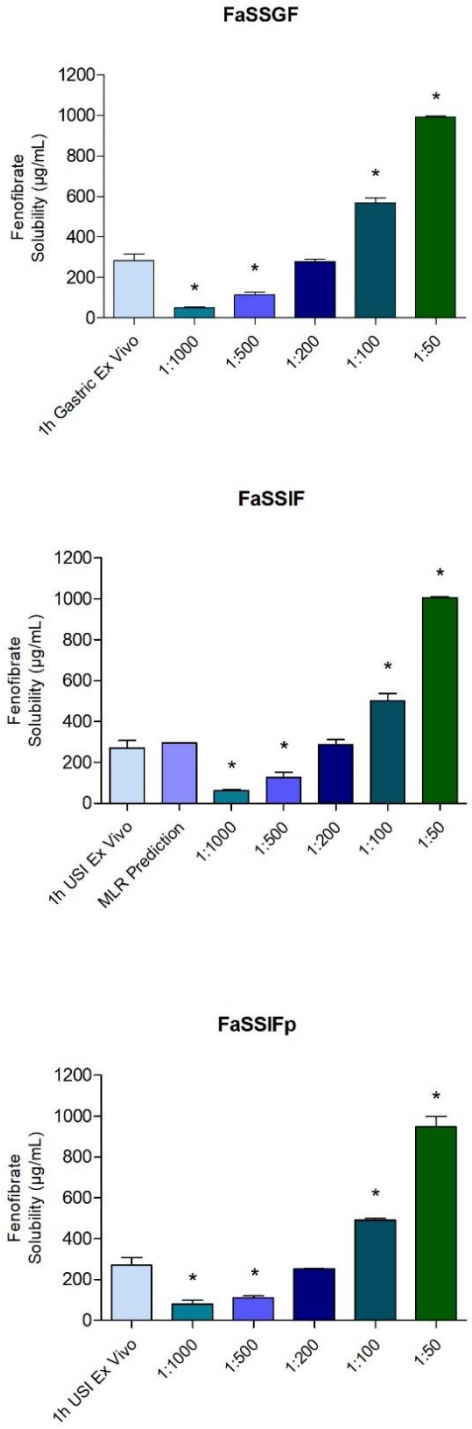


Figure 6-6. Maximum fenofibrate ex vivo solubility in Gastric and USI luminal fluid 1 h after SEDDS administration compared to FaSSGF, FaSSIF and FaSSIFp media supplemented with 1:50, 1:100, 1:200, 1:500, 1:1000 dispersions of SEDDS. A one-way ANOVA and Tukey post-test revealed no significant difference between solubility in the porcine fluids versus the three biorelevant media with a 1:200 dilution. A predicted solubility upon SEDDS dispersion in FaSSIF from MLR also displayed no significant difference from the *ex vivo* USI result. * represents a significant difference ($p < 0.05$) of mean solubility compared the *ex vivo* solubility measurement in each graph i.e., 1 h gastric *ex vivo* or 1 h USI *ex vivo*.

Discussion

The increasing demand for bio-enabling drug delivery systems has generated a complementary necessity for predictive tools to support data-driven, model-informed formulation development. Ability to confidently discriminate formulation performance using *in vitro* tools is key for successful implementation of modern bio-enabling drug delivery approaches. While combined pressures of traditional production familiarity, cost effectiveness and time constraints reinforce the importance of both accurate and efficient tools to encourage adoption of novel bio-enabling technologies. In recent decades, research has focused on standardization and increasing the physiological relevance of such tools (50). However, significant gaps in understanding limit ability to consistently predict *in vivo* drug luminal behaviour. Accordingly, the OrBiTo (Oral Biopharmaceutics Tools) project collaborators have highlighted importance of validation of *in vitro* and *in silico* models, through identifying key *in vivo* processes to be simulated and optimizing experimental inputs to reflect the identified variables (50, 55). Consequently, this research aimed to establish if, through microscopic and quantitative assessment of porcine fluids, *in vitro* simulation conditions could be improved through increasingly physiologically relevant input parameters. It is hoped that such a refinement of *in vitro* conditions can support the developability of drugs with SEDDS, through the facilitation of increasingly accurate *in vitro* dose number predictions.

The first step in achieving this aim involved obtaining an improved understanding of the landrace pig model, via a microscopic characterisation of pig luminal fluid ultrastructure. Aiming to reinforce the utility of the landrace pig model, while also aiding creation of increasingly bio-reflective *in vitro* conditions. Morphological characterisations on gastric, USI and MSI fluids were conducted using two

microscopic techniques: Cryo-TEM and Negative Stain TEM. Firstly, fasted and fed state as ultrastructure's were microscopically compared and contrasted. Distinct morphological differences were observed between the fasted and fed state samples due to increased prevalence of clustering and generally larger structures in the latter. The exact composition of the fasted gastric samples was more difficult to elucidate due to the lack of comparative studies investigating human gastric fluids, perhaps suggesting scope for future research. For example, the somewhat unexpected presence of small micelle resembling structures in both the fasted and 4 h SEDDS gastric samples, is in line previous reports of high bile salt concentrations in the landrace pig stomach compared to humans (294). Most likely reflecting reflux of bile from the pig duodenum to the stomach. In terms of intestinal samples, the fasted intestinal fluids did suggest, in addition to fiber-like structures and vesicles, evidence of spherical micelles, previously shown to be abundant in FaSSIF and fasted state human intestinal fluids (FaHIF) (257, 295, 296). Similar vesicular components of approximately 100 nm have also been found in FaSSIF (247, 257). In contrast, the more complex composition of the fed state samples displayed evidence of clustering and larger multivesicular structures. The heterogeneous fed state presentation was expected as previous research demonstrated large variability in fed state human intestinal fluids (FeHIF) compared to FaHIF ultrastructure (296).

The irregular appearance of some larger structures (>200 nm) has previously been related to their dynamic transient nature as intermediate phases (298, 299). The fed state images in this study resembled the ultrastructure of FeSSIF including micelles and structures ranging from unilamellar to multicompartmental vesicles (296). This is reflective of previous observations that while uni-, bi- and multilamellar vesicles dominate in FeHIF they are rarely seen in FaHIF (295). However, upon comparison

of the porcine fed state intestinal fluid to fed state human intestinal fluids, the structures observed appeared generally smaller in these porcine samples, as numerous elongated structures from 1-10 µm have been observed in FeHIF (296, 298, 300). While similar structures would be expected in general, significant differences in ultrastructure between pigs and humans is likely related to differences in rates of digestion, major primary bile acids (291), total phospholipid and cholesterol concentrations along with differences in bile salt : phospholipid ratios (294). When compared to previous work demonstrating capacity for landrace pigs to predict food effects (207), from these images, it is clear that such an effect is produced through large clusters of heterogeneous vesicular structures in the fed state compared to the fasted state, capable of increasing the solubility of PWSD. Accordingly, an impact of this work is that porcine GI fluid ultrastructure, while demonstrating differences to human and simulated fluids, also shares common characteristics of these fluids, aiding its simulation of human luminal fluids.

Microscopic analysis was additionally conducted to investigate how porcine luminal fluids responded to SEDDS administration. Like the food effect, understanding of the *in vivo* SEDDS solubilisation process is a key consideration for the development of predictive tools. SEDDS performance is often related to a bridging of the fasted-fed solubility gap, therefore, it was investigated if SEDDS administration led to production of colloidal species more closely resembling the fed state. Images of gastric and intestinal media samples taken periodically up to four hours post placebo SEDDS administration revealed time dependent SEDDS processing *in vivo* through differences in colloidal and lipid structures. Compared to fasted gastric samples, high concentrations of small micelles ranging from 20-40 nm and larger lipid structures were seen in the 0.5 h gastric SEDDS sample, resembling higher concentrations of

small micelles previously seen in fed state simulated intestinal fluids (FeSSIF) and FeHIF (257, 296). The 1 h SEDDS USI samples displayed a large unilamellar vesicle and clusters of lipid droplets, similar to the USI fed sample, previously observed in FeSSIF, FeSSIF-V2 and FeHIF (296, 300). While a similar lipid droplet was seen in the 2 h gastric sample. From these images it appears that the SEDDS droplets formed ranged from 100-200 nm approximately. This can be compared to smaller previous size estimates of 44.76 ± 0.303 nm and 51.7 ± 0.8 nm observed for this SEDDS dispersed in FaSSIF and simulated gastric fluid without pepsin (SGFsp) respectively (51, 235). However, effect of digestion on droplet size was not accounted for in these previous studies. This along with differences in bile salt and phospholipid constituents and concentrations between the *in vivo* porcine and human simulated media, likely explains the smaller sizes seen.

From these images it appears that SEDDS administration does not result in colloidal structures of the same complexity and size of fed state fluids, with the differences in types and concentrations of lipids present, along with the added complexity of the surfactant present in the SEDDS samples playing a role. In contrast, clusters of lipid droplets from SEDDS processing were seen, predominantly in numerous 1 h USI samples. This can be compared to previous work where small micelles and clusters of lipid droplets were microscopically observed at the beginning of lipolysis during *in vitro* SEDDS digestion and fewer lipid droplets were seen as time progressed, suggesting complete digestion (259). Clear differences between the 0.5 h to 4 h SEDDS samples were seen. When compared to the fasted state Negative Stain TEM images, similarities were perceived with the 4 h SEDDS images in terms of the predominant presence of fiber-like structures and small structures resembling micelles. Overall, this microscopic assessment suggests that the colloidal structures

formed post SEDDS ingestion, while demonstrating increased colloid numbers relative to the fasted state, appear less complex than fed state fluids. While other physiological effects such as transit times may also play a role in differences seen in the media.

Following the observation of varied colloidal structures in the gastric and intestinal fluids, and presence of lipid droplet clusters in the 1 h USI sample during microscopic analysis, it was then assessed if these qualitative observations could be correlated to time dependent changes in fenofibrate solubility from 0.5 h up to 4 h post SEDDS administration. Accordingly, the differing drug solubility's seen at the various time points can be related to time dependant digestion of the vehicle and potential on-going lipid absorption. Solubility in both the gastric and USI samples was highest at 1 h post administration, exceeding fed state solubility, before decreasing at 4 h to levels similar to fasted state gastric and USI fluid and simulated fasted media. Resultantly, for fenofibrate, which displays a high dose loading in this SEDDS (96.6 ± 3.4 mg/mL) (51). It appears to be the presence of the SEDDS lipid droplets, as seen in the 1 h USI images, which are likely to be the key reservoirs of drug, maintaining high solubilisation capacity.

In summation, the images 1 h after SEDDS administration, the time of maximum observed *ex vivo* drug solubility, appeared different in colloidal ultrastructure composition, compared to the fasted state fluid, and also the fed state fluids 2 h after feeding. In the fed state, the lipid load produced is likely higher and more diverse, reflecting the complexity of the meal composition, likely contributing to differences observed with the SEDDS samples. While it appears from this study that the increased fenofibrate solubilisation using SEDDS is primarily driven through presence of SEDDS lipid droplets. It must be acknowledged that a different SEDDS may produce

a different outcome in such a study and these results represent a snapshot reflecting the ingested type and amount of SEDDS ingested. Any solubilising effect of the SEDDS was lost at 4 h, where fenofibrate solubility mirrored fasted state gastric and USI values. In agreement, the Cryo-TEM 4 h images resembled the fasted state, with these results both suggesting that by this time the GIT had sufficiently processed the SEDDS, and digestion and absorption was complete. A contributing factor may also be that from 3 h after SEDDS administration pigs received *ad libitum* water access, potentially accelerating flushing of the SEDDS at 4 h.

Previous research has investigated the drug solubilisation effect of carvedilol in canine intestinal fluids, where administration of 2 g of LBF resulted in a significantly higher solubility in fluids collected 5-20 min after administration versus 1 g of LBF or water (301). This higher solubility was only seen for the 5-20 min samples, and not in samples taken at 0-5 min or 20-90 min after LBF administration. Differences in the timeframe of maximum observed solubility between these two studies is likely as a result of the previously discussed differences in absorption rates between canines versus pigs. Where the rate of drug absorption and gastric emptying in pigs is suggested to be marginally slower than canines (291). Meanwhile, in this study, solubility in fasted and fed state *ex vivo* USI samples, 7 ± 1 and 104 ± 19 µg/mL respectively, appeared similar to reported human values for fenofibrate solubility in FaHIF (20 ± 26 µg/mL) and FeHIF (148 ± 60 µg/mL) (249). Overall, these results reflect the fact that while inter-species differences are inevitable and some weaknesses may exist for use of the porcine model, merits for its ability to provide close predictions of human solubility are clear.

Stemming from the combination of qualitative and quantitative assessments of porcine fluids post SEDDS ingestion, solubility at 1 h post SEDDS ingestion appeared highest

for both gastric and USI fluids. Therefore, it appears that in order to accurately represent the maximum solubility estimate for dose number solubility classification, as outlined in the developability classification system (DCS), use of this 1 h media would provide the best solubility approximation when a SEDDS approach is considered. Accordingly, using knowledge obtained from the porcine fluid assessments in this study, *in vitro* testing was conducted to assess if this solubility estimate could be closely replicated *in vitro* using supplemented biorelevant media under optimal screening conditions. While use of *in vitro* models provide welcomed resources to predict *in vivo* formulation performance, they remain only as accurate as the precision of the experimental parameters upon which they are based. Even though appropriate simulation of *in vivo* relevant fluid volumes has been suggested to be critical for correct implementation of bio-predictive tools (302), typically, testing parameters utilised, including dilutions, are taken from previous research and repeated, which may not represent an accurate bio-simulation of the conditions being replicated. Resultantly, it was hypothesised that a refined biorelevant medium reflecting the highest solubility observed from the microscopic and quantitative assessments in this study, could provide a more predictive and physiologically relevant estimation of how the GIT responds to SEDDS ingestion. In order to investigate which *in vitro* dispersion conditions provided the closest estimate of the 1 h *ex vivo* result, five different ratios of SEDDS dispersed (1:50, 1:100, 1:200, 1:500 and 1:1000) in three different biorelevant media were tested. These dispersions were selected as being reflective of the current physiological volumes suggested for the human small intestine, which vary from approximately 50-1100 mL (303, 304), while approximation is complicated through complementary presence of absorption. While the testing dilutions also approximated the 250 mL BCS solubilisation parameter (64)

and 500 mL DCS dose number solubilisation parameter (13), they also considered dilutions typically used for *in vitro* dynamic dispersion testing using dissolution testing apparatus (USP 2) (51, 290). As such, this also aimed to provide justification for traditional dispersion practises used in current *in vitro* tools to reflect the maximum solubilisation effect of 1 g of SEDDS administered in an *in vivo* study, a typical desired dose for humans.

Accordingly, this work succeeded in verifying that a 1:200 dispersion of SEDDS in biorelevant media provides the closest simulation of maximum *ex vivo* solubility after 1 h upon administration of 1 g of SEDDS. SEDDS dispersed 1:200 in all three media (FaSSGF, FaSSIF and FaSSIFp) displayed no statistically significant differences from the *ex vivo* 1 h SEDDS solubility value in gastric and USI porcine fluids (Figure 6-6). Suggesting that biorelevant media containing 1:200 dispersed SEDDS should be used to accurately reflect likely maximum solubility *in vivo* after 1 h when 1 g of SEDDS is administered. Furthermore, as no statistically significant differences were found between solubility in the (1:200) FaSSGF, FaSSIF, FaSSIFp, or the *ex vivo* 1 h SEDDS gastric and USI samples, the earlier hypothesis, from the microscopy and solubility assessments, regarding the importance of the SEDDS droplets for solubilisation was reinforced. As these similar solubility values suggest that the SEDDS excipients are primarily driving fenofibrate solubility in the SEDDS dispersions, in contrast to any altered concentrations of bile salts and phospholipids or products of digestion in the respective medias. Additionally, accuracy of a previously published *in silico* tool for predicting solubility gain and resultantly dose number and DCS classification upon SEDDS dispersion in biorelevant media (Chapter 4) was verified when the predicted value (297 µg/mL) was similar to the *ex vivo* (1 h USI) and *in vitro* (1:200) solubility estimates (Figure 6-6). Therefore, accurate prediction of the *in vivo* and *in vitro*

measurements through this *in silico* modelling reinforces its applicability to reliably predict dose numbers with SEDDS. Overall, an improved *in vitro* screening tool using appropriately concentrated SEDDS dispersions accurately predicted maximum *in vivo* drug solubility, demonstrating the significance of *in vitro* tool validation.

Conclusion

Overall, implications of this work for wider research are numerous. The ability of microscopic and solubility analysis of porcine fluids to refine *in vitro* predictions with SEDDS was realised upon demonstration that solubility at 1 h post SEDDS administration was closely matched by a 1:200 dispersion of SEDDS in various biorelevant media. Resultantly, this study demonstrates that tailoring of formulation screening with refined bio-relevant inputs is of the utmost importance. While there will always remain a need for certain *in vivo* studies, characterisation of these systems can lessen dependence and aid progression to a more confirmatory, rather than exploratory role, via improvements in the predictive power of *in vitro* tools. Furthermore, this study represents the first characterisation of GI colloidal phases in pigs using advanced microscopic techniques, forming the basis for a better understanding of the landrace pig as a model for evaluating drug bioavailability from SEDDS, while providing increasing evidence for its close representation of human solubilisation capacity. Overall, through integration of qualitative and quantitative *ex vivo* porcine GI fluid characterisations and solubility estimates with *in vitro* tools, this work has demonstrated that refined predictions of *in vivo* drug solubility are achievable.

General Discussion

Overview and Summary

Declining R&D efficiency of new drug candidates has provided an incentive to reinvigorate approaches that improve drug developability (9, 16, 67). Since the late 20th century, computationally driven drug design has facilitated streamlined identification of candidates with optimal binding properties (1, 2). However, despite the drug design advances achieved by a "structure-based analysis", these candidates often display less than optimal aqueous solvation properties. This renders such drug candidates difficult to deliver using conventional formulations and candidate developability is often closely associated with the ability of bio-enabling formulations to rescue their delivery. As a result, successful adoption of these bio-enabling approaches, including LBFs, can only be achieved through scientifically informed decision making and optimised screening of potential candidates. Considering that only small quantities of a drug substance exist in the early pre-formulation stage, use of extensive screening, laboratory and monetary resources is not an optimal selection approach (3). Incorporation of predictive biopharmaceutic tools in these established screening settings can assess a candidate 'fitness for purpose' to achieve adequate exposure when formulated. By using such predictive tools, only drug candidates that show positive results *in vitro* or *in silico* will undergo subsequent laboratory testing, resulting in less material waste and associated costs.

Accordingly, one means to identify or predict suitable candidates is through the employment of computational pharmaceutics approaches. Computational pharmaceutics is an area of growing interest (132) as the digitalisation of pharmaceutical sciences, which encompasses many AI techniques (133), is a

promising means to support decision making in drug development. Computational approaches can be initiated with either a well-established understanding of physiological processes (theory-driven modelling) or by deriving statistically significant correlations between an indicator of interest and independent variables (data-driven modelling). With reference to this ever-expanding computational research area, Chapter 1 and 2 of this thesis illustrate how decisions regarding both formulation suitability and design are increasingly being informed by expert learning approaches. These include formulation DTs, retrospective statistical analyses, or AI methods, where the importance of both biopharmaceutical parameters and drug properties provides the basis for many investigations. However, such successful anticipation of developability, or more specifically as demonstrated in this thesis, "formulat-ability" is not straightforward. It requires a combined approach using a myriad of predictive *in vitro* and *in silico* tools to break away from the significant amount of trial-and-error still applied. Considering the limitations of resource-intensive experimental screening and the emergence of computational pharmaceutics, an optimised computational pharmaceutics-aided drug development paradigm shift can be achieved through:

1) Development of data-driven computational tools to predict quality target product profile (QTPP) characteristics (e.g., food effects) which inform early drug development decisions.
2) Adopting computational tools into the current development setting which predict indicators of drug developability for optimal bio-enabling formulation selection.

3) Development of reliable experimental biopharmaceutics *in vitro* tools which validate and confirm these computational pharmaceutics tools in accurately anticipating *in vivo* performance.

In light of these factors, the specific objectives of this thesis involve the development of drug property-based computational models and *in vitro* screening tools to inform drug developability. This thesis aimed to place a spotlight on the potential of emerging ML approaches to inform formulation selection. Specifically, we have assessed the capacity for ML algorithms to predict both QTPP characteristics, i.e., food effect, which typically dictate early drug development pathways, as well as tools to identify suitable candidates for LBF bio-enabling delivery systems. Special emphasis is placed on drug physicochemical and molecular descriptors as predictive modelling inputs and the potential application of ML techniques to a broad variety of biopharmaceutics datasets.

In Chapter 3, two ML algorithms were compared for the prediction of FE on bioavailability, identifying a convenient means of highlighting drug delivery problems associated with variable drug plasma levels upon concomitant food ingestion. While it is widely acknowledged in literature that FE prediction is difficult (200), such ML models can be added to the growing number of FE predictive tools to advance this research area (201). As delivery of drugs displaying significant FE is often facilitated through bio-enabling approaches, such as LBFs and SDs, Chapter 1 used a retrospective statistical analysis of commercialised drug products to identify drug property trends which may signal increased suitability of a drug for a LBF or SD approach. Both Chapter 4 and Chapter 5 subsequently focused on the development of ML computational tools predicting developability indicators which signal drug suitability for LBFs. Special emphasis was placed in these chapters on the solubility

gain achieved upon dispersion of a SEDDS in intestinal fluids (Chapter 4), as well as the maximal dose loading achieved via a thermally induced supersaturation approach (Chapter 5).

Subsequently, *ex vivo* solubility and microscopic assessments carried out as part of Chapter 6 investigated the capacity for the pig pre-clinical model to not only validate examples of the previously obtained *in silico* predictions from Chapter 4, but also to produce increasingly bio-predictive *in vitro* screening tools for LBFs. This is in acknowledgement of the fact that not only is bio-prediction a significant factor to ensure confidence can be placed *in vivo* performance simulation capacity (49), but also that *in vitro* testing remains an important aspect of pre-formulation decision making. Reliable experimental testing is still required to confirm any promising *in silico* results prior to significant drug product investment. Considering the results obtained in this current thesis, a refining of the wider drug substance to drug product developability framework incorporating both computationally informed and experimentally confirmed facets of drug developability testing was proposed. The approaches involved in assessing the aims of this thesis have been described in detail in the preceding chapters, and an overall, general discussion of the findings is now provided.

Model Performance and Lessons Learned from Computational Modelling

Model Performance Comparisons

In response to the major aims of this thesis, this work has succeeded in demonstrating the capability of ML algorithms to identify suitable drug candidates for LBFs. Throughout this thesis data-driven computational models using numerous algorithms including MLR, PLS, PLS-DA, SVM and ANN were used to predict various biopharmaceutical properties of interest. In light of growing 'big data' opportunities, availability of molecular descriptors, and the desire to avoid any previously hypothesised theoretical concepts providing bias in prediction, the current models were developed using drug properties as inputs where statistically significant relationships were obtained with the intended dependent variables. Examples of both classification (Chapter 3, Appendix 4-6) and regression models (Chapter 4, 5) were produced. These demonstrated the applicability of ML to numerous scenarios depending on the stage of drug development or type of prediction required. Classification models provide worthwhile early information to formulators guiding initial decisions, while in later stages of screening, regression models, like those in this current thesis, can provide valuable quantitative outputs to compare bio-enabling applications.

As described in Chapter 2, consistent model development and prediction validation is imperative to encourage wider application of computational pharmaceutics tools. As the predictive outputs and accuracy of these models provide the major results from this work, a comparison of the accuracy of the predictions to the experimental input values is shown in Table 8-1. It is acknowledged that singular predictions do not give the best estimation of model performance, however, the percentage difference of the

predictions, using fenofibrate as a model drug, from each model were used for illustrative purposes. The PLS-DA classifications of aDS groups (>2 or <2) are not illustrated in Table 8-1 as their development was primarily for trend analysis and no test set was used (Appendix 4-6). However, fenofibrate aDS was correctly classified for both aDS $sLBF_{Capmul}^{MC}$ and $LBF_{Maisine}^{LC}$ using PLS-DA.

As shown in Table 8-1, both the SVM and ANN FE models from Chapter 3, which displayed overall prediction accuracies of 82% and 70% respectively, correctly predicted fenofibrate to display a positive FE. Importance of stringent model validation using numerous performance indicators is also highlighted in Chapter 3 as the sensitivity results for both positive (87%) and negative (69%) FE classes using the BCS tool appeared acceptable, seemingly outperforming the SVM model. However, upon a wider look at the precision (41%, 35%) and MCC values obtained (0.3, 0.2) the poor performance of the BCS tool was revealed.

In Chapter 4, fenofibrate solubility values upon SEDDS dispersion for both the PLS and MLR-based models, which displayed r^2 values ranging from 0.69-0.81, were found to differ less than 10% different from the experimentally determined values (i.e., $SEDDS_{Migylol812}$ (PLS 8.3% and MLR 8.9% difference) and $SEDDS_{OliveOil}$ (PLS 2.3% and MLR 2.5% difference). Here, the proximity of the PLS and MLR model results verifies the merit of a simplified MLR approach using three drug properties where no pre-processing or specialised software is required. The suitability of Chapter 4's $Do_{(Predicted)}$ approach to forecast developability was furthermore validated by comparing $Do_{(Predicted)}$ to $Do_{(SEDDS)}$ values (experimentally determined values). There 80% of drugs (8 out of 10 drugs assessed) used in model development were correctly predicted to transition to a "good solubility" DCS class (I/III). Further validation of the $Do_{(Predicted)}$ approach was demonstrated when, using the database of commercial

LBF drugs from Chapter 1, it was predicted 65.4% of these drugs would offer benefits as a LBF, owing to a projected shift in DCS classification to either Class I, III or IIa. Within this 16/17 (94%) of the DCS class IIa and IIb drugs did demonstrate a class shift.

Lastly, in terms of prediction accuracy, the PLS and ANN models developed to forecast aDS in sLBFs from Chapter 5, demonstrated correlation coefficients (r^2) ranging from 0.56-0.90, while the PLS models produced comparatively poor accuracy. For all four models, modest percentage differences between the predictions and the experimentally determined values for fenofibrate were seen (Table 8-1). Overall, these results highlight how the accuracy or greater applicability of a particular ML algorithm for a dataset is dependent on many factors and no one size fits all approach is currently known or available.

Table 8-1: Collation of the computational models produced in this thesis detailing the training and testing accuracy, and the % difference between the predicted versus reference results for fenofibrate.

Model Description	**Model Type**	**Training Accuracy**	**Testing Accuracy**	**Fenofibrate % Difference**
SVM Model Predicting FE (Chapter 3)	Classification	Overall Accuracy 72%	Overall Accuracy 69%	Correct Classification
ANN Model Predicting FE (Chapter 3)	Classification	Overall Accuracy 82%	Overall Accuracy 72%	Correct Classification
PLS Model Predicting $logSR_{MC}$ (Chapter 4)	Regression	r^2 0.81	RMSE 0.36	8.3%
PLS Model Predicting $logSR_{LC}$ (Chapter 4)	Regression	r^2 0.77	RMSE 0.37	2.3%
MLR Equation Predicting $logSR_{MC}$ (Chapter 4)	Regression	r^2 0.74	RMSE 0.39	8.9%
MLR Equation Predicting $logSR_{LC}$ (Chapter 4)	Regression	r^2 0.69	RMSE 0.37	2.5%
PLS Model Predicting aDS $sLBF_{Capmul}^{MC}$ (Chapter 5)	Regression	r^2 0.56	RMSE 0.79	4.9%
PLS Model Predicting aDS $sLBF_{Maisine}^{LC}$ (Chapter 5)	Regression	r^2 0.62	RMSE 0.79	23.4%
ANN Model Predicting aDS $sLBF_{Capmul}^{MC}$ (Chapter 5)	Regression	r^2 0.90	RMSE 0.21	13.4%
ANN Model Predicting aDS $sLBF_{Maisine}^{LC}$ (Chapter 5)	Regression	r^2 0.83	RMSE 0.25	9.1%

Interpretability of the Modelling Results

While predictive values and classifications form major outputs, the modelling and statistical analysis conducted in this thesis also facilitated improved molecular basis for understanding the various biopharmaceutical processes investigated in this current thesis. Various previously held assumptions, for example regarding the importance of solubility and permeability parameters to FE on bioavailability (Chapter 3) (191), or drugs commercialised with bio-enabling strategies being more likely to be found outside Lipinski's Ro5 (3) (Chapter 1) were verified. On the other hand, modelling did uncover some unexpected insights, including that no direct link between GFA class and aDS in sLBF could be established (Chapter 5 and Appendix 4-6) or that over 55% of LBF commercial drugs exceeded the reported T_m guide of <150°C for LBF suitability (86, 115) (Chapter 1). Moreover, new insights regarding SR trends achieved upon dispersion involving the cationic, anionic or neutral state of the drug at pH 6.5 were uncovered. DLS analysis suggested surface association of charged bile salts to the oil droplets formed upon SEDDS dispersion, which may facilitate favourable interactions between cationic drugs and these charged bile salts found in FaSSIF (Chapter 4). Additionally, the significance of drug reactivity and electron density to aDS (Chapter 5) was suggested. Both of which may warrant further in-depth investigations in the future to improve understanding of these hypotheses.

An important aspect of data-driven modelling is to ensure that any lack of bias towards inclusion or omission of any drug property is avoided in model development. For example logP, has long been highlighted as a significant drug property upon which suitable drug candidates are identified, particularly considering formulation as LBFs (95). More generally, logP is an important absorption property, related to both drug solubility and permeability. Well documented relationships have been established with

permeability, pharmacological potency, and toxicity, as it reflects the key event of molecular desolvation in transfer from aqueous phases to cell membranes and protein binding sites (54, 74, 79). Furthermore, the fact that logP is said to be changing less over-time than other physical properties appears to cement its standing as an essential "drug-like" property signally successful drug development (70, 79). However, in terms of data-driven ML, depending on the dataset in question, no statistically significant correlations may be attainable with a particular parameter, even if traditional hypothesis' regarding its importance exist. Therefore, it is important not to place extra significance on this during model development, so that input parameters are chosen blindly. In these cases, any familiarity bias with certain properties may decrease the predictive accuracy of the models. For example, in Chapter 1 as part of the retrospective analysis, logP was shown to significantly differentiate both SD and LBF drugs from others, however, its inability to separate commercial LBF and SD drugs does appear to challenge the commonly held view that drugs with higher logP values are more suited to LBFs. On the other hand, when unbiased selection of input parameters forms part of ML model development, any presence of logP or any of its closely related properties is statistically warranted. Accordingly, in this current thesis logP and closely related properties were final descriptors in multiple models. For example, logP was a significant parameter for FE classification prediction in Chapter 3, representing the sole property to be significantly different between the 3 classification groups upon initial statistical analysis and providing the most important property in the ANN. While $logD_{6.5}$ was significant in both the PLS and MLR models to predict the SR upon dispersion for $SEDDS_{Migylol812}$ and $SEDDS_{OliveOil}$ and MlogP (Moriguchi's method of calculating logP) (305) was a significant property to predict aDS for $sLBF_{Capmul}^{MC}$. Overall, these observations demonstrate how data-driven

modelling can only be achieved if pre-conceived theory's regarding important properties are disregarded, while the complexity of the phenomena in question was also demonstrated as in all cases numerous drug properties were required to obtain acceptable accuracy.

Importance of Biorelevant Experimental Inputs for Informing and Confirming Computational and *In Vitro* Tools – Lessons Learned from the Pig Model.

High model predictivity, as described previously in this discussion, or indeed high analytical capacity of any tool, can only be achieved when appropriate bio-relevant input parameters are used. Over the last 30 years significant strides have been made in using *in vitro* and *in silico* tools to improve and/or confirm mechanistic understanding of what is happening in the GIT upon drug administration (53, 55). While the emergence of computational tools is described in this thesis, the continued necessity for experimental validation testing of the formulations highlighted for potential success by these computational predictions remains a critical part of pre-formulation. Even though the use of computationally models has undoubtedly streamlined the number of candidates which reach the experimental testing stage, it is vital that the experimental confirmatory analysis of any *in silico* prediction is highly reflective of the *in vivo* environment. It is undoubtedly the major aim of such screening tools to provide value in early formulation design by ensuring validity and frequent refinement of their input parameters (49, 50) thereby intensifying their ability to accurately simulate *in vivo* scenarios, providing the most "bio-predictive" estimates. Accordingly, the proceeding sections highlight the potential insights to be gained from

microscopic and quantitative analysis of the landrace pig animal model to both inform and confirm that *in silico* and *in vitro* screening methods produced as part of this thesis are bio-predictive in their predictions.

Firstly, in conjunction with model validation requirements already discussed, the *ex vivo* fenofibrate solubility analysis of pig luminal fluids detailed in Chapter 6 further sought to confirm the close proximity of our modelling predictions to the likely *in vivo* results. Results revealed the *in-silico* solubility gain upon dispersion predictions for fenofibrate in $SEDDS_{OliveOil,}$ using the PLS (283.38 µg/mL) and MLR (296.95 µg/mL) models from Chapter 4 closely approximated *ex vivo* apparent fenofibrate solubility in pig intestinal fluids collected after 1 hour of 1g of administration of the same placebo SEDDS (271 ± 36 µg/mL) (Figure 8-1). The further observation that this 1-hour porcine estimate produced the highest solubility value among all samples taken at 0.5, 1, 2, 3 and 4 hours after oral SEDDS ingestion further justifies the capacity of the computational model from Chapter 4 to reflect the likely maximum *in vivo* solubility value achieved with SEDDS. Suggesting accurate Do calculations were produced using bio-predictive *in silico* estimates of drug solubility upon SEDDS dispersion. Thereby, facilitating increased confidence in these previously discussed *in silico* results.

In addition to *in silico* tools, a major aspect to improve early drug development is the link between *in vitro* testing and the pre-clinical evaluation *in vivo*. There has been significant focus in designing and validating improved biorelevant, biopharmaceutical *in vitro* tools simulating *in vivo* formulation performance (51, 55, 302, 306). However, it does appear that often anecdotal or traditionally applied rules rather than systematic validation of current experimental methods is the status quo. Often input parameters

for *in vitro* tools such as dilution conditions to simulate *in vivo* fluid volumes, appear to be taken from previous research and repeated. As a result, this sometimes doesn't represent the most accurate bio-simulation of the conditions. In line with this thought, Chapter 6 focused on the hypothesis that comparisons to porcine luminal fluids estimates could improve the bio-predictive nature of *in vitro* screening tools for LBFs. Subsequently, upon testing of 5 different dilution conditions it was observed that a 1:200 dispersion of $SEDDS_{OliveOil,}$ in biorelevant media (FaSSIF, FaSSGF and FaSSIFp) provided the closest simulation of maximum *ex vivo* solubility after 1 h upon administration of 1 g of SEDDS, as no significant differences between the obtained solubility values were obtained. Therefore, it is suggested that this dilution condition should be used going forward. This 1:200 dilution condition was used for the experimental inputs for the dispersion modelling data in Chapter 4. As illustrated in Figure 8-1, in terms of the intestinal media solubility results, no significant difference ($p = 0.163$) was found between the *ex vivo* porcine, *in vitro* results using biorelevant media (FaSSIFp, FaSSIF) or the predicted results from Chapter 4, once again helping to strengthen confidence in the validity of the resulting *in silico* predictions.

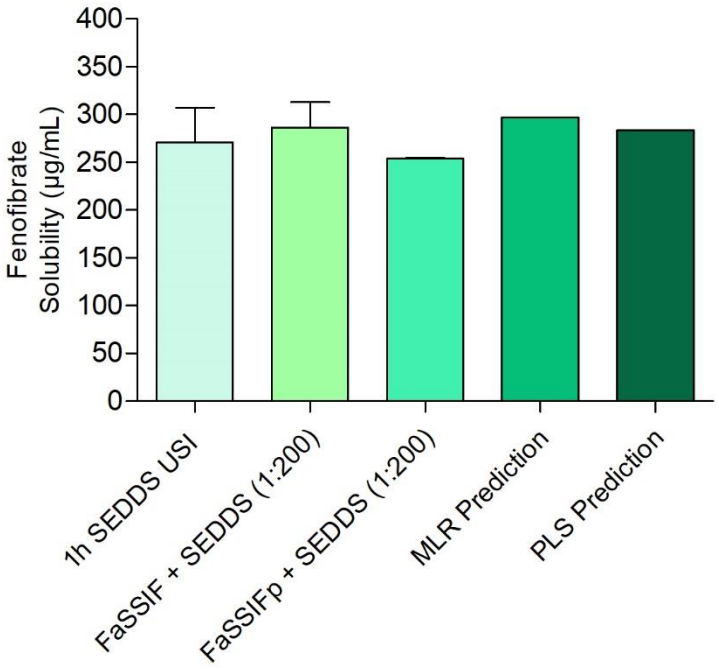


Figure 8-1: Visual illustration of the fenofibrate intestinal solubility results predicted and experimentally determined in Chapter 4 and 6. 1h SEDDS USI refers to the *ex vivo* solubility of fenofibrate in porcine upper small intestinal fluids collected 1h after placebo $SEDDS_{OliveOil}$ administration to pigs. FaSSIF + SEDDS (1:200) refers to solubility in FaSSIF plus a 1:200 dilution of dispersed SEDDS (Chapter 4). FaSSIFp + SEDDS (1:200) refers to solubility in FaSSIFp plus a 1:200 dilution of dispersed SEDDS (Chapter 6). MLR and PLS prediction refer to the predicted fenofibrate solubility values upon $SEDDS_{OliveOil}$ dispersion in FaSSIF using Chapter 4's PLS and MLR models.

In addition to providing validation to both these *in vitro* and *in silico* estimates, Chapter 6 provided increased understanding of the fate of a SEDDS formulation upon oral administration in an *in vivo* model. This work aimed to advance the utility of the landrace pig model to provide invaluable early indicators for human oral bioavailability and formulation performance. For example, to the best of our knowledge the microscopic images of the luminal fluid ultrastructure obtained in Chapter 6 provide the first visual characterisation of porcine luminal fluids, facilitating comparisons to similar images previously published for both biorelevant media and human GI fluids (257, 295-297, 300). The combination of these microscopic images and *ex vivo* solubility studies of the fluids facilitated an in-depth comparison of the

colloidal structures formed in the fasted and fed states. Furthering the previously published acknowledgment regarding ability of the pig model to simulate human FE (207). This work also uncovered substantial differences in both the complexity and concentrations of colloidal structures seen in the fasted state and upon SEDDS administration. Combination of these microscopic differences and the fact that each biorelevant media (FaSSIF, FaSSGF and FaSSIFp) displayed similar fenofibrate solubility upon SEDDS dispersion also pointed to the importance of the SEDDS excipients as drug reservoirs and as the primary drivers of fenofibrate solubility in contrast to any altered concentrations of bile salts and phospholipids or products of digestion in the respective medias.

Moreover, this work has provided insights to strengthen knowledge regarding previously acknowledged physiological differences between the pig and humans (62). For example, the images presented in Chapter 6 showing presence of small micelle resembling structures in both the fasted and 4 h SEDDS gastric samples also lend further support to the previously held hypothesis by Henze *et al.* regarding reflux of bile from the pig duodenum to the stomach resulting in high bile salt concentrations in the landrace pig stomach compared to humans (294). In summary, the analysis conducted in Chapter 6 provided useful novel insights into the pig as a predictive model of human bioavailability, while succeeding in both producing an increasingly bio-predictive *in vitro* screening tools for LBFs and verifying the accuracy of the *in silico* predictions for fenofibrate from Chapter 4.

Interplay Between Computational Modelling and Conventional Biopharmaceutics Related Classifications Systems

In recognition of the previously acknowledged significance of drug classification systems to inform drug developability and the likely rate limiting steps for oral bioavailability, this thesis has identified three ways in which commonly used classifications systems (BCS, DCS, rDCS and BDDCS) and computational pharmaceutics tools are intertwined.

Firstly, predictive estimates from conventional classification systems can be utilised to establish the predictive accuracy of ML models by providing comparative estimates. Accordingly, a drug's BCS class has previously been suggested as a useful predictive tool to anticipate FE (191, 214). However, using the database of newly licensed drugs in Chapter 3, only a 46% overall accuracy was achieved, with poor MCC values of 0.2, 0.2 and 0.3 calculated for the FE categories. This result was substantially lower than results achieved for both ML ANN and SVM classification models, therefore, providing a good baseline comparator of their predictive ability. Additionally, the close reflection of the significant properties for the ANN model, in particular logP, T_PSA and HBD, to the stated importance of solubility and permeability parameters to BCS drug classifications also aided in verification of the significance of these parameters to early pre-clinical decision-making.

Secondly, in Chapter 1, by classifying the dataset according to BDDCS class, previously uncovered applications for bio-enabling formulations can be uncovered. In Chapter 1, using the BDDCS, it was revealed that, as expected, the highest proportion of commercial LBF (76%) and SD (60%) drugs belonged to BDDCS Class 2. However, contrary to the generally taken view that LBFs are suited for lipophilic and

PWSD (14), results revealed that the second highest proportion of LBF drugs were found to be in the highly soluble BDDCS Class 1 category, as a lower dosage strength requirement or the ease of large scale LBF manufacture was suggested as a hypothesis to explain this unexpected result. Therefore, the use of the BDDCS tool as part of the retrospective analysis in Chapter 1 aided in highlighting potential uncovered or somewhat hidden further applications of LBFs.

Thirdly, Chapter 4 demonstrated how predictions of SR upon SEDDS dispersion can be combined with established classification systems to advance the broader applicability of computational pharmaceutics tools. Chapter 4 explored suitability of linking the biopharmaceutical $Do_{(Predicted)}$ approach to the framework provided by the DCS and rDCS and hence, providing a tool for guiding developability of a SEDDS strategy. Incorporation of DCS class transition predictions with SEDDS into the rDCS, as part of the initial "standardised investigations" was also suggested (67). As part of the rDCS guide, various other properties of the drug including ionisation potential, would next trigger customised *in vitro* investigations. Embedding these computational models within a broader developability framework such as the rDCS demonstrates effectively how the wider applicability of *in silico* predictions can be strengthened for an external formulation scientist. Undoubtedly, a clear visualisation of potential for formulation success within an established biopharmaceutics framework, such as the rDCS, has a wider significance in a development setting, compared to a purely numerical predicted output from a model.

Model Integration and Computational Accessibility

This potential to embed ML models within widely known drug classification systems to support developability provides only one example of methods by which computational pharmaceutics tools can be integrated into the current drug development landscape. The various tools detailed in this thesis demonstrate importance of a collaborative, accessible and overall user-friendly approach to reinforce the appeal of computational algorithms. The quote "essentially all models are wrong but some are useful" often attributed to statistician George E. P. Box, rings true in this case. Since modelling will always be associated with a certain degree of uncertainty, it should be recognised that any type of model-based drug development approach, in particular the comparatively youthful field of computational pharmaceutics, should not be naively expected to predominantly replace *in vitro* and *in vivo* assessments. Therefore, at present the continued need to merge computational tools with appropriate levels of reliable *in vitro* confirmatory testing is acknowledged. However, while these computational tools may not completely revolutionise how *in vivo* performance is predicted, their adoption can significantly optimise the costly pre-formulation process and mitigate against declining R&D productivity.

There are multiple ways in which the growing data science and research communities can unite and promote widespread application of these tools. Examples of such means are demonstrated across this current thesis. Firstly, as seen in Chapter 1 significant insights can be gained from retrospective analysis of drug product datasets. While much of the data needed for such analysis is already available, it is often currently not collected in a way that allows helpful ML projects to be performed. It is now the job of interested parties to collate, harness and structure such valuable data assets.

Retrospective analyses like that detailed in Chapter 1 could be repeated and updated periodically to ensure their enduring impact. Moreover, the potential for such commercial datasets is not limited to a singular analysis, as demonstrated in Chapter 4 where the previously collated LBF commercial product database from Chapter 1 was used to externally validate the SR computational modelling approach accuracy to predict drug DCS classifications. In this instance, it was shown how the same dataset can be used as part of both a "top down" and "bottom up" computational analysis.

While additional examples of the wider applicability of the computational models developed in this thesis have previously been discussed, including connections to well established classification systems and examples of linear and non-linear ML approaches, the importance of not producing "stand-alone" predictions is paramount. Combined and sequential use of these computational tools including predictions of both QTPP characteristics of interest (Chapter 3), followed by predictions to determine the most appropriate bio-enabling formulation application (Chapters 1, 4, 5 and 6) provides a synergistic value to these *in silico* models. Both this and the place of these computational assessments within the wider drug developability framework will be discussed in greater detail later in this discussion.

Going one step further, it could also be suggested that improved mechanistic understanding of some of the physiological phenomena underpinning the various *in silico* predictions could be achieved by combining predictions, simulations, and imaging from various sources. For example, a formulation scientist trialling use of a SEDDS approach could combine computational predictions of drug loading in LBF excipients (114, 115, 160, 163), predicted SR upon SEDDS dispersion in intestinal

fluids (Chapter 4), SR trends in ionisation of compounds observed (Chapter 4), MD simulation of SEDDS behaviour in the GIT (164-166) and microscopic images post SEDDS ingestion (Chapter 6). This would facilitate an increased understanding of the *in vivo* processing of the prospective SEDDS and the relationship to formulation performance. Furthermore, while the models obtained throughout this thesis focus on ML and data-driven predictions, it is likely that while data-driven and theory-based approaches occupy either end of the modelling spectrum, realistically the most accurate models will contain both theory and data-based aspects hand-in-hand. Therefore, combinations of different approaches, as described above, will aid to achieve the best performance.

In addition, the value of any model is determined by its practicalities for its intended purpose. Therefore, the importance of accessible and easily disseminated modelling approaches, often suggested to be important drivers for increased adoption of ML (68), is also addressed throughout this thesis. This is exemplified via 1) development of drug property-related MLR equations in Chapter 4 requiring no data-processing or specialised software for predictions, 2) the easily interpretable and accessible nature of the four final modelling parameters in the FE ANN model (Chapter 3) and 3) the ability for the produced ANNs to be exported as predictive model markup language (PMML) files. PMML is a gold standard approach to share models allowing the models to be compatible with a wide range of ML software's, not only the software on which the networks were produced. This facilitates external predictions to be made by other researchers without any manual manipulation or prior knowledge of the networks being necessary. Moreover, the preference to develop models which contain easily interpretable drug properties, such as the MLR equations produced in Chapter

4, highlights the issue regarding a necessity to strike a balance between model interpretability versus statistical accuracy, which can sometimes be at odds. In that particular case, while the overall accuracy of the comparative PLS-based models using 5-6 drug properties appear more robust versus the MLR equations predicting SR, it must be decided in the individual case if an increased training/testing accuracy or the omission of any data scaling or software requirements is more favoured. Without question, the most appropriate model depends on numerous factors including the stage of development, costs, accessibility to software or indeed availability of computational expertise or molecular descriptors.

Finally, in response to the growth of big data analytics across the pharmaceutical industry, ML models built and validated using hundreds, if not thousands of input samples remain the ultimate goal. However, it must not be forgotten that in certain circumstances, whether limited by resources or availability of high-throughput analytical methods, particularly in an academic setting, such large datasets are currently unrealistic in certain aspects of pharmaceutics. Indeed, if the science of computational pharmaceutics is to move to the mainstream studies such as some of those presented in this thesis, involving relatively smaller datasets, are required to disseminate the necessary computational expertise and identify potential modelling opportunities. These studies built on smaller datasets can provide a foundation upon which larger subsequent datasets can be employed and again highlight the necessity for continued experimental confirmatory studies for identified candidates. Overall, in summary, as exemplified by these examples, it is hoped that by extending the accessibility of these models, synergistic adoption of numerous tools will help to guide drug developability and support computational pharmaceutics in achieving its full potential.

Computationally Informed and Experimentally Confirmed Developability Testing to Reconsider the Drug Substance to Drug Product Paradigm

Stemming from the significant need for improved computational tools, a primary aim of this thesis was to investigate if data-driven ML models, predicting numerous drug developability indicators, could provide an increasingly structured approach to formulation selection. In line with the need to integrate these computational models into the wider drug developablity setting, sequential application of these tools within a refined drug substance to drug product development framework will be discussed (Figure 8-3). In that case use of these tools can provide guidance on how to proceed strategically when presented with a seemingly difficult to deliver drug, by both highlighting potential barriers to development and subsequently seeks to identify delivery solutions to overcome said barriers. Before discussing the wider applications for these tools within the drug developability space, an exemplary case study detailing the model predictions using fenofibrate is shown in Figure 8-2, illustrating the preliminary information obtainable from the computational models. Fenofibrate, as used previously throughout this discussion, a typical BCS class II compound, was chosen here as conventional formulations display low and variable bioavailability in the fasted state relative to the fed state, due to low solubility and resultant slow dissolution (307). Thus, it is seen as a reliable model for assessment of FE and formulation decision-making.

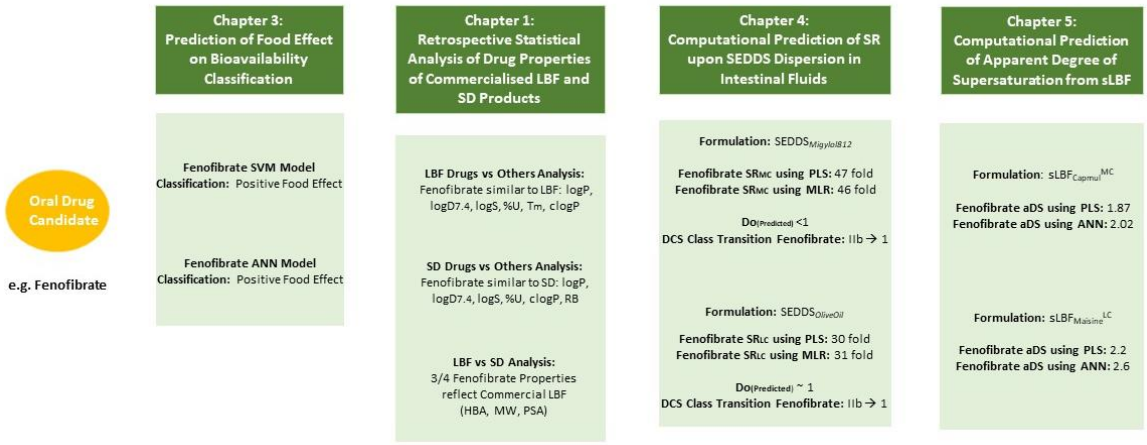


Figure 8-2: Visual representation of the modelling results for each computational model produced in this thesis for the model compound fenofibrate.

As seen in Figure 8-2, after drug candidate identification, a first prediction strategy within this computational suite would be to allocate either a predicted positive, negative or no FE on bioavailability classification. Early prediction of FE would provide pertinent information to formulation scientists (92). This could facilitate identification of potential delivery problems, as possibility of either a significant positive or negative FE at this stage would identify either a need to formulate the drug with a more sophisticated technology to overcome the FE or to redesign the drug itself. Accordingly, Chapter 3 of this thesis identified two ML methods which accurately classified drugs according to FE. Figure 8-2 demonstrates that both SVM and ANN models correctly predicted fenofibrate as belonging to the "positive" FE category, as it is well known that fenofibrate bioavailability increases significantly with increased solubilisation and/or dissolution in the fed state (207).

Next as both ML FE models predicted a positive FE for fenofibrate, subsequent application of other computational tools can indicate whether certain bio-enabling formulation strategies, which are reported to overcome FE (92), would offer a high

likelihood of success for this drug. This anticipation of "successful formulation potential" would generally consist of several successive analyses. In this case, the first of which being the top-down retrospective statistical analysis of drug property trends of commercial LBF and SD products as described in Chapter 1. The limitations and somewhat crude nature of this sort of approach, including a "successful formulation bias" have been previously identified in this thesis. However, it is believed its adoption would provide an early indicator, or a "formulation likeness filter" with little cost, of which bio-enabling strategy would be more likely to be successful with this compound prior to further in-depth investigations of applicable excipient types.

Application of the analysis in Chapter 1 can be two-fold, as preliminary comparisons of statistically significant differences between both LBF and SD approaches versus Others suggest if the properties of the drug more closely reflect those which have used bio-enabling formulations versus a conventional strategy. Next, subsequent statistical comparisons between trends in LBF versus SD commercial drugs properties identifies which of these bio-enabling strategies is more suited, using these previous property trends as a guide. In addition, further analysis of trends in dosage forms and commercialisation rates per decade, also identified in Chapter 1, can likewise provide useful information to formulators. As part of this retrospective statistical analysis of commercial LBF and SD drugs, compared to drugs not licensed using either approach, 8 and 11 of the well-established drug properties analysed were found to be significantly different between both LBF and SD approaches versus Others. As illustrated in Figure 8-2, it is suggested that the properties of fenofibrate resembled trends in drugs commercialised using both bio-enabling approaches in terms of its values for 7 drug properties including logP, %U, logS, logD, T_m, clogP and RB count. Subsequently, when focus shifted to drug properties which were significantly different

between LBF versus SDs, fenofibrate appeared to more closely reflect trends in commercial LBF drugs in ¾ cases i.e., lower number of HBA (three), lower MW (360.83 g/mol) and lower PSA (52.6 $Å^2$). Therefore, using Chapter 1's retrospective analysis as a guide to inform future success, it is anticipated that even though commercial examples exist of both LBF and SD fenofibrate products, the drug properties of fenofibrate more closely conform to previous commercial LBF properties.

Next, as this retrospective analysis from Chapter 1 did signal likely success with a LBF strategy for fenofibrate, formulators may be prompted to try to anticipate which type of LBF application, including those classified by the LFCS (21) or beyond. A common strategy may be to assess the potential for SEDDS formulation to facilitate adequate drug solubility, through their self-emulsifying properties (14). Prediction of drug solubility in LBF excipients and dose loading capacity has provided a significant research focus for many with numerous computational approaches including predictions from drug structure (114, 115, 160-163, 264), as previously described in this thesis. Therefore, it is also acknowledged that predictions from these tools could also be employed at this stage in the wider drug developability setting. However, it has long been recognized that successful LBF capability assessment needs to include aspects of *in vivo* formulation performance (88, 308). This acknowledges the interplay between any formulation excipients and the GIT, which is of high complexity, as modifications in the GIT upon SEDDS ingestion are crucial in determining formulation performance. While significant strides have been made in understanding this complex formulation-drug-intestinal fluid interplay through the use of, for example, MD simulations (164-166), to the extent of our knowledge the ML models produced in Chapter 4 of this thesis are the first examples of ML LBF models to

incorporate formulation dispersion. These provide the first quantitative models to estimate solubility increases (i.e., solubility ratios) achieved by a drug upon dispersion of a SEDDS in intestinal fluids, to determine SEDDS suitability.

PLS models of 5-6 drug properties and subsequent MLR equations of only 3 easily interpretable drug properties were described in Chapter 4. It can be seen in Figure 8-2 that both methods predicted high SR of >30 fold for fenofibrate using both a MCT and LCT-based SEDDS. Therefore, Chapter 4's work also demonstrated a two-fold application of modelling as development of multiple models can provide a useful comparison of the most appropriate formulation excipient blend (MCT versus LCT-based). Such a significant solubility increase is to be relatively expected as fenofibrate is often used as a model compound for the assessment of LBFs, with numerous preclinical studies demonstrating improvement in fenofibrate bioavailability from these systems (51, 309). After predicting SR this approach can next be employed to calculate the biopharmaceutical dose number (Do) produced in intestinal fluids, guiding the dose that is effectively solubilised and if a DCS Class transition to "good solubility" is likely**.** Importantly, from a practical point of view, only prior knowledge regarding solubility of the candidate in FaSSIF and the dose would be required to obtain such an important indicator of success using this tool. In a particular development setting, depending on the individual dosage strength requirements, the suggested dose obtained from such a $Do_{(Predicted)}$ approach i.e., if < 1, could prompt manual experimental screening of this SEDDS ability to deliver the selected dose.

Conversely, if $Do_{(Predicted)}$ results (i.e. $Do_{(Predicted)}$ >1) from Chapter 4 or indeed other previously published computational tools which predict initial drug loading capacity in LBFs (115), suggest that a traditional LBF approach presents a high level of risk, a next logical step is to analyse the potential for other LBF strategies to facilitate

delivery. Dose loading limitations are a critical restraint to the more widespread use of traditional LBF systems (6, 95). As the dosage levels achievable are restricted by the inherent drug solubility in the lipid vehicle or in the lipid reservoirs produced upon formulation processing. This solubility is often limited for "brick dust" molecules which exhibit high crystal lattice energies. Numerous advanced initiatives are emerging to combat these limitations including ionic liquids/lipophilic salts (310, 311), hybrid systems (312) or lipophilic prodrugs (313). Chapter 5 of this thesis focused on the development of computational models predicting the solubility increases achieved through thermal induced supersaturation i.e., aDS of a drug, in a sLBF system. sLBFs involve heating of a drug–lipid mixture to overcome the drug's crystal lattice energy, while during heating and upon cooling the drug is maintained in a supersaturated state (6, 34). Supersaturated systems have previously been described according to the aDS ratio (275, 276), which has also been considered a major factor impacting sLBF ability to maintain drug supersaturation upon storage (32). Therefore, we believed aDS prediction was a useful indicator of achievable dose loading levels in early development. This computational tool can compare any gain in dose loading levels achieved to the non-supersaturated LBF. As demonstrated in Chapter 5, an ANN modelling approach was shown to outperform a PLS approach using 15 and 11 properties to predict aDS for both a MCM and LCM sLBF ($sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$). Once again demonstrating the potential for modelling to compare different fatty acid chain lengths.

For fenofibrate, as outlined in Figure 8-2, aDS ratios of 2.02 and 2.6 were predicted for $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$ respectively. This ANN modelling results suggest use of the MCM formulation may achieve higher dose loading for this drug (approximately 150 mg/mL versus 130 mg/mL), where a similar initial dosage level

to the $SEDDS_{Migyolol812}$ formulation from Chapter 4, previously published, was observed (143.8 ± 10.9 mg/mL) (51). However, formulation long-term stability and precipitation potential, or applicability of precipitation inhibitors for these sLBF (314) should be next considered by the formulation scientist, in addition to likely *in vivo* solubility. While limitations of these models as stand-alone assessments, including omission of SEDDS digestion and long-term stability of sLBFs are acknowledged, as demonstrated via this case study, valuable information to optimise early decisions in LBF development can be obtained using these models.

These preceding paragraphs have illustrated how the computational tools developed in this thesis can collectively provide vital drug developability information. However, in a much broader sense the work in this thesis raises the question of where this computational suite of tools fits into the modern drug substance to drug product development framework. With a view to the future, it is suggested that the knowledge accrued in this thesis could be used to redefine the framework for transitioning from drug substance to drug product. With this in mind, a refining of the current drug developability and formulation decision pathways, incorporating both computationally informed and experimentally confirmed aspects of developability testing could initiate a paradigm shift (Figure 8-3). Previously the biopharmaceutics risk assessment roadmap (BioRAM) was highlighted as a critical mechanism to identify the necessary information, patient factors and decision steps for optimum drug delivery.

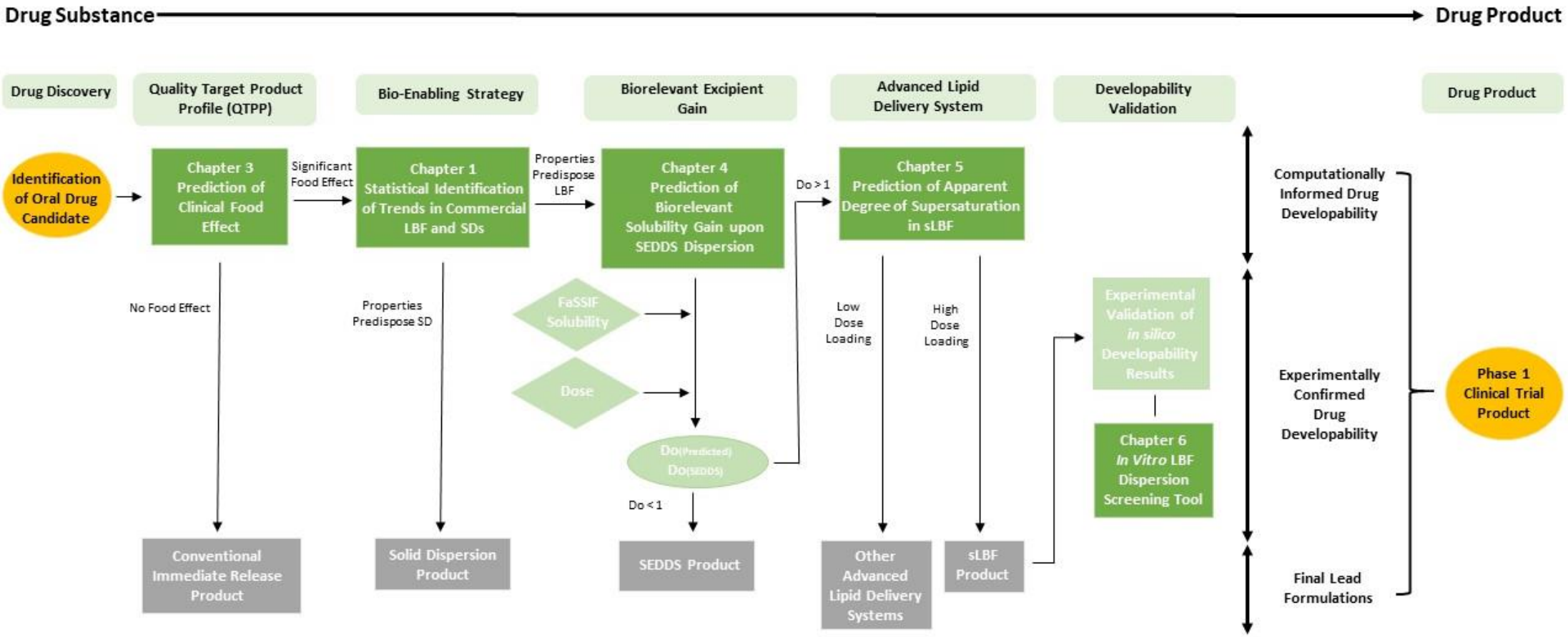


Figure 8-3: Visual incorporation of the computational and *in vitro* tools developed in this thesis as part of a refined drug substance to drug product development paradigm. Here both computationally informed and experimentally confirmed aspects of drug developability are incorporated.

In the BioRAM use of modelling and simulation techniques, in line with those produced in this thesis, were highlighted as critical sources of information for optimum drug delivery (315). It can clearly be identified that the computational tools produced in this thesis reflect key decision steps identified in the BioRAM. A visualisation of the stages in developability testing where the computational and *in vitro* tools produced in this thesis may be applied within the wider testing framework can be seen in Figure 8-3. Here, in line with the "learn and confirm" approach followed by the BioRAM; computational "learning" tools can be integrated with appropriate experimental "confirmatory" studies to identify the optimal delivery approach.

Firstly, at the drug discovery phase after identification of an oral drug candidate, the fundamental considerations of the BioRAM are the quality target product profile (QTPP) characteristics which encompass key product attributes deemed clinically relevant for achieving product success (315). One such characteristic of clinical interest is the food effect on bioavailability. Depending on the therapeutic class or target population early knowledge of food effects may lead to a "go/no-go" decision or initiation of investigations into alternate dosage forms to mitigate this effect. Therefore, even at the earliest stages of drug discovery computational tools such as the ANN and SVM FE classification models produced in Chapter 3 can be used to provide vital early information regarding the feasibility of a conventional formulation versus need for an advanced bio-enabling approach.

Next a central tenet of the BioRAM involves understanding the optimal delivery of a drug substance through developing a suitable formulation approach (315). As part of the BioRAM the significance of leveraging prior information in drug development decisions to achieve effective delivery is highlighted. In line with this, in this thesis the retrospective statistical analysis conducted in Chapter 1 employed prior data to

identify drug property trends in commercial LBF and SD drugs. Such an approach can be successfully employed at this stage in drug development to highlight if a prospective candidate reflects the physicochemical properties of drugs previously successfully developed using these bio-enabling approaches (Figure 8-3). Next, sequentially along this revised framework, the computational models developed in Chapters 4 and 5 can also be applied for bio-enabling formulation selection to provide early indicators of the likelihood of success with SEDDS or sLBF systems, depending on the SR and dose loading levels achieved for the candidates. At this point experimentally determined inputs such as FaSSIF solubility and dose are required. However, as it is acknowledged that at this early point in development the exact dose is unlikely to be known, the dose range of 5mg – 50mg – 500mg as suggested by the rDCS could be used for predictions (67).

In line with the suggestions of the BioRAM, a refined drug developability paradigm will certainly include numerous tools but will also likely evolve from an initially computationally informed approach to an increasingly experimentally confirmed drug development approach once the initial computational outputs have been established. This change in significance from computational to experimental testing can be explained by the main intended purpose of both the BioRAM and the computational models applied in this thesis. Both primarily serve as early risk assessments which highlight the likelihood of success for a particular formulation approach. Therefore, as illustrated in Figure 8-3, based on the *in silico* results and knowledge generated, any emerging understanding should then be validated in experimental confirmatory studies to ascertain whether the formulation approaches highlighted could result in acceptable drug product performance. In doing this only a streamlined number of candidates with the highest chance of success will reach the experimental testing stage.

This highlights that while the utility and strength of computational tools in the development paradigm continues to grow, at present, there still remains a requirement for structured initial *in vitro* studies to confirm the *in silico* predictions before significant financial investment is initiated. Examples of such *in vitro* approaches at this point in development include the *in vitro* LBF screening tool detailed in Chapter 6 of this thesis (Figure 8-3). Subsequently, following the application of the various confirmatory assessments, as illustrated in Figure 8-3, application of the various computationally tools included in this thesis can result in an optimal phase 1 clinical trial product.

Overall, both the intuitive linking of the tools produced in this current thesis and their place in a refined drug developability framework has been demonstrated. By providing a structure to initial clinical and formulat-ability decision making with little cost or manual effort, these "team players" in the vast computational pharmaceutics toolbox can advance the field of computational pharmaceutics through their application at numerous points in drug development. This work has effectively demonstrated the need for early predictors of developability and how knowledge of drug properties and biopharmaceutical parameters, can aid the transition from drug substance to drug product. In doing so capturing both computationally informed and experimentally confirmed aspects of drug development.

Overall Conclusion and Future Perspectives

This thesis has, firstly, demonstrated the capability of ML to guide formulation strategy with relatively minimal manual or financial effort. While the overarching concept of computational pharmaceutics has been spoken about for the better part of a decade, over the last few years in particular, momentum is increasing towards an integrated model-informed approach to guide drug development strategies. The computational models produced in this thesis, including accurate predictions of FE on bioavailability, SR upon SEDDS dispersion in intestinal media and aDS upon heating, represent useful additions to the emerging computational pharmaceutics toolbox. These tools support the increasing adoption of bio-enabling formulations, in particular LBFs. The easily interpretable nature of these tools, including predictive models which do not require specialised software, highlights the importance of both useability and dissemination of computational expertise to facilitate further expansion of computational pharmaceutics-driven development. To further elucidate the wider applicability of these models, future studies incorporating more drug datasets of varying chemical and structural composition and the investigation of deep learning potential are required (158).

Secondly, while the modelling predictions form major outputs of this thesis, this work has also aided further recognition of the significance of drug properties to numerous phenomena relevant for formulation testing. It was demonstrated how analysis of drug properties can provide a mechanistic tool to elucidate formulation performance. Throughout this thesis increased molecular understanding of which properties form influential players for various aspects of formulat-ability were identified. In some cases, new or unexpected drug properties were highlighted for certain phenomena for which in-depth understanding has not been fully elucidated to this point. Accordingly,

future studies can gain a deeper understanding for the presence of these properties in our final models, extending past the preliminary hypotheses for their inclusion presented in this thesis.

Thirdly, this thesis has examined the utility of the landrace pig model to inform increasingly bio-relevant input parameters to improve the bio-predictive capacity of *in vitro* and *in silico* tools. Through comparisons to *ex vivo* solubility levels in porcine fluids, an *in vitro* screening tool which closely mimicked likely *in vivo* performance was developed. Furthermore, the ability of the pig to act as a model of human bioavailability was strengthened as the first microscopic analysis of porcine fluids was provided. This facilitated comparisons of the colloidal structures formed in the fed and fasted state. Even though *ex vivo* solubility results in porcine fluids aided in verification of the bio-predictive nature of the related computational modelling predictions, further studies are now needed using different formulation types and ionisable drugs, to further investigate appropriate *in vitro* parameters which closely mimic *in vivo* performance.

Finally, this thesis proposed a refined drug substance to drug product development paradigm, encompassing both computationally informed and experimentally confirmed decision making. It was shown that through sequential application of the computational and *in vitro* tools produced in this thesis, informed decisions regarding both the necessity for or suitability of a particular LBF approach can be anticipated. This thesis has also advanced computational pharmaceutics through incorporation of drug developability. Capacity to integrate these tools into standard developability testing demonstrates that while a solid foundation is present, substantial effort is required to shift computational pharmaceutics tools from “nice to have” to a “must have”.

Appendix 1: Chapter 1

Table 1-1 Results of Statistical Analysis comparing LBF, SD and Others using all BDDCS Classes (Total).

Drug Property	Descriptors	LBF	SD	Others	Statistical Tests	p-value LBF vs SD	p-value LBF vs Others	p-value SD vs Others
clogP	*n*	49	37	763	Levene's Test	0.04	0.03	0.298
	Median	4.94	4.49	2.49	Welch's/t-test	0.14^{W}	0.00^{W}	0.00^{t}
	Mean	5.30	4.49	2.29	Mean Difference	0.82	3.01	2.20
	SD	2.97	2.04	2.37	95% Confidence Interval	(L) -3.17	(L) 2.14	(L) 1.42
	Q1, Q3,	3.32, 7,32	3.24, 4.49	0.81, 3.86		(U) 1.95	(U) 3.88	(U) 2.98
	Min, Max	-0.73, 14.36	-1.63, 7.63	-6.66, 10.97				
	Variance	8.85	4.15	5.613				
Hydrogen Bond Acceptors	*n*	49	37	763	Levene's Test	0.37	0.03	0.45
	Median	4	6	4	Bootstrap	0.011	0.85	0.00
	Mean	4.76	6.87	4.64	Mean Difference	-2.11	0.12	2.26
	SD of Mean	4.01	2.72	3.02	95% Confidence Interval	(L) -3.43	(L) -0.89	(L) 1.33
	Q1, Q3	2, 6	5, 9.5	3, 6		(U) -0.70	(U) 1.36	(U) 3.12
	Min, Max	1, 23	2, 13	0, 4				
	Variance	16.11	7.398	9.092				
Hydrogen Bond Donors	*n*	49	37	763	Levene's Test	0.77	0.45	0.714
	Median	1	3	2	Bootstrap	0.32	0.74	0.07
	Mean	1.92	2.27	1.82	Mean Difference	-0.35	0.09	0.45
	SD of Mean	1.86	1.43	1.78	95% Confidence Interval	(L) -1.05	(L) -0.39	(L) -0.05
	Q1, Q3	1, 3	1, 3.5	1,2		(U) 0.34	(U) 0.63	(U) 0.96
	Min, Max	0, 10	0, 4	0, 23				
	Variance	3.45	2.04	3.18				
$logD_{7.4}$	*n*	49	37	488	Levene's Test	0.15	0.003	0.7
	Median	3.87	3.59	1.34	Bootstrap	0.52	0.00	0.00
	Mean	3.82	3.46	1.25	Mean Difference	0.36	2.57	2.21
	SD of Mean	2.89	2.40	2.15	95% Confidence Interval	(L) -0.72	(L) 1.71	(L) 1.38
	Q1, Q3	1.48, 5.65	2.15, 5.26	-0.11, 2.65		(U) 1.45	(U) 3.43	(U) 2.98
	Min, Max	-3.2, 11.35	-5.4, 7.05	-8.86, 10.40				
	Variance	8.36	5.76	4.63				
logP	*n*	49	37	454	Levene's Test	0.22	0.45	0.33
	Median	4.50	4.37	2.36	Bootstrap	0.25	0.00	0.00
	Mean	4.66	4.16	2.22	Mean Difference	0.49	2.44	1.94
	SD of Mean	2.16	1.78	1.99	95% Confidence Interval	(L) -0.29	(L) 1.79	(L) 1.31
	Q1, Q3	3.31, 6.15	3.16, 5.69	0.92, 3.66		(U) 1.31	(U) 3.07	(U) 2.55
	Min, Max	0.28, 10	-1.80, 6.92	-8.83, 7.80				
	Variance	4.66	3.18	3.96				
logS	*n*	46	35	587	Levene's Test	0.24	0.28	0.008

Drug Property	Descriptors	LBF	SD	Others	Statistical Tests	p-value LBF vs SD	p-value LBF vs Others	p-value SD vs Others
	Median	-4.8	-5.3	-2.92	Welch's Test/ t-test	0.11[t]	0.00[t]	0.00[w]
	Mean	-4.38	-4.95	-2.81	Mean Difference	0.58	-1.57	-2.15
	SD of Mean	1.64	1.49	1.73	95% Confidence Interval	(L) 0.58	(L) -2.09	(L) -2.63
	Q1, Q3	-5.54, -3.70	-5.7, -4.4	-4.14, -1.34		(U) 0.357	(U) -1.06	(U) -1.63
	Min, Max	-6.50, 0.25	-8.85, -0.57	-7.44, 1.70				
	Variance	2.70	2.22	2.968				
Maximum Dosage Strength (mg)	*n*	44	37	760	Levene's Test	0.79	0.39	0.46
	Median	62.5	100	75	Bootstrap	0.50	0.195	0.30
	Mean	118.59	144.33	195.79	Mean Difference	-25.75	-77.21	-51.46
	SD	141.02	181.56	761.68	95% Confidence Interval	(L) -106.33	(L) -160.1	-134.01
	Q1, Q3	1.94, 200	40, 200	10, 250		(U) 44.78	(U) -7.78	30.48
	Min, Max	0.0005, 500	1, 1000	0.04, 20000				
	Variance	19885.23	32964.28	580148.89				
Melting Point (^{O}C)	*n*	47	30	652	Levene's Test	0.01	0.20	0.05
	Median	151	170.5	180.5	Bootstrap	0.231	0.035	0.54
	Mean	160.81	175.97	181.18	Mean Difference	-15.18	-20.38	-5.21
	SD	64.14	44.83	58.8	95% Confidence Interval	(L) -39.26	(L) -39.91	(L) -22.22
	Q1, Q3	116.5, 211	141, 207.86	139, 222.5		(U) 8.98	(U) -1.14	(U) 11.81
	Min, Max	38, 284	80.5, 271	43, 374				
	Variance	4114.07	2009.88	3457.69				
Molecular Weight (g/mol)	*n*	49	37	763	Levene's Test	0.12	0.001	0.00
	Median	396.65	493.58	329.63	t-test/Bootstrap	0.009	0.011	0.00
	Mean	448.20	586.63	354.63	Mean Difference	-138.43	93.57	231.99
	SD of Mean	216.82	230.92	148.61	95% Confidence Interval	(L) -235.25	(L) 37.08	(L) 158.46
	Q1, Q3	314.61, 517.1	405.47, 785.47	263.79, 419.39		(U) -40.70	(U) 156.41	(U) 310.92
	Min, Max	144.21, 1202.61	129.17, 1113.2	46.07, 1681.91				
	Variance	47011.07	53322.02	22086.19				
pDose	*n*	44	37	760	Levene's Test	0.001	0.000	0.007
	Median	3.86	3.88	3.71	Bootstrap	0.115	0.063	0.737
	Mean	4.27	3.50	3.83	Mean Difference	0.407	0.45	0.04
	SD of Mean	1.49	0.66	0.90	95% Confidence Interval	(L) -0.07	(L) 0.36	(L) -0.18
	Q1, Q3	3.29, 4.94	3.37, 4,17	3.16, 4.45		(U) 0.91	(U) 0.90	(U) 0.26
	Min, Max	2.11, 8.60	2.11, 5,57	1.37, 7.00				
	Variance	2.21	0.44	0.82				
Percentage Excreted Unchanged in Urine (%)	*n*	40	33	667	Levene's Test	0.86	0.00	0.00
	Median	0.5	0.05	4.2	Bootstrap	0.70	0.01	0.03
	Mean	7.33	5.68	19.47	Mean Difference	1.65	-12.14	-13.79
	SD of Mean	16.5	18.24	27.78	95% Confidence Interval	(L) -7.21	(L) -17.07	(L) -18.77
	Q1, Q3	0.10, 5.75	0, 1.25	0.5, 30		(U) 9.23	(U) -6.41	(U) -7.18
	Min, Max	0, 69	0, 99	0, 100				
	Variance	272.16	332.84	771.91				
pKa (strongest acid)	*n*	46	29	624	Levene's Test	0.34	0.52	0.44
	Median	10.44	9.7	10.33	Bootstrap	0.38	0.73	0.41
	Mean	10.20	9.12	9.90	Mean Difference	1.08	0.30	-0.78
	SD of Mean	5.66	4.86	0.21	95% Confidence Interval	(L) -1.18	(L) -1.38	(L) -2.61

Drug Property	Descriptors	LBF	SD	Others	Statistical Tests	p-value LBF vs SD	p-value LBF vs Others	p-value SD vs Others
	Q1, Q3	0.27, 22	4.09, 12.63	4.77, 13.98		(U) 3.41	(U) 1.99	(U) 1.06
	Min, Max	4.75, 13.86	0, 19.90	-12.00, 19.96				
	Variance	32.07	23.57	26.99				
Polar	*n*	49	37	762	Levene's Test	0.82	0.12	0.26
Surface	*Median*	52.9	112.85	72.91	Bootstrap	0.003	0.85	0.00
Area	*Mean*	79.68	125.92	81.48	Mean Difference	-46.24	-1.91	44.33
(Å^2)	*SD of Mean*	67.94	52.74	57.91	95% Confidence Interval	(L) -71.83	(L) -19.95	(L) 26.83
	Q1, Q3	37.3, 102.15	84.76, 180.59	46.53, 104.09		(U) -19.74	(U) 18.36	(U) 61.81
	Min, Max	17.10, 364.00	23.68, 212.97	1.18, 772.46				
	Variance	4616.27	2781.26	3340.33				
Rotatable	*n*	49	37	746	Levene's Test	0.95	0.013	0.024
Bonds	*Median*	5	7	4	Bootstrap	0.041	0.06	0.00
	Mean	6.6	8.76	5.2	Mean Difference	-2.14	1.41	3.56
	SD of Mean	4.82	4.67	4.02	95% Confidence Interval	(L) -4.20	(L) 0.06	(L) 2.04
	Q1,Q3	3, 10.5	5.5, 12.5	2, 7		(U) -0.09	(U) 2.81	(U) 5.04
	Min, Max	0, 18	0, 18	0, 32				
	Variance	23.2	21.8	16.17				
Rule of 5	*n*	49	37	763	Pearson Chi-Square/	0.22^P	0.006^F	0.000^F
Violations	*Mean*	0.82	1.03	0.269	Fischer's Exact Test			
	SD of Mean	0.88	0.96	0.62				

Results of the pairwise comparisons completed using BDDCS I-IV classification groups. B = Bootstrap, t = t-test, W = Welch's test, P = Pearson Chi-Square, F = Fischer's Exact Test. Bootstrap 95% Confidence Interval based upon 5000 stratified bootstrap samples. (L) and (U) refer to lower and upper 95% confidence limits. For non-categorical variables showing normal distribution, when Levene's test was not significant, 95% Confidence intervals and sig. Level for groups comparison were based on 'equal variance assumed' calculations i.e independent samples t-test (2 sided). When Levene's test was significant, 95% Confidence intervals and sig. Level for group's comparison were based on 'equal variance not-assumed' calculations i.e Welch's test. For non-categorical variables not showing normal distribution the bootstrap method was used (5000 samples). Categorical variables i.e., Ro5, were analysed using Chi-Square tests. If 1 or more cells had an expected count below 5, Fisher's exact test was employed. A p-value of 0.05 was used as the significance level for all tests. SD refers to Standard Deviation of the Mean.

Table 1-2: Rule-of-5 Violations versus Drug Group Cross Tabulation (All BDDCS Classes):

			Drug Group			Total
			LBF	**SD**	**Others**	
Ro5	**No Greater than 1**	**Count**	40	26	714	780
		% of Group Total	81.6%	70.3%	93.6%	91.9%
	Greater than 1	**Count**	9	11	49	69
		% of Group Total	18.4%	29.7%	6.4%	8.1%
Total		**Count**	49	37	763	849
		% of Group Total	100.00%	100.00%	100.00%	100.00%

Table 1-3: Results of Statistical Analysis comparing LBF, SD and Others using BDDCS Class II/IV (Low Solubility).

Drug Property	Descriptors	LBF	SD	Others	Statistical Tests	p-value LBF vs SD	p-value LBF vs Others	p-value SD vs Others
clogP	*n*	38	30	307	Levene's Test	0.16	0.005	0.33
	Median	4.99	5.05	3.36	Welch's/ t-test	0.21[W]	0.000[W]	0.000[t]
	Mean	5.62	4.92	3.31	Mean Difference	0.70	2.31	1.61
	SD of Mean	0.47	1.57	0.12	95% Confidence Interval	(L) -0.39	(L) 1.34	(L) 0.86
	Q1, Q3	3.76, 7.36	3.82, 6.02	2.19, 4.40		(U) 1.79	(U) 3.27	(U) 2.35
	Min, Max	-0.73, 14.36	1.91, 7.63	-2.42, 10.97				
	Variance	8.25	2.45	4.03				
Hydrogen Bond Acceptors	*n*	38	30	307	Levene's Test	0.97	0.36	0.37
	Median	4	6	4	Bootstrap	0.00	0.31	0.00
	Mean`	4.34	7	4.81	Mean Difference	-2.66	-0.47	2.19
	SD of Mean	2.88	2.56	2.60	95% Confidence Interval	(L) -3.90	(L) -1.40	(L) 1.26
	Q1, Q3	2, 6	5, 10	3, 6		(U) -1.39	(U) 0.58	(U) 3.16
	Min-Max	1, 13	3, 12	0, 18				
	Variance	8.29	6.55	6.78				
Hydrogen Bond Donors	*n*	38	30	307	Levene's Test	0.43	0.23	0.04
	Median	1	2.50	1	Bootstrap	0.09	0.81	0.03
	Mean	1.68	2.27	1.63	Mean Difference	-0.58	0.06	0.64
	SD of Mean	1.38	1.46	1.23	95% Confidence Interval	(L) -1.23	(L) -0.36	(L) 0.11
	Q1, Q3	1, 3	1, 4	1, 2		(U) 0.09	(U) 0.54	(U) 1.17
	Min-Max	0, 5	0, 4	0, 7				
	Variance	1.90	2.13	1.52				
$logD_{7.4}$	*n*	38	30	181	Levene's Test	0.04	0.02	0.30
	Median	3.92	4.05	2.85	Bootstrap	0.80	0.001	0.000
	Mean	3.90	4.04	4.04	Mean Difference	-0.14	1.82	1.96
	SD of Mean	2.85	1.67	2.01	95% Confidence Interval	(L) -1.21	(L) 0.92	(L) 1.28
	Q1, Q3	2.10, 5.58	2.75, 5.39	0.73, 3.52		(U) 0.94	(U) 2.74	(U) 2.66
	Min-Max	-3.20, 11.35	1.28, 7.05	-3.68, 10.40				
	Variance	8.12	2.79	4.04				
logP	*n*	38	30	175	Levene's Test	0.05	0.12	0.31
	Median	4.51	4.62	3.12	Bootstrap	0.56	0.00	0.00
	Mean	4.84	4.60	3.06	Mean Difference	0.24	1.78	1.54
	SD of Mean	2.05	1.30	1.66	95% Confidence Interval	(L) -0.54	(L) 1.08	(L) 1.02
	Q1, Q3	3.72, 6.29	3.63, 5.73	2.24, 4.18		(U) 1.03	(U) 2.49	(U) 2.07
	Min-Max	0.28, 10	2.18, 6.92	-1.56, 7.80				
	Variance	4.19	1.69	2.76				
logS	*n*	36	29	228	Levene's Test	0.74	0.93	0.63
	Median	-5.13	-5.4	-4.2	Bootstrap	0.24	0.00	0.00
	Mean	-4.91	-5.29	-4.23	Mean Difference	0.38	-0.69	-1.06
	SD of Mean	1.14	1.31	1.03	95% Confidence Interval	(L) -0.29	(L) -1.05	(L) -1.54
	Q1,Q3	-5.70, -4.2	-5.80, -4.90	-4.9, -3.44		(U) 1.03	(U) -0.29	(U) -0.56

Drug Property	Descriptors	LBF	SD	Others	Statistical Tests	p-value LBF vs SD	p-value LBF vs Others	p-value SD vs Others
	Min-Max	-6.5, -1.21	-8.85, -0.57	-7.44, -1.00				
	Variance	1.29	1.71	1.07				
Maximum Dosage Strength (mg)	*n*	35	30	307	Levene's Test	0.19	0.008	0.00
	Median	75	100	100	Bootstrap	0.96	0.008	0.003
	Mean	116.64	118.01	195.50	Mean Difference	0-1.37	-78.86	-77.50
	SD of Mean	133.81	104.11	209.24	95% Confidence Interval	(L) -57.57	(L) -126.11	(L) -120.90
	Q1, Q3	10, 200	37.5, 200	30, 300		(U) 57.13	(U) -30.98	(U) -32.83
	Min-Max	0.0005, 500	1, 400	0.45, 300				
	Variance	17904.70	10838.05	43781.20				
Melting Point (^{O}C)	*n*	36	24	257	Levene's Test	0.04	0.20	0.15
	Median	153	173.75	182	Welch's/t-test	0.25^{W}	0.051^{t}	0.77^{t}
	Mean	162.97	179.93	183.46	Mean Difference	-16.96	-20.49	-3.53
	SD of Mean	65.49	46.81	57.73	95% Confidence Interval	(L) -44.88	(L) -41.06	(L) -27.44
	Q1, Q3	117.88, 223.55	143.25, 211.63	141.75, 224.00		(U) 11.50	(U) 0.08	(U) 20.38
	Min-Max	38, 284	80.5, 271	52, 349.84				
	Variance	4288.33	2191.41	3332.99				
Molecular Weight (g/mol)	*n*	38	30	307	Levene's Test	0.17	0.007	0.000
	Median	398.64	581.65	375.87	Bootstrap	0.002	0.129	0.000
	Mean	449.49	618.37	394.59	Mean Difference	-168.88	54.91	223.79
	SD of Mean	207.06	215.47	138.62	95% Confidence Interval	(L) -266.68	(L) -7.11	(L) 145.70
	Q1, Q3	315.45, 530.36	431.08, 812.76	296.54, 451.62		(U) -68.55	(U) 127.64	(U) 306.04
	Min-Max	153.14, 1202.61	346.34, 1113.20	136.11, 1058.06				
	Variance	42874.38	46426.99	19214.60				
pDose	*n*	35	30	307	Levene's Test	0.004	0.000	0.04
	Median	3.86	3.95	3.51	Bootstrap	0.34	0.026	0.037
	Mean	4.18	3.93	3.67	Mean Difference	0.25	0.52	0.27
	SD of Mean	1.38	0.59	0.77	95% Confidence Interval	(L) -0.24	(L) 0.07	(L) 0.04
	Q1, Q3	3.38, 4.48	3.46, 4.21	3.09, 4.15		(U) 0.78	(U) 0.99	(U) 0.50
	Min-Max	2.11, 8.32	3.04, 5.57	2.29, 6.03				
	Variance	1.91	0.36	0.59				
Percentage Excreted Unchanged in Urine (%)	*n*	31	27	262	Levene's Test	0.13	0.001	0.000
	Median	0.5	0.03	1.5	Bootstrap	0.39	0.014	0.000
	Mean	3.98	1.36	1.36	Mean Difference	2.61	-7.77	-10.38
	SD of Mean	11.81	4.64	21.41	95% Confidence Interval	(L) -0.78	(L) -11.88	(L) -13.38
	Q1, Q3	0.05, 2.2	0, 0.5	0.29, 10		(U) 7.15	(U) -2.73	(U) -7.41
	Min-Max	0, 65	0, 24	0, 100				
	Variance	139.67	21.56	458.18				
pKa (strongest acid)	*n*	37	25	272	Levene's Test	0.23	0.14	0.70
	Median	10.6	9.33	10.29	Bootstrap	0.45	0.81	0.44
	Mean	10.11	9.04	9.85	Mean Difference	1.06	0.25	-0.81
	SD of Mean	0.99	5.01	5.13	95% Confidence Interval	(L) -1.68	(L) -1.78	(L) -2.89
	Q1, Q3	4.25, 14.04	3.99, 12.63	4.74, 13.78		(U) 3.78	(U) 2.4	(U) 1.25
	Min-Max	0.27, 22	0, 19.90	-12, 19.96				

Drug Property	Descriptors	LBF	SD	Others	Statistical Tests	p-value LBF vs SD	p-value LBF vs Others	p-value SD vs Others
	Variance	36.4	25.06	26.33				
Polar Surface Area (Å^2)	*n*	38	30	306	Levene's Test	0.56	0.22	0.05
	Median	55.4	116.43	76.15	Bootstrap	0.00	0.37	0.00
	Mean	74.63	130.08	82.85	Mean Difference	-55.45	-8.23	47.22
	SD of Mean	54.39	49.79	43.78	95% Confidence Interval	(L) -79.78	(L) -24.99	(L) 28.55
	Q1, Q3	37.3, 98.56	90.16, 182.69	54.8, 104.60		(U) -29.24	(U) 10.32	(U) 65.61
	Min-Max	20.23, 279	46.53, 204	1.18, 266.66				
	Variance	2957.95	2478.83	1917.02				
Rotatable Bonds	*n*	38	30	306	Levene's Test	0.76	0.04	0.13
	Median	5	7	5	Bootstrap	0.050	0.274	0.001
	Mean	6.53	8.8	5.62	Mean Difference	-2.27	0.90	3.18
	SD of Mean	4.88	4.39	4.10	95% Confidence Interval	(L) -4.46	(L) -0.65	(L) 1.51
	Q1, Q3	3, 11	5.75, 12.25	3, 7		(U) -0.08	(U) 2.53	(U) 4.87
	Min-Max	0, 18	3, 18	0, 24				
	Variance	23.8	19.27	4.10				
Rule of 5 Violations	*n*	34	27	239	Pearson Chi-Square/	0.159^P	0.086^F	0.001^F
	Mean	0.9412	1.148	0.343	Fischer's Exact Test			
	SD	0.8507	0.9488	0.6542				

Results of the pairwise comparisons completed using BDDCS II/IV classification groups. B = Bootstrap, t = t-test, W = Welch's test, P = Pearson Chi-Square, F = Fischer's Exact Test. Bootstrap 95% Confidence Interval based upon 5000 stratified bootstrap samples. (L) and (U) refer to lower and upper 95% confidence limits. For non-categorical variables showing normal distribution, when Levene's test was not significant, 95% Confidence intervals and sig. Level for groups comparison were based on 'equal variance assumed' calculations i.e independent samples t-test (2 sided). When Levene's test was significant, 95% Confidence intervals and sig. Level for group's comparison were based on 'equal variance not-assumed' calculations i.e Welch's test. For non-categorical variables not showing normal distribution the bootstrap method was used (5000 samples). Categorical variables i.e., Ro5, were analysed using Chi-Square tests. If 1 or more cells had an expected count below 5, Fisher's exact test was employed. A p-value of 0.05 was used as the significance level for all tests. SD refers to Standard Deviation of the Mean.

Table 1-4 Rule-of-5 Violations versus Drug Group Cross Tabulation (BDDCS Class II/IV)

			Drug Group			Total
			LBF	**SD**	**Others**	
Ro5	**No Greater than 1**	**Count**	31	20	279	330
		% of Group Total	81.60%	66.70%	90.90%	88.0%
	Greater than 1	**Count**	7	10	28	45
		% of Group Total	18.40%	33.30.%	9.10%	12.00%
Total		**Count**	38	30	307	375
		% of Group Total	100.00%	100.00%	100.00%	100.00%

Table 1-5: Tabular representation of SD commercial products.

Trade Name	Drug	Dosage Form/Strength	Excipients*	Method of Manufacturer
Afeditab CR®	Nifedipine	Tablet (30mg)	Poloxamer/PVP	Spray Drying
Afinitor®	Everolimus	Tablet (2.5,5, 7.5, 10mg)	HPMC	Spray Drying
Astagraf XL®	Tacrolimus	Capsule (0.5, 1, 5mg)	HPMC	Wet Granulation
Belsomra®	Suvorexant	Tablet (5, 10, 15, 20mg)	Polyvinylpyrrolidone/ Vinyl Acetate Copolymer (Copovidone)	Melt Extrusion
Certican®	Everolimus	Tablet (0.25, 0.5, 0.75, 1mg)	HPMC	Spray Drying
Cesamet®	Nabilone	Capsule (1mg)	Povidone	Solvent Evaporation
Cokiera®	Dasabuvir/ Ombitasvir/ Paritaprevir/ Ritonavir	Tablet (200/8.33/50/33.33mg)	Copovidone	Melt Extrusion
Crestor®	Rosuvastatin Calcium	Tablet (5, 10, 20, 40mg)	HPMC	Spray Drying
Cymbalta®	Duloxetine	Capsule (30, 60mg (+20mg FDA))	HPMCAS	
Deltyba®	Delamanid	Tablet (50mg)	Hypromellose Phthalate (HPMCP)	
Envarsus XR®	Tacrolimus	Tablet (0.75, 1, 4mg)	HPMC	Melt Granulation
Epclusa®	Sofosbuvir/ Velpatasvir	Tablet (400/100mg)	Copovidone	Spray Drying
Eucreas®	Vildagliptin/ Metformin HCL	Tablet (50/850mg + 50/1000mg)	HPC	Hot Melt Extrusion
Fenoglide®	Fenofibrate	Tablet (40, 120mg)	PEG 6000, Poloxamer 188	Spray Melt
Galvumet®	Vildagliptin /Metformin HCL	Tablet (50/850mg + 50/1000mg)	HPC	Hot Melt Extrusion
Gris-PEG®	Griseofulvin	Tablet (125, 250mg)	PEG 400 and 8000, Povidone	Melt-Extrusion
Harvoni®	Ledipasvir/ Sofosbuvir	Tablet (90/400, 45/200mg)	Copovidone	Spray Drying
Incivek®	Telaprevir	Tablet (375mg)	HPMCAS	Spray Drying
Incivo®	Telaprevir	Tablet (375mg)	HPMCAS	Spray Drying
Intelence®	Etravirine	Tablet (25, 100, 200mg)	HPMC	Spray Drying
Isoptin SR-E 240®	Verapamil	Tablet (240mg)	HPMC/HPC	Spray Drying
Kaletra®	Lopinavir/Ritonavir	Tablet (100/25, 200/50mg)	PVP	Melt Extrusion
Kalydeco®	Ivacaftor	Tablet (75, 150mg)	HPMCAS	Spray Drying
Mavyret®	Glecaprevir/ Pibrentasvir	Tablet (40/100mg)	Copovidone (Type K 28)	Melt Extrusion
Modigraf®	Tacrolimus	Granules for Oral Suspension (0.2,1mg)	HPMC	Spray Drying
Nimotop®	Nimodipine	Tablet (30mg)	PEG	Spray Drying/ Fluid Bed
Nivadil®	Nilvadipine	Capsule (16mg,8mg)	HPMC	Spray Drying
Norvir®	Ritonavir	Tablet (100mg)	PVP VA 64	Melt Extrusion
Noxafil®	Posaconazole	Tablet (100mg)	HPMCAS	Melt Extrusion
Onmel®	Itraconazole	Tablet (200mg)	PVP VA 64	Melt-Extrusion
Orkambi®	Lumacaftor/ Ivacaftor	Tablet (100mg/125mg, 200mg/125mg)	HPMCAS	Spray Drying
Prograf®	Tacrolimus	Capsule (0.5, 1, 3, 5mg)	HPMC	Spray Drying
Rezulin®	Troglitazone	Tablet (200, 300, 400mg)	PVP	Spray Drying
Samsca®	Tolvaptan	Tablet (15, 30 + 60mg)	HPMC	Granulation

Trade Name	Drug	Dosage Form/Strength	Excipients*	Method of Manufacturer
Shui linjia	Silibinin	Capsule (70mg)	Lecithin	
Sporanox®	Itraconazole	Capsule (100mg)	HPMC	Fluid Bed Bead Layering
Stivarga®	Regorafenib	Tablet (40mg)	Povidone K25	
Venclexta®	Venetoclax	Tablet (10, 50, 100mg)	Copovidone	Melt Extrusion
Viekira XR®	Dasabuvir/ Ombitasvir/ Paritaprevir/ Ritonavir	Tablet (200/8.33/50/33.33mg)	Copovidone	Melt Extrusion
Votubia®	Everolimus	Tablet (2.5, 5, 10mg)	HPMC	Spray Drying
Zelboraf®	Vemurafenib	Tablet (240mg)	HPMCAS	Solvent/Anti-Solvent Precipitation
Zepatier®	Elbasvir/ Grazoprevir	Tablet (50/100mg)	TPGS, Copovidone, HPMC	Spray Drying
Zortress®	Everolimus	Tablet (0.25, 0.5, 0.75, 1mg)	HPMC	Spray Drying

Data obtained from FDA Drug Label (from Drugs@FDA database), European Summary of Pharmaceutical Characteristics (SPC), Health Products Regulatory Authority (HPRA) National Drug Authorisation SPC or Therapeutic Goods Administration (TGA) product information. *Excipients listed refer only to selected relevant excipients from the total excipients of the drug products which contribute directly to the transformation and/or stability of a drug as a SD.

Table 1-6: Tabular representation of LBF commercial products.

Trade Name	Drug	Dosage Form/Strength	Excipients*
Absorica®	Isotretinoin	Hard Gelatine Capsule (10,20,25,30,35,40mg)	Sorbitan Monooleate, Soybean Oil and Stearoyl Polyoxylglycerides
Accutane®	Isotretinoin	Soft Gelatine Capsule (10,20,40mg)	Beeswax, Hydrogenated Soybean Oil Flakes, Hydrogenated Vegetable Oil, Soybean Oil
Advil Cold and Sinus®	Ibuprofen	Liquid Gel Capsule (200mg/30mg)	Fractionated Coconut Oil, Poly Ethylene Glycol
Agenerase®	Amprenavir	Soft Gelatine Capsule (50, 150mg)	Polyethylene Glycol 1000 Succinate (TPGS), Polyethylene Glycol 400 (PEG 400), Propylene Glycol
Aloxi®	Palonosetron	Soft Gelatine Capsule (0.5mg)	Mono- and di-glycerides of Capryl/Capric acid, Glycerin, Polyglyceryl Oleate, Water, and Butylated Hydroxyanisole
Amitiza®	Lubiprostone	Soft Gelatine Capsule (8, 24mcg)	Medium-Chain Triglycerides
Aptivus®	Tipranivir	Soft Gelatine Capsule (250mg)	Macrogolglycerol Ricinoleate, Ethanol, Mono/diglycerides of Caprylic/Capric acid, Propylene Glycol.
Aptivus®	Tipranivir	Oral Solution (100mg/mL)	Macrogol, Polyethylene Glycol, Propylene Glycol, Mono/Diglycerides of Caprylic/Capric Acid, Polyoxyl 35 Caster Oil, Vitamin E Polyethylene Glycol Succinate (TPGS).
Avodart®	Dutasteride	Soft Gelatine Capsule (0.5mg)	Mono- and Diglycerides of Caprylic/Capric acid
Cipro®	Ciprofloxacin	Oral Suspension (250mg/mL, 500mg/5mL)	Medium Chain Triglycerides
Claravis®	Isotretinoin	Liquid Filled Hard Shell Capsule (10,20,30,40mg)	Hydrogenated Vegetable Oil, Polysorbate 80, Soybean Oil.
Clarityn®	Loratadine	Soft Gelatine Capsule (10mg)	Caprylic/Capric Glycerides, Glycerin, Polysorbate 80.
Convulex®	Valproic Acid	Soft Gelatine Capsule (150, 300, 500mg)	Macrogol 6000, Glycerol Monostearate 44-55 Type II
Depakene®	Valproic Acid	Soft Gelatine Capsule (250mg)	Corn Oil
Detrol La®	Tolterodine Tartrate	Extended Release Gelatine Capsule (2, 4mg)	Medium Chain Triacylglycerides, Oleic Acid, Gelatin.
Drisdol®	Ergocalciferol	Liquid Filled Hard Shell Capsule (1.25mg)	Glycerin, Soybean Oil, Edible Vegetable Oil.
Epadel®(36)	Ethyl Eicosapentaenoate	Soft Gelatine Capsule (500mg)	Alpha Tocopherol
Fenogal®	Fenofibrate	Hard Gelatine Capsule (200mg)	Lauryl Macroglycerides, Macrogol 20,000
Fortovase®	Saquinavir	Soft Gelatine Capsule (200mg)	Medium Chain Mono- and Diglycerides.
Gengraf®	Cyclosporin	Hard Gelatine Capsule (25, 100mg) (50mg discontinued)	Polyethylene Glycol, Polyoxyl 35 Castor Oil, Polysorbate 80, Propylene Glycol, Ethanol.
Gengraf®	Cyclosporin	Oral Solution (100mg/mL)	Polyoxyl 40, Hydrogenated Castor Oil, Polysorbate 80, Propylene Glycol

Trade Name	Drug	Dosage Form/Strength	Excipients*
Glakay®	Menatetrenone	Soft Gelatine Capsule (15mg)	Carnauba Wax, Hydrogenated Oil, Glyceryl Monooleate, PG Esters of Fa, Glycerin.
Hectorol®	Doxercalciferol	Soft Gelatine Capsule (0.5, 1, 2.5mcg)	Ethanol, Fractionated Triglyceride of Coconut Oil
Heminevrin®	Clomethiazole	Soft Gelatine Capsule (192mg)	Medium Chain Triglycerides, Glycerol
Hycamtin®	Topotecan	Liquid Filled Hard Shell Capsule (0.25, 1mg)	Hydrogenated Vegetable Oil, Glyceryl monostearate
Infree®	Indomethacin	Capsule (100, 200mg)	Cremophor RH 60
Juvela N®	Tocopherol Nicotinate	Soft Gelatine Capsule (200mg)	Carnauba Wax, Medium Chain Triglycerides, Glycol Esters of Fatty Acids, Glycerin.
Kaletra®	Lopinavir/ Ritonavir	Soft Gelatine Capsule (133.3mg/33.3mg)	Glycerin, Oleic Acid, Polyoxyl 35 Castor Oil, Propylene Glycol.
Kaletra®	Lopinavir/ Ritonavir	Oral Solution (80+20mg/mL)	Ethanol, Glycerin, Polyoxyl 40 Hydrogenated Castor Oil, Propylene Glycol.
Ketas®	Ibudilast	Sustained Release Granules (10mg)	Hydrogenated Castor Oil, Macrogol 6000, Cremophor RH 60.
Lamprene®	Clofazimine	Soft Gelatine Capsule (50, 100mg)	Beeswax, Glycerin, Lecithin, Plant Oils, Propylene Glycol.
Lipofen®	Fenofibrate	Hard Shell Capsule (50, 150mg) (100mg discontinued)	Gelucire 44/14, Polyethylene Glycol 20,000, Polyethylene Glycol 8000, Propylene Glycol
Lovaza®	Omega-3 Acid Ethyl Esters	Soft Gelatine Capsule (900mg/gram)	Soybean Oil.
Marinol®	Dronabinol	Soft Gelatine Capsule (2.5, 5, 10mg)	Sesame Oil.
MXL®	Morphine	Prolonged Release Capsule (30, 60, 90, 120,150,200mg)	Hydrogenated Vegetable Oil BP, Macrogol 6000 Ph Eur
Navelbine®	Vinorelbine	Soft Gelatine Capsule (20, 30, 80mg)	Anhydrous Ethanol, Glycerol Macrogol 400
Neoral®	Ciclosporin	Soft Gelatine Capsule (25, 50, 100mg)	Alpha-tocopherol, Ethanol, Propylene Glycol, Glycerol, Corn oil-mono-di-triglycerides, Macrogolglycerol hydroxystearate / Polyoxyl 40 hydrogenated castor oil.
Neoral®	Ciclosporin	Oral Solution (100 mg/mL)	Alpha–tocopherol, Ethanol, Propylene Glycol, Corn oil-mono-di-triglycerides, Macrogolgylcerol Hydroxystearate / Polyoxy 40 Hydrogenated Castor Oil.
Nimotop®	Nimodipine	Soft Gelatine Capsule (30mg)	Glycerin, Peppermint oil, Polyethylene Glycol 400
Norvir®	Ritonavir	Oral Solution (80 mg/mL)	Polyoxyl 35 Castor oil, Propylene Glycol, Ethanol.
Norvir®	Ritonavir	Soft Gelatine Capsule (100 mg)	Ethanol, Oleic Acid, Polyoxyl 35 Castor Oil.
Ofev®	Nintedanib	Soft gelatine capsule (100mg, 150mg)	Triglycerides (Medium-Chain), Hard Fat Lecithin (soya)
One-Alpha®	Alfacalcidol	Soft Gelatine Capsule (1mcg)	Sesame Oil (refined)
Panimun Bioral®	Cyclosporin	Soft Gelatine Capsule (25, 50, 100mg)	Ethanol, Propylene Glycol, Corn Oil Mono/Di/Tri-Glycerides, Macrogolglycerol hydroxystearate / Polyoxyl 40 Hydrogenated Caster Oil, Ethanol.

Trade Name	Drug	Dosage Form/Strength	Excipients*
Pentasa®	Mesalazine	Extended-Release Capsule (250, 500mg)	Acetylated Monoglyceride, Castor Oil
Prometrium®	Progesterone	Soft Gelatine Capsule (100, 200,300mg)	Peanut Oil, Glycerin, Lecithin.
Rapamune®	Sirolimus	Oral Solution (1mg/mL)	Polysorbate 80 (E433), Phosal 50 PG (Phosphatidylcholine, Propylene Glycol, Mono-and Diglycerides, Ethanol, Soya Fatty Acids and Ascorbyl Palmitate).
Rayaldee®	Calcifediol	Extended-Release Capsule (0.03mg)	Mixture of Lipophilic Emusifier with a HLB <7 and an absorption enhancer, oily vehicle - mineral oil, liquid paraffins or squalene.
Restandol Testocaps®	Testosterone	Soft Gelatine Capsule (40mg)	Castor Oil and Propylene Glycol Monolaurate (E477)
Roaccutane®	Isotretinoin	Soft Gelatine Capsule (10, 20mg)	Beeswax, Soya-Bean Oil (refined), Soya-Bean Oil (hydrogenated). Soya-bean Oil (Partially Hydrogenated)
Rocaltrol®	Calcitriol	Soft Gelatine Capsule (0.25, 0.5mcg)	Fractionated Triglycerides of Coconut Oil
Sandimmune®	Ciclosporin	Oral Solution (100 mg/mL)	Alcohol dissolved in Olive Oil, Ph. Helv./Labrafil M 1944 CS (Polyoxyethylated Oleic Glycerides) Vehicle
Sandimmune®	Ciclosporin	Soft Gelatine Capsule (25, 50 and 100mg)	Corn Oil, Linoleoyl Macrogolglycerides, Glycerol, Ethanol.
Selbex®	Teprenone	Hard Gelatine Capsule (50mg)	Alpha-tocopherol, Macrogol 6000
Solufen®	Ibuprofen	Hard Gelatine Capsule (200mg)	Gelucire 44/14
Sustiva®	Efavirenz	Oral Solution (30mg/mL)	Medium Chain Triglycerides
Targretin®	Bexarotene	Soft Gelatine Capsule (75mg)	Polysorbate 20, PEG400
Thorens®	Cholecalcifer-ol	Oral Drops Solution (10000IU/mL, 25000IU/2.5mL)	Refined Olive Oil
Tirosint®	Levothyroxine	Soft Gelatine Capsule (0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.112, 0.137, 0.088, 0.174, 0.200, 0.013mg)	Glycerin
Uvedose®	Cholecalcifer-ol	Oral Solution (100,000IU/2mL)	Glycolyzed Polyoxyethylenated Glycerides
Vesanoid®	Tretinoin	Soft Gelatine Capsule (10mg)	Beeswax, Hydrogenated Soybean Oil Flakes, Hydrogenated Vegetable Oils and Soybean Oil
Vyndaqel®	Tafamidis	Soft Gelatine Capsule (20mg)	Macrogol 400, Polysorbate 20, Butylated hydroxytoluene
Xtandi®	Enzalutamide	Soft Gelatine Capsule (40mg)	Caprylocaproyl Polyoxylglycerides.
Zantac®	Ranitidine	Soft gelatine capsule (150, 300mg)	Medium Chain Triglycerides, Gelucire 33/01
Zemplar®	Paricalcitol	Soft Gelatine Capsule (1, 2mcg)	Medium Chain Triglycerides (fractionated from coconut oil or palm kernel oil), Alcohol
Zipsor®	Diclofenac Potassium	Soft Gelatine Capsule (25mg)	ProSorb (proprietary combination of Polyethylene Glycol 400, Glycerin, Sorbitol, Povidone, Polysorbate 80, and Hydrochloric Acid), Isopropyl Alcohol, and Mineral Oil

Trade Name	Drug	Dosage Form/Strength	Excipients*
Zmax®	Azithromycin	Extended-Release Oral Suspension (27mg/mL)	Glyceryl Behenate

Data obtained from FDA Drug Label (from Drugs@FDA database), European Summary of Pharmaceutical Characteristics (SPC), Health Products Regulatory Authority (HPRA) National Drug Authorisation SPC or Medicines and Healthcare Products Regulatory Agency (MHRA) SPC unless otherwise stated.*Excipients listed refer only to selected relevant excipients from the total excipients of the drug products which include both lipophilic and hydrophilic excipients types as classified by the lipid formulation classification system.

Appendix 2: Chapter 3

Table 2-1: Compilation of licensed oral medicines from 2016-2020 and their $AUC_{fed/fasted}$ ratio, clinical recommendation regarding food take, BCS and food effect (FE) Classification. LF, MF and HF refer to Low Fat, Medium Fat and High Fat Meals respectively, * refers to the 90% CI limits and ** refers to the range of values quoted.

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	$AUC_{fed/fasted}$	BCS Class
2020	Avapritinib	Ayvakit/Ayvakyt	Taken on an empty stomach, at least one hour before and two hours after a meal	Positive	1.27/1.29	2
2020	Glasdegib	Daurismo	Taken with or without food	No FE	0.84	4
2020	Lefamulin	Xenleta	Taken on an empty stomach, at least 1 hour before or 2 hours after a meal	No FE	0.82	3
2020	Pralsetinib	Gavreto	Taken on an empty stomach (no food intake for at least 2 hours before and at least 1 hour after taking)	Positive	2.22	2
2020	Osilodrostat	Isturisa	Taken with or without food	No FE	0.89	1
2020	Filgotinib	Jyseleca	Taken with or without food	No FE	1	2
2020	Ivacaftor Tezacaftor Elexacaftor	Kaftrio	Taken with fat-containing food	Positive Positive No FE	1.9-2.5** 2.5-4** 1	2 2 4
2020	Selumetinib	Koselugo	Take on an empty stomach. Do not consume food 2 hours before each dose or 1 hour after each dose	No FE	0. 62	4
2020	Nifurtimox	Lampit	Taken with food	Positive	1.71	2
2020	Siponimod	Mayzent	Taken with or without food	No FE	1	2
2020	Bempedoic Acid	Nilemdo	Taken with or without food	No FE	1	2
2020	Darolutamide	Nubeqa	Taken with food	Positive	2-2.5	2
2020	Bempedoic Acid Ezetimibe	Nustendi/Nexlizet	Taken with or without food	No FE No FE	1 1	2 2
2020	Azacitidine	Onureg	Take with or without food	Negative	0.79	1
2020	Elagolix Sodium Estradiol Norethindrone Acetate	Oriahnn	No instructions with regard to food intake	Negative No FE No FE	0.75 1 1.23	3 1 2
2020	Alpelisib	Piqray	Taken immediately after food, at approximately the same time each day	Positive	LF 1.77 HF 1.73	2
2020	Pretomanid	Pretomanid FGK	Taken with food	Positive	1.88	2

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	$AUC_{fed/fasted}$	BCS Class
2020	Selpercatinib	Retevmo	Taken with or without food	No FE	1	
2020	Solriamfetol	Sunosi	Taken with or without food	No FE	1	1
2020	Capmatinib Hydrochloride	Tabrecta	Taken with or without food	Positive	LF 1 HF 1.46	2
2020	Fostamatinib	Tavlesse	Taken with or without food	No FE	1.23	4
2020	Dolutegravir Sodium	Tivicay pd	Taken with or without food	Positive	1.66	4
2020	Tucatinib	Tukysa	Taken with or without food	Positive	1.5	2
2020	Solifenacin Succinate	Vesicare LS	Avoid taking with food due to bitter taste	No FE	1	1
2020	Enzalutamide	Xtandi	Taken with or without food	No FE	1	2
2020	Ozanimod	Zeposia	Taken with or without food	No FE	1	2
2019	Isotretinoin	Absorica Ld	Taken with or without food	No FE	1.2	2
2019	Lumateperone Tosylate	Caplyta	Taken with food	No FE	1.09	1
2019	Ivabradine	Corlanor	Taken with food	Positive	1.2-1.4**	1
2019	Trientine Dihydrochloride	Cufence	Take this medicine with water only. Avoid eating or drinking (except water) for 1 hour before, or 2 hours after taking.	Negative	0.55	3
2019	Lemborexant	Dayvigo	Taken immediately before going to bed	No FE	1.18	2
2019	Avatrombopag Maleate	Doptelet	Taken with food	No FE	LF 1.0 HF 1.0	4
2019	Dolutegravir Sodium Lamivudine	Dovato	Taken with or without food	Positive No FE	1.33 1	4 3
2019	Triclabendazole	Egaten	Taken with food	Positive	2	2
2019	Gilteritinib Fumarate	Xospata	Taken with or without food	No FE	0.9	4
2019	Riluzole	Xservan	Taken at least 1 hour before or 2 hours after a meal	No FE	0.85	2
2019	Apalutamide	Erleada	Taken with or without food	No FE	1	2
2019	Colchicine	Gloperba	Taken with or without food	No FE	0.93	3
2019	Ledipasvir Sofosbuvir	Harvoni	Taken with or without food	No FE Positive	1 ~2	2 3
2019	Fedratinib Hydrochloride	Inrebic	Taken with or without food	No FE	LF 1.24 HF 1.24	2
2019	Lorlatinib	Lorviqua	Taken with or without food	No FE	1.05	4
2019	Lusutrombopag	Mulpleo	Taken with or without food	No FE	1	4
2019	Istradefylline	Nourianz	Taken with or without food	Positive	1.64	2
2019	Voxelotor	Oxbryta	Taken with or without food	Positive	1.42	2
2019	Naldemedine	Rizmoic/Symproic	Taken with or without food	No FE	1	4

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	$AUC_{fed/fasted}$	BCS Class
2019	Amifampridine	Ruzurgi	Taken with or without food	No FE	1	3
2019	Talazoparib	Talzenna	Taken with or without food	No FE	1	2
2019	Elexacaftor Ivacaftor Tezacaftor	Trikafta (Copackaged)	Taken with fat containing food	Positive Positive No FE	1.9-2.5** 2.5-4** 1	4 2 2
2019	Pexidartinib Hydrochloride	Turalio	Taken on an empty stomach, at least 1 hour before or 2 hours after a meal or snack	Positive	2	2
2019	Ubrogepant	Ubrelvy	Taken with or without food	No FE	1	4
2019	Larotrectinib	Vitrakvi	Taken with or without food	No FE	1	1
2019	Dacomitinib	Vizimpro	No instructions with regard to food intake	No FE	1	2
2019	Ceritinib	Zykadia	Taken with food	Positive	LF 1.64 HF 1.39	4
2019	Sotagliflozin	Zynquista	Taken once daily before the first meal of the day	Positive	1.5	2
2018	Brigatinib	Alunbrig	Taken with or without food	No FE	1	1
2018	Tafenoquine Succinate	Arakoda	Taken with food	Positive	1.41	
2018	Bictegravir Emtricitabine Tenofovir Alafenamide	Biktarvy	Taken with or without food	No FE No FE Positive	1.24 1 1.64	2 1 3
2018	Estradiol Progesterone	Bijuva	Taken with food	No FE Positive	1 1.79	1 2
2018	Encorafenib	Braftovi	Taken with or without food	No FE	1	2
2018	Duvelisib	Copiktra	Taken with or without food	Negative	0.63	4
2018	Doravirine Lamivudine Tenofovir Disoproxil Fumarate	Delstrigo	Taken with or without food	No FE No FE Positive	1.1 0.93 1.27	2 3 3
2018	Baloxavir Marboxil	Xofluza	Taken with or without food	No FE	0.64	2
2018	Ibrutinib	Imbruvica	No instructions with regard to food intake	Positive	2	2
2018	Dolutegravir Sodium Rilpivirine Hydrochloride	Juluca	Taken with a meal	Positive Positive	1.87 1.72	4 2
2018	Tolvaptan	Jynarque	Taken with or without food	No FE	1	4
2018	Tafenoquine Succinate	Krintafel	Taken with food	Positive	1.41	
2018	Binimetinib	Mektovi	Taken with or without food	No FE	1	2
2018	Mexiletine Hydrochloride	Namuscla	Should be swallowed with water. In case of digestive intolerance, capsules should be taken during a meal.	No FE	1	1

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	$AUC_{fed/fasted}$	BCS Class
2018	Neratinib	Nerlynx	Taken with food, preferably in the morning	Positive	2.2	4
2018	Omadacycline Tosylate	Nuzyra	Fast for at least 4 hours and then take	No FE	0.39	3
2018	Elagolix Sodium	Orilissa	Taken with or without food	No FE	0.76	3
2018	Ivacaftor Lumacaftor	Orkambi	Mixed with one teaspoon (5 mL) of age-appropriate soft food or liquid and the mixture completely consumed.	Positive	3 2	2 2
2018	Doravirine	Pifeltro	Taken with or without food	No FE	1.16	2
2018	Letermovir	Prevymis	Taken with or without food	No FE	0.99	2
2018	Tacrolimus	Prograf	Taken consistently with or without food.	No FE	0.63	2
2018	Rucaparib Camsylate	Rubraca	Taken with or without food	Positive	1.38	2
2018	Brexpiprazole	Rxulti	Taken with or without food	No FE	1	2
2018	Sarecycline Hydrochloride	Seysara	Taken with or without food	Negative	0.73	3
2018	Ertugliflozin	Steglatro	Taken with or without food	No FE	1	1
2018	Ertugliflozin Sitagliptin	Steglujan	Taken with or without food	No FE No FE	1 1	1 2
2018	Ivacaftor Tezacaftor	Symdeko/Symkevi	Taken with fat-containing food	Positive No FE	3 1	2 2
2018	Ivosidenib	Tibsovo	Taken with or without food. Do not administer with a high fat meal due to increase in concentration	Positive	1.98	2
2018	Riluzole	Tiglutik Kit	Taken at least 1 hour before or 2 hours after a meal	No FE	0.91	2
2018	Tecovirimat	Tpoxx	Taken within 30 minutes after a full meal of moderate or high fat	Positive	1.39	2
2018	Abemaciclib	Verzenios	Taken with or without food	No FE	1.09	3
2018	Abiraterone Acetate	Yonsa	Taken with or without food	Positive	4.4	4
2017	Alectinib	Alecensa	Taken with food	Positive	3	4
2017	Deutetrabenazine	Austedo	Taken with food	No FE	1	2
2017	Betrixaban	Bevyxxa	Taken with food	Negative	LF 0.39 HF 0.52	3
2017	Acalabrutinib	Calquence	Taken with or without food	No FE	1	2
2017	Spironolactone	Carospir	Taken with or without food, but should be taken consistently with respect to food	Positive	1.9	2
2017	Allopurinol Lesinurad	Duzallo	Taken with food	No FE No FE	1 1	4 2

Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	$AUC_{fed/fasted}$	BCS Class
2017	Deflazacort	Emflaza	Taken with or without food. Tablet and Suspension	No FE	1	
2017	Tofacitinib	Xeljanz	Taken with or without food	No FE	1	3
2017	Telotristat Etiprate	Xermelo	Taken with food	No FE	3.64	
2017	Pirfenidone	Esbriet	Taken with food	No FE	0.84	1
2017	Tivozanib	Fotivda	Taken with or without food	No FE	1	2
2017	Valbenazine Tosylate	Ingrezza	Taken with or without food	No FE	0.87	1
2017	Deferasirox	Jadenu Sprinkle	Taken on an empty stomach or with a light meal	No FE	LF 1 HF 1.18	2
2017	Ribociclib Succinate	Kisqali	Taken with or without food	No FE	1	4
2017	Macimorelin Acetate	Macrilen	Taken after fasting for at least 8 hours	Negative	0.51	
2017	Cladribine	Mavenclad	Taken with or without food	Negative	1	3
2017	Glecaprevir Pibrentasvir	Maviret	Taken at the same time with food	Positive Positive	1.83-2.63** 1.4-1.53**	4 4
2017	Pitavastatin Sodium	Nikita	Taken with or without food	No FE	1	2
2017	Ritonavir	Norvir	Should be mixed with soft food	Negative	0.51	4
2017	Baricitinib	Olumiant	Taken with or without food	No FE	0.86	3
2017	Valsartan	Prexxartan	No instructions with regard to food intake	No FE	0.92	2
2017	Cariprazine	Reagila	Taken with or without food	No FE	1.12	2
2017	Edoxaban	Roteas	Taken with or without food	No FE	1	4
2017	Oxycodone Hydrochloride	Roxybond	No instructions with regard to food intake	No FE	1.23	3
2017	Midostaurin	Rydapt	Taken with food	No FE	1.6	2
2017	Tenofovir Alafenamide Darunavir Cobicistat Emtricitabine	Symtuza	Taken with food	No FE Positive Positive No FE	1.20 1.52 1.4 1	3 2 2 1
2017	Tenofovir Alafenamide	Vemlidy	Taken with food	Positive	1.51-1.81**	3
2017	Niraparib	Zejula	Taken with or without food	No FE	1	1
2017	Pitavastatin Magnesium	Zypitamag	Taken with or without food	No FE	1	2
2016	Brivaracetam	Briviact	Taken with or without food	No FE	0.95	1
2016	Emtricitabine Tenofovir Alafenamide	Descovy	Taken with or without food	No FE Positive	1 1.17-1.77**	1 3

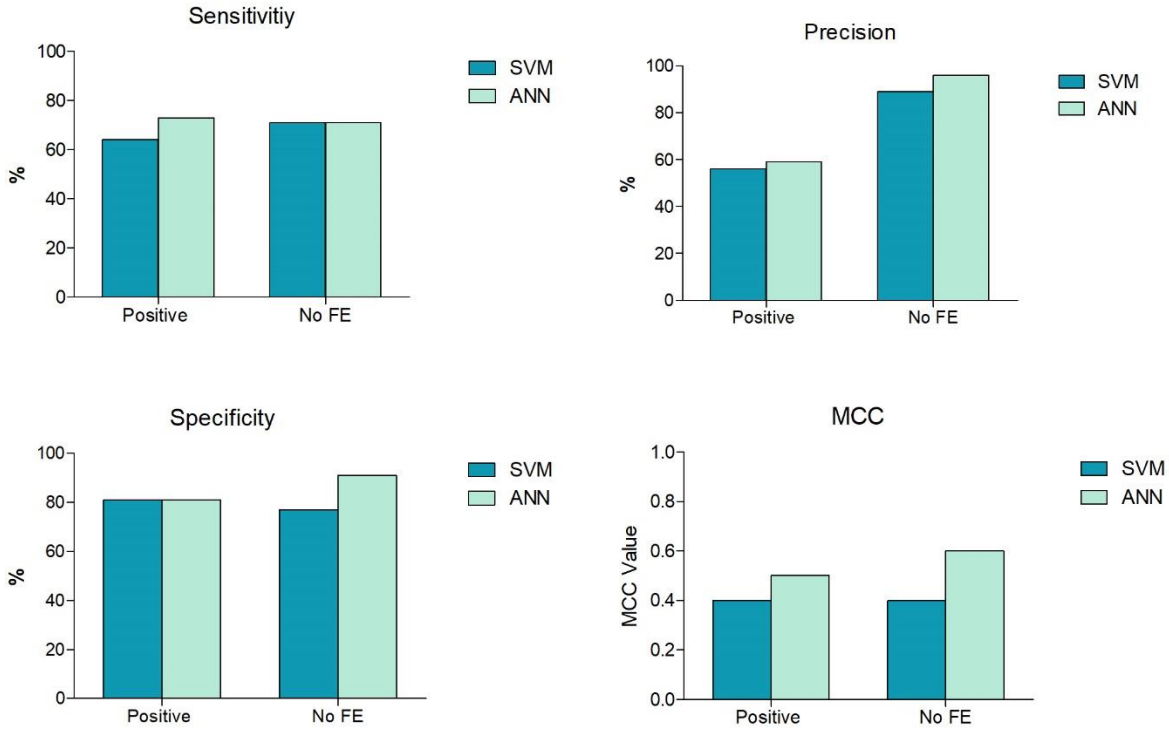
Year Licensed	Generic Name	Commercial Name	Clinical Recommendation	Food Effect Classification	$AUC_{fed/fasted}$	BCS Class
2016	Sofosbuvir Velpatasvir	Epclusa	Taken with or without food	Positive No FE	1.78 1.21	3 4
2016	Migalastat Hydrochloride	Galafold	Food should not be consumed at least 2 hours before and 2 hours after taking to give a minimum 4 hours fast	Negative	0.63-0.58**	3
2016	Empagliflozin Lignagliptin	Glyxambi	Taken with or without food	No FE No FE	0.84 1	3 3
2016	Palbociclib	Ibrance	Taken with food	No FE	1.2	2
2016	Lenvatinib Mesilate	Kisplyx	Taken at about the same time each day, with or without food	No FE	1	2
2016	Trifluridine Tipiracil Hydrochloride	Lonsurf	No instructions with regard to food intake	No FE Negative	1 0.6	3 3
2016	Sacubitril Valsartan	Neparvis	Taken with or without food	No FE No FE	1 1	4 2
2016	Ixazomib	Ninlaro	Taken at least 1 hour before or at least 2 hours after food	Negative	0.72	3
2016	Pimavanserin	Nuplazid	Taken with or without food	No FE	1.08	
2016	Obeticholic Acid	Ocaliva	Taken with or without food	No FE	1	2
2016	Emtricitabine Rilpivirine Hydrochloride Tenofovir Alafenamide	Odefsey	Taken with food	No FE Positive Positive	0.88 (0.85-0.9)* 1.72 (1.49-1.99)* 1.53 (1.39-1.69)*	1 2 3
2016	Opicapone	Ongentys	Should not eat food for 1 hour before and for at least 1 hour after intake.	Negative	0.69	2
2016	Saxagliptin Dapagliflozin Propanediol Monohydrate	Qtern	Taken with or without food	Positive No FE	1.27 1	3 3
2016	Dronabinol	Syndros	Administer the first dose on an empty stomach at least 30 minutes before eating. Subsequent doses can be taken without regard to meals.	Positive	2.5	2
2016	Osimertinib Mesylate	Tagrisso	Taken with or without food	No FE	1.06	3
2016	Eluxadoline	Truberzi	Taken with food	Negative	0.4	3
2016	Selexipag	Uptravi	Taken with food	No FE	1.1	2
2016	Venetoclax	Venclyxto	Taken with a meal	Positive	MF 3.4 HF 5.1-5.3**	4
2016	Elbasvir Grazoprevir	Zepatier	Taken with or without food	No FE Positive	0.89 1.5	2 2

Table 2-2: Results of Statistical Analysis comparing Positive, Negative and No FE groups for drugs licensed from 2016-2020

Drug Property	Descriptors	Negative	No FE	Positive	Statistical Tests	Neg vs No FE	Neg vs Pos	No FE vs Pos
S+logP	*n*	17	80	44	Levene's Test	0.08	0.15	0.81
	Median	2.09	2.80	3.61	Bootstrap/t-test	0.018[B]	0.002[B]	0.021[t]
	Mean	1.50	2.90	3.67	Mean Difference	-1.41	-2.17	-0.76
	SD of Mean	2.29	1.72	1.77	95% Confidence	(L) -2.56	(L) -3.41	(L) 1.41
	Q1, Q3,	-0.45, 3.33	2.02, 3.90	2.71, 4.80	Interval	(U) -0.34	(U) -0.98	(U) -0.11
	Min, Max	-2.43, 4.21	-1.38, 7.49	-1.34, 7.25				
	Variance	5.26	2.97	3.13				
Hydrogen	*n*	17	80	44	Levene's Test	0.075	0.26	0.34
Bond	*Median*	4	2	2	Bootstrap	0.00	0.00	0.86
Donors	*Mean*	3.53	1.90	1.86	Mean Difference	1.63	1.67	0.04
	SD of Mean	1.55	1.09	1.15	95% Confidence	(L) 0.85	(L) 0.86	(L) -0.36
	Q1, Q3	2.5, 4.5	1, 3	1, 3	Interval	(U) 2.39	(U) 2.48	(U) 0.46
	Min, Max	0, 6	0, 4	0, 4				
	Variance	2.39	1.18	1.33				
Hydrogen	*n*	17	80	44	*Levene's Test*	*0.60*	*0.42*	*0.04*
Bond	*Median*	8	6	7	*Bootstrap*	*0.06*	*0.67*	*0.09*
Acceptors	*Mean*	7.71	6.28	7.30	*Mean Difference*	*1.43*	*0.41*	*-1.02*
	SD of Mean	2.69	2.57	3.49	*95% Confidence*	*(L) -0.04*	*(L) -1.47*	*(L) -2.22*
	Q1, Q3	5.5, 10	5, 7.75	5, 9	*Interval*	*(U) 2.91*	*(U) 2.30*	*(U) 0.18*
	Min, Max	4, 13	1, 15	2, 16				
	Variance	7.221	6.61	12.17				
logD$_{7.4}$	*n*	17	80	44	Levene's Test	0.05	0.16	0.73
	Median	1.12	2.18	2.93	Bootstrap/t-test	0.09[B]	0.003[B]	0.003[t]
	Mean	0.82	1.98	3.01	Mean Difference	-1.16	-2.19	-1.03
	SD of Mean	2.65	1.78	1.93	95% Confidence	(L) -2.65	(L) -3.66	(L) -1.71
	Q1, Q3	-0.57, 2.93	0.99, 3.13	1.75, 4,17	Interval	(U) 0.22	(U) -0.80	(U) -0.35
	Min, Max	-5.18, 4.21	-2.87, 6.64	-2.26, 7.25				
	Variance	7.07	3.17	3.71				
Polar Surface	*n*	17	80	44	Levene's Test	0.16	0.87	0.05
Area	*Median*	107.41	90.37	100.01	Bootstrap	0.00	0.07	0.047
	Mean	125.09	89.12	104.26	Mean Difference	35.92	20.77	-15.15
	SD of Mean	35.87	33.26	43.93	95% Confidence	(L) 18.03	(L) -1.06	(L) -30.40
	Q1, Q3	92.88, 157.65	70.47, 108.48	77.36, 132.71	Interval	(U) 54.07	(U) 42.94	(U) 0.19
	Min, Max	76.10, 178.36	22.00, 188.80	29.46, 199.58				
	Variance	1286.99	1105.95	1930.20				
Dose (mg)	*n*	17	80	44	Levene's Test	0.25	0.009	0.258
	Median	80	50.00	200.00	Bootstrap	0.64	0.001	0.005
	Mean	105.36	121.15	242.57	Mean Difference	-15.79	-137.20	-121.41
	SD of Mean	96.00	205.71	193.88	95% Confidence	(L) -79.20	(L) -212.56	(L) -195.66
	Q1, Q3	25.00, 175.00	10.00, 143.75	100.00, 375.00	Interval	(U) 48.38	(U) -62.23	(U) -44.74
	Min, Max	1.00, 300.00	0.20, 1340.00	2.00, 800.00				
	Variance	9216.31	42314.80	37588.93				
S+Sw	*n*	17	80	44	Levene's Test	0.18	0.18	0.86

Drug Property	Descriptors	Negative	No FE	Positive	Statistical Tests	Neg vs No FE	Neg vs Pos	No FE vs Pos
	Median	-0.49	-0.99	-1.54	Bootstrap/t-test	0.13B	0.003B	0.002t
	Mean	-0.38	-0.92	-1.56	Mean Difference	0.54	1.18	0.64
	SD	1.39	1.07	1.05	95% Confidence	(L) -0.10	(L) 0.51	(L) 0.24
	Q1, Q3	-1.49, 0.22	-1.55, -0.31	-2.35, -0.76	Interval	(U) 1.25	(U) 1.91	(U) 1.03
	Min, Max	-2.05, 2.98	-3.61, 2.02	-4.42, 1.01				
	Variance	1.94	1.14	1.097				
MAD	*n*	17	80	44	Levene's Test	0.07	0.07	0.78
	Median	2.46	2.78	1.02	t-test	0.96	0.06	0.001
	Mean	2.77	2.75	2.11	Mean Difference	0.15	0.67	0.64
	SD	1.54	1.06	1.02	95% Confidence	(L) -0.60	(L) -0.02	(L) 0.25
	Q1, Q3	1.57, 3.54	2.15, 3.38	1.26, 2.71	Interval	(U) 0.63	(U) 1.33	(U) 1.03
	Min, Max	0.57, 6.14	0.02, 5.84	-0.13, 4,34				
	Variance	2.36	1.13	1.04				
Molecular	*n*	17	80	44	Levene's Test	0.09	0.86	0.04
Weight	*Median*	457.69	426.31	446.95	Bootstrap	0.47	0.41	0.02
(g/mol)	*Mean*	453.74	419.89	497.65	Mean Difference	33.85	-43.91	-77.78
	SD of Mean	184.44	134.86	180.56	95% Confidence	(L) -59.07	(L) -147.60	(L) -141.53
	Q1, Q3	302.62, 570.61	348.99, 491.83	368.95, 557.87	Interval	(U) 126.99	(U) 60.81	(U) -22.05
	Min, Max	146.24, 804.04	109.13, 889.02	287.22, 1113.21				
	Variance	34018.56	18188,43	32601.00				
Dose/Solubility	*n*	17	80	44	Levene's Test	0.57	0.26	0.28
Ratio	*Median*	2.14	2.47	3.86	t-test	0.28	0.00	0.00
	Mean	2.10	2.47	3.71	Mean Difference	-0.37	-1.61	-1.24
	SD of Mean	1.36	1.23	1.20	95% Confidence	(L) -1.03	(L) -2.32	(L) -1.70
	Q1, Q3	1.08, 3.22	1.62, 3.26	3.32, 4.40	Interval	(U) 0.30	(U) -0.90	(U) -0.79
	Min, Max	-0.89, 4.05	-0.54, 5.83	0.18, 7.02				
	Variance	1.86	1.52	1.43				
Rotatable	*n*	17	80	44	Levene's Test	0.15	0.89	0.02
Bonds	*Median*	6	5	5	Bootstrap	0.26	0.96	0.20
	Mean	6.65	5.50	6.60	Mean Difference	1.15	0.079	-1.07
	SD of Mean	5.26	3.41	4.97	95% Confidence	(L) -0.86	(L) -2.81	(L) -2.79
	Q1, Q3	2.50, 8.50	3.23, 7.75	3.00, 9.00	Interval	(U) 3.15	(U) 2.97	(U) 0.49
	Min, Max	0.00, 22.00	0.00, 15.00	0.00, 24.00				
	Variance	27.62	11.60	24.72				

Figure 2-1: Visual comparison of the sensitivity, precision, specificity and MCC performance metrics calculated for the test set of the optimum SVM and ANN models produced in this study to predict FE classification.



Appendix 3: Chapter 4

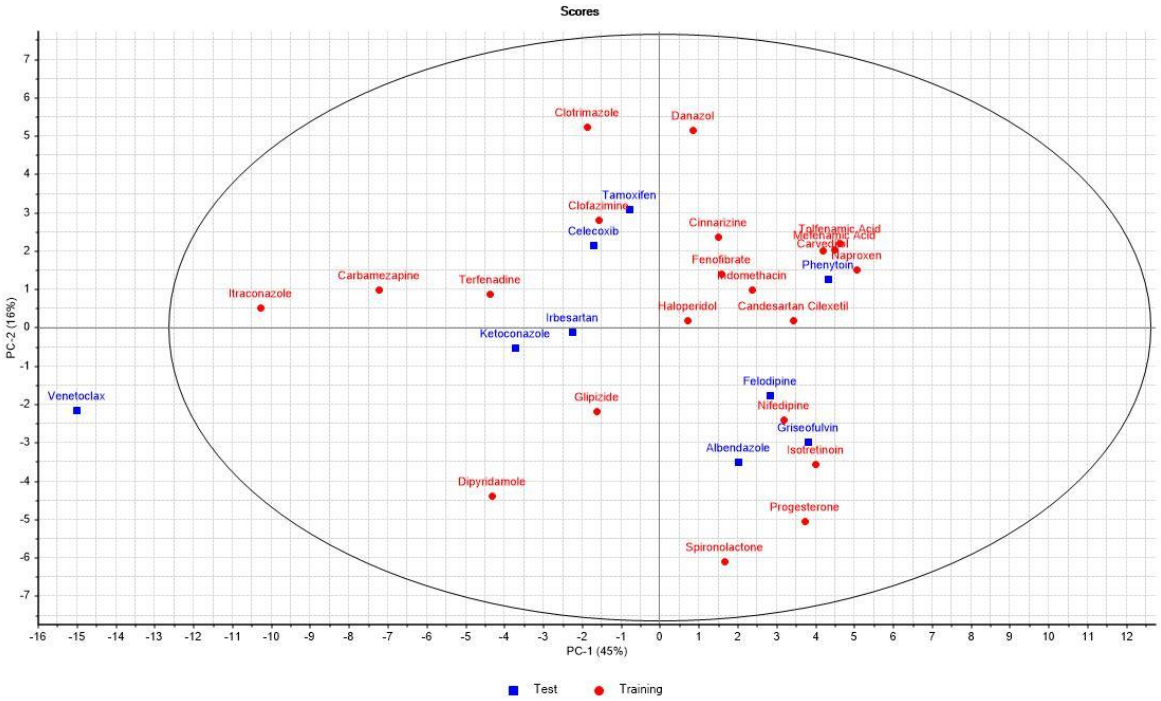
Table 3-1: Apparent solubility values in $PhB_{Ph6.5,}$ FaSSIF, FeSSIF, FaSSIF-$SEDDS_{Migyolo812}$, FaSSIF-$SEDDS_{OliveOil}$, and SR for $SEDDS_{Migylol812}$, $SEDDS_{OliveOil}$, and FeSSIF versus FaSSIF. FE = Food Effect.

Drug Compound	$PhB_{pH6.5}$ ± SD (µg/mL)	FaSSIF ± SD (µg/mL)	FeSSIF ± SD (µg/mL)	$FaSSIF - SEDDS_{Migylol812}$ ± SD (µg/mL)	$FaSSIF - SEDDS_{OliveOil}$ ± SD (µg/mL)	$logSR_{MC}$	$logSR_{LC}$	logFE
Albendazole	0.9 ± 0.4	1.9 ± 0.0	6.1 ± 0.1	9.62 ± 1.28	7.44 ± 1.634	0.704	0.593	0.507
Candesartan Cilexetil	-	8.26	10	138.9 ± 9.6	138.49 ± 4.63	1.227	1.225	0.083
Carbamazepine	227.1 ± 22.9	266.1 ± 31.4	524.1 ± 25.0	388.127 ± 13.46	379.83 ± 16.47	0.164	0.152	0.294
Carvedilol	46	55.9	305.0 ± 2.0	634.098 ± 5.46	442.88 ± 32.49	1.055	0.898	0.737
Celecoxib	1.54234 ±0.51	34.09 ± 5.12	226	579.82 ± 33.83	240.02 ± 16.07	1.230	0.848	0.823
Cinnarizine	1.4	13.4	112 ± 2.0	228.9 ± 10.84	194.05 ± 9.45	1.232	1.161	0.922
Clofazimine	-	6.2	29.6	57.8 ± 1.68	47.12 ± 4.24	0.969	0.881	0.679
Clotrimazole	2.3 ± 0.3	3.5 ± 0.4	71.1 ± 6.0	225.43 ± 17.8	208.95 ± 11.27	1.809	1.776	1.311
Danazol	0.3 ± 0.05	9.6729 ±1.89	28.8 ± 0.4	59.09 ± 1.44	41.61 ± 3.79	0.786	0.633	0.473
Dipyridamole	6.35	11.56	137.2 ± 6.2	42.90 ± 1.19	37.873 ± 0.907	0.568	0.516	1.074
Felodipine	1.187	54.278	237.0 ± 1.0	337.47 ± 29.1	245.36 ± 14.63	0.794	0.655	0.640
Fenofibrate	0.3 ± 0.0	9.6 ±1.4	40.4 ± 2.9	482.39 ± 47.175	286.55 ± 26.29	1.706	1.480	0.624
Glipizide	22.5 ± 0.6	31.3 ± 3.3	4.3 ± 0.2	35.47 ± 0.62	32.523 ± 1.27	0.054	0.017	-0.862
Griseofulvin	15	20 ±0.9	29.2 ± 3.4	37.35 ± 0.8675	34.15 ± 0.57	0.272	0.233	0.164
Haloperidol	77.81	110.51	120.9 ± 7.3	347.44 ± 25.48	243.65 ± 14.83	0.497	0.343	0.039
Indomethacin	219.0 ± 78.0	443.0 ± 10.0	109.0 ± 7.0	811.48 ± 9.1	794.21 ± 4.77	0.263	0.253	-0.609
Irbesartan	102.0 ± 4.0	112.0 ± 3.4	261	306.29 ± 31.19	277.5 ± 37.22	0.436	0.395	0.367
Isotretinoin	-	52.21	321	188.3 ±6 .90	155.53 ± 6.30	0.556	0.474	0.789
Itraconazole	-	0.33	0.7	4.763 ± 0.16	2.89 ± 0.5	1.159	0.943	0.327
Ketoconazole	6.5	25.91 ± 0.70	403.3 ± 16.5	109.32 ± 5.65	98.21 ± 7.43	0.625	0.579	1.192
Mefenamic acid	-	60	649	212.46 ± 13.292	198.6 ± 13.19	0.549	0.520	1.034
Naproxen	230.26	492.29	401	2356.17 ± 95.78	2255.05 ± 75.76	0.672	0.661	-0.089
Nifedipine	11.5	27.8	46.1 ± 1.0	124.78 ± 9.93	101.72 ± 9.65	0.652	0.564	0.220
Phenytoin	39.07	42.84	283	61.51 ± 7.24	56.29 ± 5.58	0.158	0.117	0.820
Progesterone	11.16	25.56	78.6 ± 16.2	89.34 ± 5.98	62.72 ± 1.21	0.544	0.389	0.488
Spironolactone	22	25.8	46.0 ± 2.5	61.9 ±7.08	47.8 ± 6.05	0.380	0.267	0.251
Tamoxifen	5.9	156	236.0 ± 13.0	1081.47 ± 56.36	882.21 ± 20.79	0.839	0.753	0.180
Terfenadine	13.6 ± 1.3	89.0 ± 4.0	256	371.23 ± 15.46	329.95 ± 4.28	0.620	0.569	0.459
Tolfenamic acid	27.404	62.779	41.0 ± 0.5	311.56 ± 16.03	224.37 ± 7.13	0.696	0.553	-0.185
Venetoclax	0.04	20.729 ±0.51	28.4 ± 2.2	246.340 ± 25.75	138.83 ± 16.37	1.075	0.825	0.137

Table 3-2: RP-HPLC/UV methods for the 6 drugs completed using the Shake Flask Method with HPLC-UV analysis.

Drug	Column	A	B	Ratio	Temp (°C)	Flow Rate (mL/min)	Inj. Vol (µL)	λ (nm)
Danazol	Symmetry C18 5 µm, 4,6 x 150 mm	ACN	Water	55:45	25	1	50	286
Ketoconazole	Symmetry C18 5 µm, 4,6 x 150 mm	Phosphate buffer 10 mM, pH 8.5	ACN	40:60	25	0.8	50	297
Venetoclax	Zorbax Eclipse Plus-C18 column (5 µm, 4.6 mm x 150 mm) including Zorbax 156 Eclipse Plus-C18 guard column (5 µm, 4.6 mm x 12.5 mm)	ACN + 0.5 % TFA	Water + 0.5 % TFA	53:47	40	1	50	316
Fenofibrate	Symmetry C18 5 µm, 4,6 x 150 mm	NaAc 25 mM, pH 5.0	ACN	20:80	25	1	50	287
Celecoxib	Symmetry C18 5 µm, 4,6 x 150 mm	ACN + 0,15%TEA, pH3	Water	55:45	25	1	20	254
Griseofulvin	Symmetry C18 5 µm, 4,6 x 150 mm	ACN	Water	55:45	25	1	50	292

Figure 3-1: Principal component analysis (PCA) scores plot detailing the chemical space occupied by the Training and Test Sets of the dataset. Training set is shown in red and test set is shown in blue.



61% of the variation in the dataset is explained by PC-1 and PC-2 and Venetoclax is outside the 95% confidence level and was therefore placed in the test set.

Table 3-3: Preliminary studies testing the two solubility methods employed

Drug Solubility in FaSSIF-SEDDS$_{Migylol812}$ and FaSSIF-SEDDS$_{OliveOil}$ completed for both shake flask and µDISS methods using Danazol. Solubilities were obtained using FaSSIF-V2 for the µDISS method, which contains a smaller concentration of lecithin, due to powder availability at that time, therefore a ratio of solubility in MC/LC was calculated to test similarity of results instead of direct comparisons.

	Shake Flask	**uDiss**	**Ratio MC/LC Solubility**
FaSSIF-SEDDS$_{Migylol812}$	59.089 µg/mL (±1.44)	37.130 µg/mL (±0.589)	Shake Flask = 1.42
FaSSIF-SEDDS$_{OliveOil}$	41.612 µg/mL (± 3.79).	22.878 µg/mL (±1.138)	µDISS = 1.6

Appendix 4: Chapter 5

Table 4-1. Equilibrium solubility values (µg/mL) and aDS for the dataset of 21 drugs using Capmul MCM and Maisine CC. AT refers to Ambient Temperature. Data in brackets refer to standard deviation (SD) of solubility value or standard error (SE) of aDS ratios. * denotes data obtained from a previous publication.

Drug Compound	AT Solubility Capmul MCM	AT Solubility Maisine CC	Solubility 60ºC $sLBF_{Campul}^{MC}$	Solubility 60ºC $sLBF_{Maisine}^{LC}$	aDS $sLBF_{Capmul}^{MC}$	aDS $sLBF_{Maisine}^{LC}$
Carvedilol	45.07 (2.94)	10.28 (0.70)	127.90 (2.77)	28.91 (1.79)	2.84 (0.11)	2.81 (0.15)
Celecoxib	49.78 (5.61)*	13.30 (1.11)*	88.10 (6.67)	36.16 (2.83)	1.77 (0.14)	2.72 (0.18)
Cinnarizine	35.97 (1.03)*	29.27 (1.01)*	81.67 (3.98)	86.64 (6.54)	2.27 (0.07)	2.96 (0.14)
Clotrimazole	192.06 (5.91)	91.05 (3.73)	311.16 (13.89)	170.82 (5.27)	1.62 (0.05)	1.88 (0.06)
Danazol	16.32 (4.52)	11.19 (0.64)	42.91 (1.85)	22.41(0.95)	2.63 (0.43)	2.00 (0.08)
Dipyridamole	10.17 (0.98)	1.38 (0.17)	32.22 (3.12)	4.71 (0.61)	3.17 (0.25)	3.42 (0.36)
Felodipine	74.41 (4.63)	36.79 (1.04)	125.73 (12.40)	49.71 (6.19)	1.69 (0.11)	1.35 (0.10)
Fenofibrate	76.95 (6.66)	50.11 (1.38)	136.98 (11.10)	144.38 (8.82)	1.78 (0.12)	2.88 (0.11)
Fenofibric Acid	11.81 (4.77)	7.80 (0.26)	36.28 (1.55)	18.45 (0.56)	3.07 (0.72)	2.37 (0.06)
Griseofulvin	4.70 (0.51)	1.92 (0.01)	12.55 (0.52)	3.68 (0.26)	2.67 (0.18)	1.91 (0.08)
Haloperidol	31.37 (1.34)	7.63 (0.94)	46.03 (1.43)	14.22 (0.45)	1.47 (0.04)	1.86 (0.14)
Ibuprofen	237.34 (14.85)	128.67 (4.05)	616.10 (103.62)	436.41 (46.36)	2.60 (0.27)	3.39 (0.22)
Indometacine	21.31 (0.70)	7.84 (0.30)	35.53 (1.55)	14.45 (0.25)	1.67 (0.05)	1.84 (0.04)
Itraconazole	1.99 (0.02)	0.53 (0.01)	5.12 (0.49)	1.16 (0.03)	2.57 (0.14)	2.18 (0.03)
JNJ-2a	283.07 (15.58)*	47.50 (1.24)	293.35 (27.73)	76.48 (2.92)	1.04 (0.07)	1.61 (0.04)
Ketoconazole	104.92 (3.32)	29.62 (0.25)	221.22 (2.98)	66.14 (5.06)	2.11 (0.04)	2.23 (0.10)
Naproxen	39.21 (4.83)	17.27 (1.98)	50.73 (6.04)	25.90 (0.45)	1.29 (0.13)	1.50 (0.10)
Niclosamide	8.73 (0.55)	3.21 (0.25)	14.96 (0.96)	5.44 (0.22)	1.71 (0.09)	1.70 (0.09)
Progesterone	97.92 (9.08)	49.49 (0.84)	163.96 (19.86)	112.75 (6.06)	1.67 (0.15)	2.28 (0.07)
Sulfalazine	3.79 (0.08)	0.20 (0.03)	4.89 (0.13)	0.32 (0.03)	1.29 (0.03)	1.62 (0.17)
Venetoclax	2.45 (0.41)	2.44 (0.16)	6.43 (0.28)	3.62 (0.84)	2.62 (0.12)	1.48 (0.21)

Table 4-2. RP-HPLC/UV Methods utilised in this study. ACN refers to Acetonitrile, MeOH refers to Methanol, TFA refers to Trifluoroacetic Acid, H_2O refers to Water and NaAC refers to Sodium Acetate. * denotes the fact that the mobile phase was adjusted to pH 2.5 with 5 %v/v Orthophosphoric acid.

Drug	Column	A	B	Ratio	Temp (°C)	Flow Rate (ml/min)	Inj. Vol (µL)	λ (nm)
Danazol	Symmetry C18 5 µm, 4,6 x 150 mm	ACN	H_2O	55:45	25	1	50	286
Ketoconazole	Symmetry C18 5 µm, 4,6 x 150 mm	Phosphate buffer 10 mM, pH 8.5	ACN	40:60	25	0.8	50	297
Venetoclax	Zorbax Eclipse Plus-C18 column (5 µm, 4.6 mm x 150 mm) including Zorbax 156 Eclipse Plus-C18 guard column (5 µm, 4.6 mm x 12.5 mm)	ACN + 0.5 % TFA	H_20 + 0.5 % TFA	53:47	40	1	50	316
Carvedilol	Symmetry C18 5 µm, 4,6 x 150 mm	ACN	NaAc 25 mM, pH 5.0	60:40	25	1	50	280
Clotrimazole	Symmetry C18 5 µm, 4,6 x 150 mm	MeOH	ACN	95:5	25	1	50	255
Griseofulvin	Symmetry C18 5 µm, 4,6 x 150 mm	ACN	H_2O	55:45	25	1	50	292
Dipyridamole	Gemini 5µ C18 4,6 x 250 mm	ACN	H_2O	60:40	40	1	20	282
Felodipine	Gemini 5µ C18 4,6 x 250 mm	ACN	H_2O	70:30	25	1	20	360
Haloperidol	Symmetry C18 5 µm, 4,6 x 150 mm	ACN:MeOH	NaAc 25 mM, pH 5.0	65:35	25	0.8	50	247
Naproxen	Symmetry C18 5 µm, 4,6 x 150 mm	ACN + 0.1% TFA	H_2O + 0.1% TFA	60:40	25	1	50	254
Niclosamide	Symmetry C18 5 µm, 4,6 x 150 mm	95:5:0.1 ACN:H_2O:TFA	5:95:0.1 ACN:H_2O:TFA	75:25	25	1	50	332
Progesterone	Symmetry C18 5 µm, 4,6 x 150 mm	ACN	NaAc 25 mM, pH 5.5	85:15	25	1	50	254
Sulfalazine	Symmetry C18 5 µm, 4,6 x 150 mm	ACN:H_2O:TFA	H_2O:ACN:TFA	60:40	25	1	50	357
Fenofibric Acid	Symmetry C18 5 µm, 4,6 x 150 mm	ACN*	H_20*	70:30	25	1	20	286

Table 4-3. Equilibrium solubility and aDS_{2h} in $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$ and the aDS ratio difference from average aDS and average aDS_{2h} used to investigate the short-term stability of the sLBF after cooling at AT.

	Solubility 60ºC (+2h) Capmul MCM	**Solubility 60ºC (+2h) Maisine CC**	**aDS_{2h} $sLBF_{Capmul}^{MC}$**	**aDS_{2h} $sLBF_{Maisine}^{LC}$**	**aDS Ratio Unit Change**	**aDS Ratio Unit Change**
Carvedilol	116.44 (6.57)	27.75 (1.02)	2.58 (0.13)	2.70 (0.12)	0.25	0.11
Celecoxib	101.8 (0.0)	39.6 (0.0)	2.04 (0.13)	2.97 (0.14)	0.27	0.25
Cinnarizine	100.1 (0.0)	97.5 (0.0)	2.78 (0.05)	3.33 (0.07)	0.51	0.34
Clotrimazole	308.30 (1.28)	172.73 (0.90)	1.61 (0.03)	1.74 (0.04)	0.01	0.13
Danazol	39.99 (3.54)	22.57 (1.88)	2.45 (0.41)	2.02 (0.12)	0.18	0.01
Dipyridamole	29.55 (4.25)	4.53 (0.49)	2.91 (0.29)	3.29 (0.31)	0.26	0.13
Felodipine	125.19 (5.94)	45.89 (9.05)	1.68 (0.08)	1.25 (0.14)	0.01	0.10
Fenofibric Acid	32.74 (3.17)	17.86 (0.93)	2.77 (0.66)	2.29 (0.08)	0.30	0.08
Griseofulvin	12.21 (0.15)	3.70 (0.24)	2.59 (0.16)	1.92 (0.07)	0.07	0.01
Haloperidol	45.37 (0.24)	13.47 (0.79)	1.45 (0.04)	1.76 (0.14)	0.02	0.10
JNJ-2A	259.5 (0.0)	65.8 (0.0)	0.92 (0.03)	1.38 (0.02)	0.12	0.23
Ketoconazole	224.87 (11.29)	63.82 (3.81)	2.14 (0.07)	2.15 (0.08)	0.03	0.08
Naproxen	45.42 (7.01)	20.39 (2.82)	1.16 (0.13)	1.18 (0.12)	0.14	0.32
Niclosamide	14.06 (1.02)	5.13 (0.23)	1.61 (0.09)	1.60 (0.08)	0.10	0.10
Progesterone	133.30 (16.12)	99.09 (7.27)	1.36 (0.12)	2.00 (0.09)	0.31	0.28
Sulfalazine	4.69 (0.23)	0.30 (0.03)	1.24 (0.04)	1.49 (0.16)	0.05	0.13
Venetoclax	7.13 (1.29)	2.89 (0.52)	2.91 (0.32)	1.18 (0.13)	0.29	0.30

Figure 4-1. Unabridged Abbreviations of Independent Input Variables used in the final PLS and ANN models. Extra information about the descriptors used can be found at this reference (316).

Variables in PLS model predicting aDS $sLBF_{Capmul}^{MC}$:

- **VMcGowan:** Mc Gowan's characteristic volume (317).
- **N_Hydrogn:** Number of hydrogens**.**
- **SHCH_321:** Atom-type hydrogen E-state index for -CH3, -CH2- and >CH- groups (saturated aliphatic carbon) (256).
- **EEM_Afc:** Sum of absolute values of scaled sigma Fukui indices on C (281, 282).
- **EEM_AFnp:** Sum of absolute values of scaled sigma Fukui indices on nonpolar atoms.
- **EEM_NFc:** Minimum scaled sigma Fukui index on C.
- **SHaaCH:** Atom-type hydrogen E-state index for aCHa groups (aromatic carbons).
- **Pi_FMi4:** Fourth component of the autocorrelation vector of scaled pi Fukui- indices (electrophilic).

Variables in PLS model predicting aDS $sLBF_{Maisine}^{LC}$:

- **EEM_NFc**: Minimum scaled sigma Fukui index on C.
- **EEM_NFnp**: Minimum scaled sigma Fukui index on nonpolar atoms.
- **HIVI-TC:** pIC50 in log(mol/L) for inhibition of the HIV-1 Integrase 3'-processing.
- **N_FrRotB**: Number of freely rotatable bonds.
- **NPA_Q2**: Second component of the autocorrelation vector of NPA partial atomic charges.
- **Pi_AQo**: Sum of absolute values of pi partial atomic charges on O.
- **Pi_AQc:** Sum of absolute values of pi partial atomic charges on C.
- **Pi_FPI3:** Third component of the autocorrelation vector of scaled pi Fukui+ indices (nucleophilic)
- **Pi_FMi6:** Sixth component of the autocorrelation vector of scaled pi Fukui- indices (electrophilic).

Variables in MLP-ANN predicting aDS $sLBF_{Capmul}^{MC}$:

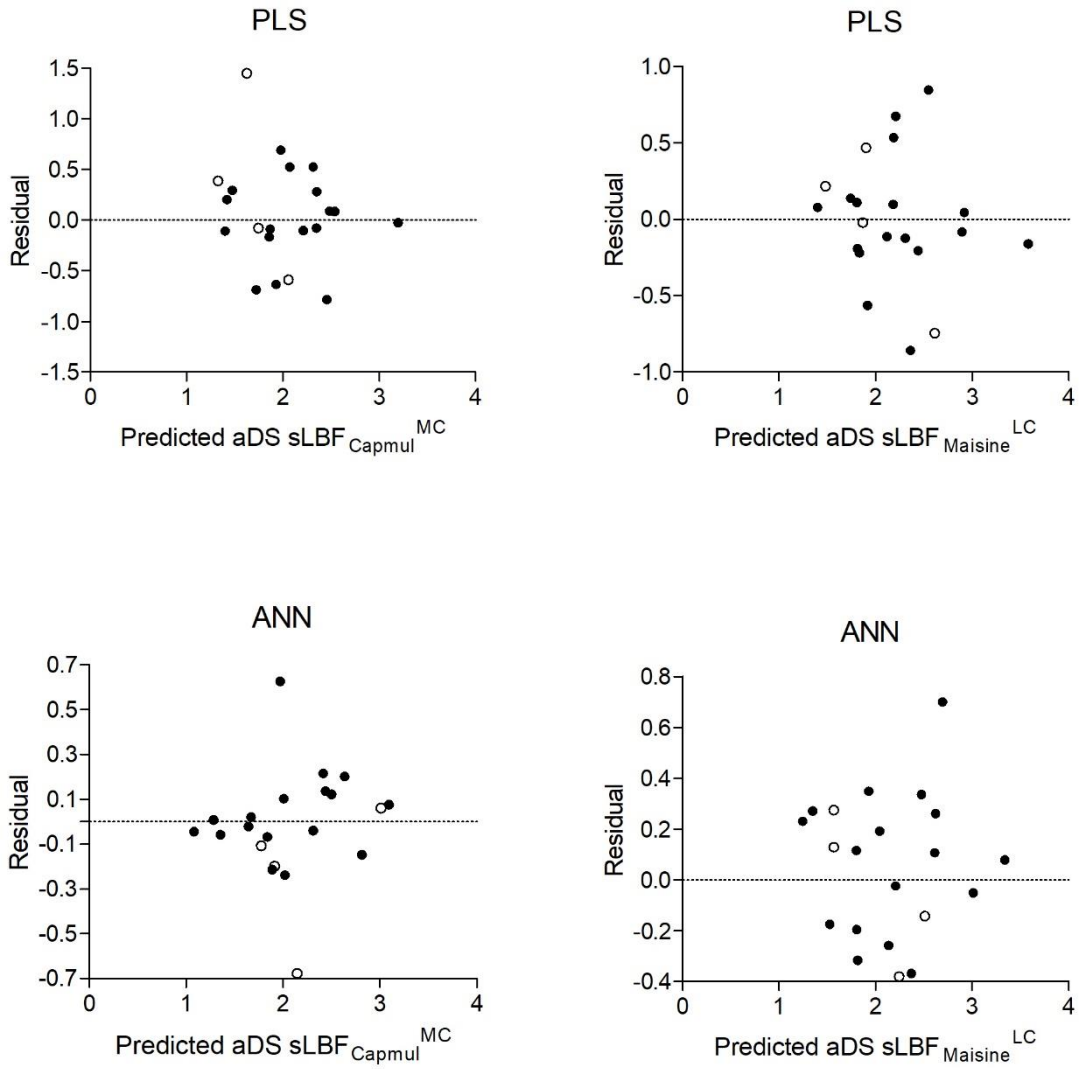
- **Pi_FPl5:** Fifth component of the autocorrelation vector of pi Fukui(+) indices.
- **SolFactor:** Universal salt solubility factor based on S+Sw model.
- **N_CYPAtoms:** Number of potential atoms that can be oxidized by CYP P450 enzymes.
- **EEM_F4:** Fourth component of the autocorrelation vector of sigma Fukui indices.
- **Pi_FPl3:** Third component of the autocorrelation vector of pi Fukui(+) indices.
- **NPA_Q6:** Sixth component of the autocorrelation vector of estimated NPA partial atomic charges (284, 285).
- **MlogP:** Moriguchi estimation of log P (318).

- **MolVol:** Liquid molal volume (cm^3/mol) at the normal boiling point is based on Schroeder's method.
- **NPA_Q1**: First component of the autocorrelation vector of estimated NPA partial atomic charges.
- **S+S_Intrinsic**: Intrinsic water solubility (mg/mL) based on S+Sw and S+pH_Satd models.
- **EqualEta:** Equalized molecular hardness.
- $\Delta \mathbf{H}_{fus}$: - Enthalpy of fusion in kJ/mol obtained from literature.
- **M_CX**: Summation of numbers of carbon and halogen atoms weighted by C:1.0, F:0.5, Cl:1.0, Br:1.5, and I:2.0.
- **Pi_MinQ**: Minimum Hückel pi atomic charge.
- **N_Electr**. Total number of electrons in a molecule.

Variables in MLP-ANN predicting aDS $sLBF_{Maisine}^{LC}$

- **N_Bonds**: Number of bonds.
- **Pi_FPl1:** First component of the autocorrelation vector of pi Fukui(+) indices.
- **T_Rads**: topological equivalent of Rads_3D.
- **MaxQ:** Maximal PEOE Partial Atomic Charge.
- **N_Atoms**: Number of atoms.
- **Pi_FMi1**: First component of the autocorrelation vector of pi Fukui(-) indices.
- **HBDch**: Sum of Estimated NPA Partial Atomic Charges on HB Donor Hydrogens.
- **F_AromB**: Aromatic bonds as fraction of total bonds.
- **NPA_Q2**: Second component of the autocorrelation vector of estimated NPA partial atomic charges.
- **SsssCH**: Atom-type E-state index for >CH- groups.
- **NPA_Q5**: Fifth component of the autocorrelation vector of estimated NPA partial atomic charges.

Figure 4-2. Predicted by residual plots for $sLBF_{Capmul}^{MC}$ and $sLBF_{Maisine}^{LC}$ using PLS and ANN modelling



Appendix 4-6: Pilot study to investigate if PLS-DA could classify drugs according to an aDS cut-off of 2 using glass-forming ability (GFA) or selected GFA related solid-state properties.

Introduction

This study design explored if PLS-DA could classify drugs according to an aDS cut-off of 2 using glass-forming ability (GFA) or selected GFA related solid-state properties. This facilitated investigation if previously published correlations between GFA class (a drugs propensity to vitrify on cooling), and aDS ratios (275, 319) are also observed for sLBFs. A high or low aDS value was arbitrarily defined by a critical value of 2, as a suitable indicator of whether a sLBF approach may be a viable formulation option.

Methods

Partial Least Squares Discriminant Analysis (PLS-DA)

To investigate if GFA class or related properties could be correlated to aDS, PLS-DA was applied using Unscrambler XI (Camo Analytics, US). Two aDS groups of >2 or <2 were obtained, forming the response variable. GFA classifications, widely known drug physicochemical properties previously related to GFA or used in previous *in silico* models for LBFs were included as descriptor variables (46, 124, 235, 320). Variables were mean centred and standardised through scaling by standard deviation. Included variables were Molecular Weight (MW), Acidic (A), Basic (B), or Neutral (N) drugs, Partition Coefficient (logP), GFA Class 1 (GI) ,Class 2 (GII) or Class 3 (GIII), T_m, T_g, ΔH_{fus}, ΔS_{fus}, Hydrogen Bond Acceptors (HBA), Hydrogen Bond Donors (HBD), Polar Surface Area (PSA), T_m/T_g, T_{rg}, Number of Bonds (N_B), Number of Freely Rotatable Bonds (N_Ro), Fraction of Single Bonds (F_SB),

Fraction of Double Bonds (F_DB), Fraction of Aromatic Bonds (F_AB), Atom-Type Cumulative Electrotopological State (E-state) Index for Methylene Carbons (S-C), Partition Coefficient at pH 6.5 (logD) and Number of Aliphatic Rings (N_AR). A Martens' uncertainty Test identified the important variables (250), involving a "jackknifing" procedure and production of sub-models to identify non-significant variables.

Results

Classification of Drugs by aDS using PLS-DA

The possibility to relate aDS classifications and GFA-associated properties using PLS-DA was tested. While MVA is commonly used for quantitative prediction, classification predictions facilitate broad analysis of property trends. We defined an aDS cut-off of 2 as an arbitrary value for suitable supersaturation and hence as indicator of whether a sLBF approach may be a viable formulation option. The aim was to achieve class separation in the score plots (Figure 4-6-1). Separation was evident for both sLBFs as two clusters were observed in blue (<2) and red (>2). However, Haloperidol (No. 11) and Sulfasalazine (No. 20) likely exhibited drug specific properties, resulting in poor separation from the scores plot area predominated by >2 aDS (Figure 4-6-1).

In terms of $sLBF_{Capmul}^{MC}$, 55% of the variability was explained using 2 principal components (PC). From the uncertainty test, MW, N_B, N_Ro and S-C were most important for classification (Figure 4-6-2). Higher values of MW, N_B, N_Ro were observed in general for the >2 group. For $sLBF_{Maisine}^{LC}$, a higher percentage of variability was explained (62%), using 2 PC and ΔH_{fus}, ΔS_{fus} and F_SB were found to be the most important properties. Drugs classified as >2 tended to display a higher ΔS_{fus}, as lower values of ΔS_{fus} were found to the left

of the scores plot. Differences in the most important properties between the two sLBF may reflect the higher number of drugs which displayed a higher aDS (>2) in $sLBF_{Maisine}^{LC}$.

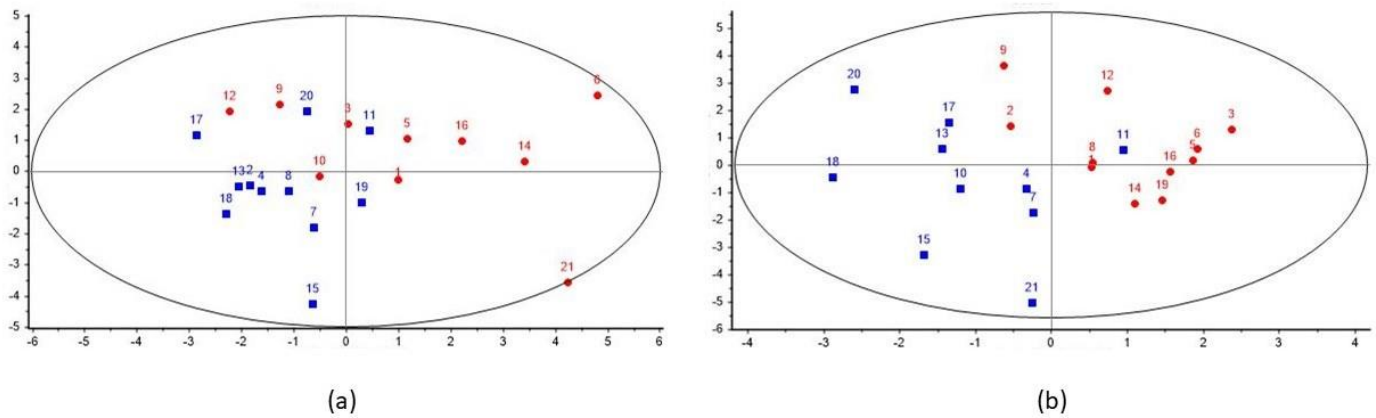


Figure 4-6-1: Scores plots from the PLS-DA classification aDS >2/<2 for $sLBF_{Capmul}^{MC}$ (a) and $sLBF_{Maisine}^{LC}$ (b) using GFA related properties. Drugs (numbered according to their listing in Chapter 5, Table 5-I) are classified according to aDS <2 (blue) or >2 (red) where better separation is achieved for $sLBF_{Maisine}^{LC}$ (b).

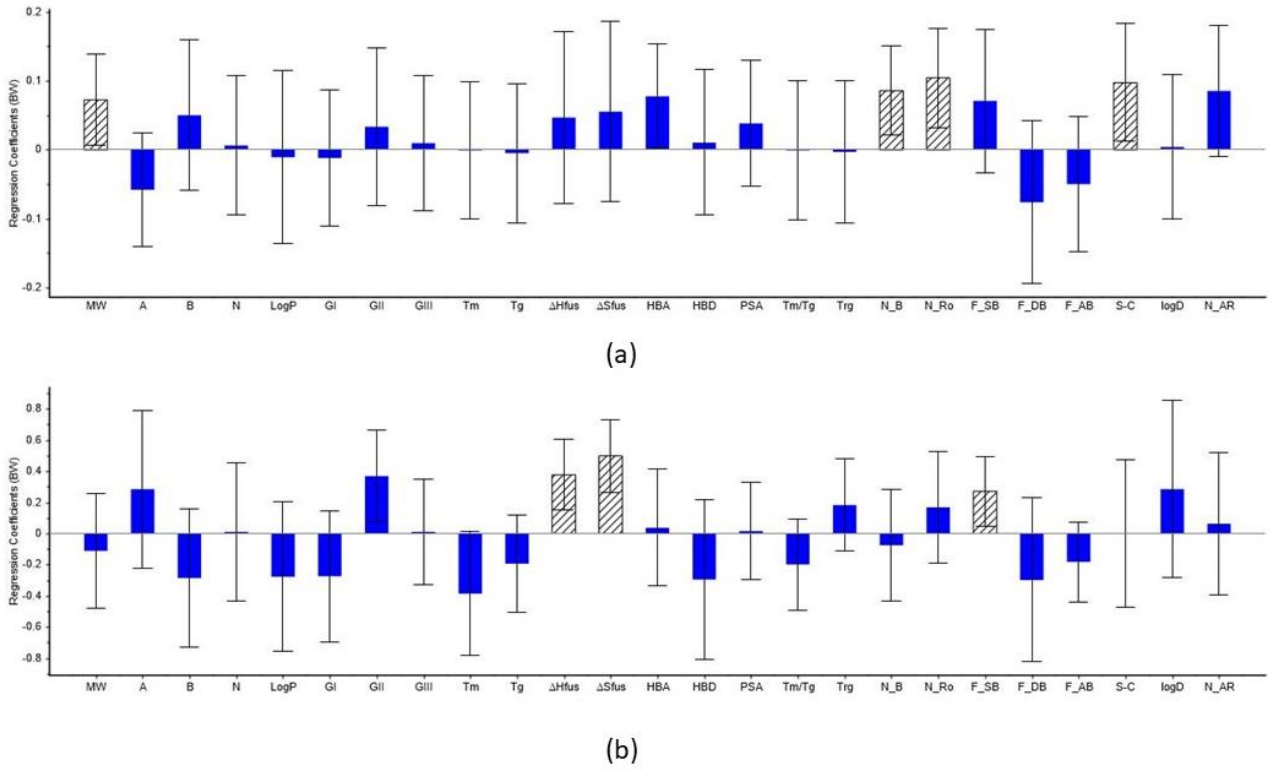


Figure 4-6-2: Important variables chart from the PLS-DA aDS classification of $sLBF_{Capmul}^{MC}$ (a) and $sLBF_{Maisine}^{LC}$ (b) demonstrating the weighted regression coefficients of the input descriptors. The error bars show the 95% confidence interval and the diagonal black lines denote the variables calculated to be the most important for classification according to the Martens' uncertainty test.

Discussion and Conclusion

This work analysed if trends between aDS and GFA-related properties could be elucidated via aDS classification modelling. GFA was previously correlated to solvent shift mediated aDS using PCA, where supersaturation potential of class 3 drugs was higher than class 1 drugs (275). Here, PLS-DA revealed GFA class and related properties better explained variability for $sLBF_{Maisine}^{LC}$ versus $sLBF_{Capmul}^{MC}$ (62% versus 55%). General separation of the aDS groups (>2/<2) was achieved. Molecular weight (MW), number of bonds (N_B), number of freely rotatable bonds (N_RB) and Atom-Type Cumulative Electrotopological State (E-state) Index for Methylene Carbons (S-C) were the most important properties for aDS classification in $sLBF_{Capmul}^{MC}$. Increasing size, incorporating N_B and MW, along with N_RB were previously positively correlated to GFA. There MW > 300g/mol and higher RB count were common characteristics of glass-forming compounds (46, 98, 118, 220, 321). While, MW and N_RB were observed as significant properties for commercial success with LBF (222), S-C was significant in modelling the solubility gain upon SEDDS dispersion (235). For $sLBF_{Maisine}^{LC}$, important classification properties were Entropy of Fusion (ΔS_{fus}), Enthalpy of Fusion (ΔH_{fus}) and Fraction of Single Bonds (F_SB). ΔS_{fus} and ΔH_{fus} were previously linked to GFA, where class 1 drugs displayed higher values compared to class 3 (118, 320). Inclusion of F_SB may indicate decreased saturation or aromaticity may influence aDS. Aromatic ring structures were previously related to high crystallisation tendencies and poor GFA. Therefore, despite lack of evidence of a direct trend between aDS and GFA classes (Chapter 5, Figure 5-2) and presence of only ΔH_{fus} in the quantitative models, properties previously correlated to GFA were somewhat influential in classifying drugs by aDS. However, a larger dataset including an external test set is needed to verify these results.

Abbreviation List

ΔH_{fus}	Enthalpy of Fusion
ΔS_{fus}	Entropy of Fusion
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicology
aDS	Apparent Degree of Supersaturation
AI	Artificial Intelligence
ANN	Artificial Neural Networks
API	Active Pharmaceutical Ingredient
AT	Ambient Temperature
AUC	Area Under the Curve
BioRAM	Biopharmaceutics Risk Assessment Roadmap
BCS	Biopharmaceutics Classification System
BDA	Big Data Analytics
BDDCS	Biopharmaceutical Drug Disposition Classification System
bRo5	Beyond Rule of Five
CBP	Computational Biopharmaceutical Profiling
clogP	Calculated Octanol/Water Partition Coefficient
C_{max}	Peak Plasma Concentration
COSMO-RS	Conductor like Screening Model for Real Solvents
Cryo-TEM	Cryogenic Transmission Electron Microscopy
D/S	Dose Solubility Ratio
DCS	Developability Classification System
DLS	Dynamic Light Scattering
Do	Dose Number
DSC	Differential Scanning Calorimetry
DTs	Decision Trees
EEM_Afc	Sum of Absolute Values of Sigma Fukui Indices on C
EEM_AFnp	Sum of Absolute Values of Sigma Fukui Indices on Nonpolar Atoms
EEM_F2	Second Component of the Autocorrelation Vector of Sigma Fukui Indices
EEM_F4	Fourth Component of the Autocorrelation Vector of Sigma Fukui Indices
EEM_NFc	Minimum Sigma Fukui Index on C
EEM_NFnp	Minimum Sigma Fukui Index on Nonpolar Atoms
EMA	Eurpean Medicines Agency
EPAR	European Public Assessment Report
EqualEta	Equalized Molecular Hardness
eRo5	Extended Rule of Five
F_DB	Fraction of Double Bonds
F_HBP	Population Average Number of Protons Available for Hydrogen Bonding Divided by the Number of Non-Hydrogen Atoms
F_SB	Fraction of Single Bonds

FArom_B,	Aromatic Bonds as a Fraction of Total Bonds
FaSSGF	Fasted State Simulated Gastric Fluids
FaSSIF	Fasted State Simulated Intestinal Fluid
FaSSIFp	Fasted State Simulated Intestinal Fluid of Pigs
FDA	Food and Drug Administration
FE	Food Effect
FeHIF	Fed State Human Intestinal Fluids
FeSSIF	Fed State Simulated Intestinal Fluid
FN	False Negative
FP	False Positive
GFA	Glass Forming Ability
GIT	Gastrointestinal Tract
HBA	Hydrogen Bond Acceptors
HBD	Hydrogen Bond Donors
HBDch	Sum of NPA Partial Atomic Charges on HB Donor Hydrogen Atoms
HIF	Human Intestinal Fluids
HIVI-TC	pIC50 in log(mol/L) for inhibition of the HIV-1 Integrase 3'-processing
HPLC	High Performance Liquid Chromatography
HPRA	Health Products Regulatory Authority
LBF	Lipid-Based Formulations
LCT	Long Chain Monoglycerides
LFCS	Lipid Formulation Classification System
$logD_{7.4}$, $logD_{6.5}$	Partition Coefficient at pH 7.4 or 6.5
logHLC	Logarithm of the Air-Water Partition Coefficient
logP	Octanol/Water Partition Coefficient
logS	Logarithm of Aqueous Solubility
LR	Linear Regression
LSI	Ileum
M&S	Modelling and Simulation
M_CX	Summation of Numbers of Carbon and Halogen Atoms Weighted by C:1.0, F:0.5, Cl:1.0, Br:1.5, and I:2.0.
MAD	Maximum Absorbable Dose
MaxQ	Maximum PEOE Partial Atomic Charge
MCC	Matthew’s Correlation Coefficient
MCT	Medium Chain Triglycerides
MDS	Maximum Dosage Strength
MD Simulations	Molecular Dynamic Simulations
ML	Machine Learning
MlogP	Moriguchi Estimation of logP
MLP	Multilayer Perception
MLR	Multiple Linear Regression
MolVol	Liquid molal volume

MSI	Middle Small Intestine
MW	Molecular Weight
N_AR	Number of Aliphatic Rings
N_Atoms	Number of Heavy Atoms
N_Bonds	Number of Bonds
N_CYPAtoms	Number of potential atoms that can be oxidized by CYPP450 enzymes
N_Electr	Total Number of Electrons in a Molecule
N_Hydrgn	Number of Hydrogens
N_FrRotB	Number of Freely Rotatable Bonds
NME	New Molecular Entity
NPA_Q1	First Component of the Autocorrelation Vector of Estimated NPA Partial Atomic Charges
NPA_Q2	Second Component of the Autocorrelation Vector of Estimated NPA Partial Atomic Charges
NPA_Q5	Fifth Component of the Autocorrelation Vector of Estimated NPA Partial Atomic Charges
NPA_Q6	Sixth Component of the Autocorrelation Vector of Estimated NPA Partial Atomic Charges
OrBiTo	Oral Biopharmaceutics Tools
PBPK	Physiological based Pharmacokinetic Modeling
PC	Principal Component
PCA	Principal Component Analysis
PC-SAFT	Perturbed Chain Statistical Associating Fluid
PDI	Polydispersity Index
pDose	LOG_{10}(Maximum Dose Strength)
$PhB_{pH6.5}$	Phosphate Buffer pH 6.5
PiAQc	Sum of absolute values of pi partial atomic charges on C
PiAQo	Sum of absolute values of pi partial atomic charges on O
Pi_FMi1	First Component of the Autocorrelation Vector of pi Fukui(-) indices
Pi_FMi6	Sixth component of the autocorrelation vector of scaled pi Fukui-indices (electrophilic)
Pi_FPl1	First Component of the Autocorrelation Vector of pi Fukui(+) indices
Pi_FPl3	Third Component of the Autocorrelation Vector of pi Fukui(+) indices
Pi_FPl5	Fifth Component of the Autocorrelation Vector of pi Fukui(+) indices
Pi_MinQ	Minimum Hückel pi Atomic Charge
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
PMML	Predictive Model Markup Language
PSA	Polar Surface Area

PWSD	Poorly Water-Soluble Drug
QSAR	Quantitative Structure Activity Relationships
QTTP	Quality Target Product Profile
R&D	Research and Development
RB	Rotatable Bonds
rDCS	Refined Developability Classification System
RMSE	Root Mean Square Error
Ro5	Lipinski Rule-of-5
S+logD	Simulations Plus model of log D based on S+logP
S+logP	Simulations Plus model of log P
S+S_Intrins	Intrinsic Water Solubility
S+Sw	Aqueous Solubility
SD	Solid Dispersion
SE	Standard Error
SEDDS	Self-Emulsifying Drug Delivery Systems
SGF_{sp}	Simulated Gastric Fluid Without Pepsin
SHaaCH	Atom-type hydrogen E-state index for aCHa groups (aromatic carbons)
SHCH_321	Atom-Type Hydrogen E-State Index for -CH3, -CH2- and >CH Groups
SLAD	Solubility Limited Absorbable Dose
sLBF	Supersaturated Lipid Based Formulation
SolFactor	Universal Salt Solubility Factor
SR_{LC}	Solubility Ratio Long Chain Triglyceride-based SEDDS
SR_{MC}	Solubility Ratio Medium Chain Triglyceride-based SEDDS
SssCH2	Atom-Type Cumulative Electrotopological State (E-State) Index for Methylene Carbons
SsssCH	Atom-Type E-State Index for >CH- Groups
SVM	Support Vector Machine
T_Rads	Topological Version of the Projected Radius of the Molecule Averaged Over All Orientations
Te	Test Set
T_g	Glass Transition Temperature
T_m	Melting Point
T_{max}	Time to Peak Plasma Concentration
TN	True Negative
TP	True Positive
T_PSA	Topological Polar Surface Area
Tr	Training Set
T_{rg}	Reduced Glass Transition Temperature
U%	Percentage Excreted Unchanged in Urine
USI	Upper Small Intestine
VMcGowan	McGowan's Characteristic Volume

References

1. Pina AS, Hussain A, Roque AC. An historical overview of drug discovery. *Methods Mol Biol*. 2009;572:3-12.

2. Bergström CAS, Porter CJH. Understanding the Challenge of Beyond-Rule-of-5 Compounds. *Adv Drug Deliv Rev*. 2016;101:1-5.

3. Bergstrom CAS, Charman WN, Porter CJH. Computational prediction of formulation strategies for beyond-rule-of-5 compounds. *Adv Drug Deliv Rev*. 2016;101:6-21.

4. Kalepu S, Nekkanti V. Insoluble drug delivery strategies: review of recent advances and business prospects. *Acta Pharm Sin B.* 2015;5(5):442-53.

5. Fagerberg JH, Al-Tikriti Y, Ragnarsson G, Bergstrom CA. Ethanol effects on apparent solubility of poorly soluble drugs in simulated intestinal fluid. *Mol Pharm.* 2012;9(7):1942-52.

6. Ditzinger F, Price DJ, Ilie AR, Kohl NJ, Jankovic S, Tsakiridou G, et al. Lipophilicity and hydrophobicity considerations in bio-enabling oral formulations approaches - a PEARRL review. *J Pharm Pharmacol*. 2019;71(4):464-82.

7. Yu LX, Amidon GL, Polli JE, Zhao H, Mehta MU, Conner DP, et al. Biopharmaceutics classification system: the scientific basis for biowaiver extensions. *Pharm Res*. 2002;19(7):921-5.

8. Porter CJ, Trevaskis NL, Charman WN. Lipids and lipid-based formulations: optimizing the oral delivery of lipophilic drugs. *Nat Rev Drug Discov*. 2007;6(3):231-48.

9. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov*. 2012;11(3):191-200.

10. Di L, Fish PV, Mano T. Bridging solubility between drug discovery and development. *Drug Discovery Today*. 2012;17(9):486-95.

11. Savjani KT, Gajjar AK, Savjani JK. Drug solubility: importance and enhancement techniques. *ISRN Pharm*. 2012;2012:195727.

12. Saxena V, Panicucci R, Joshi Y, Garad S. Developability assessment in pharmaceutical industry: An integrated group approach for selecting developable candidates. *J Pharm Sci.* 2009;98(6):1962-79.

13. Butler JM, Dressman JB. The developability classification system: application of biopharmaceutics concepts to formulation development. *J Pharm Sci.* 2010;99(12):4940-54.

14. Feeney OM, Crum MF, McEvoy CL, Trevaskis NL, Williams HD, Pouton CW, et al. 50years of oral lipid-based formulations: Provenance, progress and future perspectives. *Adv Drug Deliv Rev*. 2016;101:167-94.

15. Holm R. Bridging the gaps between academic research and industrial product developments of lipid-based formulations. *Adv Drug Deliv Rev*. 2019;142:118-27.

16. 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.

17. Boyd BJ, Bergström CAS, Vinarov Z, Kuentz M, Brouwers J, Augustijns P, et al. Successful oral delivery of poorly water-soluble drugs both depends on the intraluminal behavior of drugs and of appropriate advanced drug delivery systems. *Eur. J. Pharm. Sci*. 2019;137:104967.

18. Williams HD, Trevaskis NL, Charman SA, Shanker RM, Charman WN, Pouton CW, et al. Strategies to address low drug solubility in discovery and development. *Pharmacol Rev*. 2013;65(1):315-499.

19. Buckley ST, Frank KJ, Fricker G, Brandl M. Biopharmaceutical classification of poorly soluble drugs with respect to "enabling formulations". *Eur J Pharm Sci.* 2013;50(1):8-16.

20. 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-S8.

21. Pouton CW. Formulation of poorly water-soluble drugs for oral administration: physicochemical and physiological issues and the lipid formulation classification system. *Eur J Pharm Sci*. 2006;29(3-4):278-87.

22. Pouton CW, Porter CJH. Formulation of lipid-based delivery systems for oral administration: Materials, methods and strategies. *Adv Drug Deliv Rev*. 2008;60(6):625-37.

23. Mu H, Hoy CE. The digestion of dietary triacylglycerols. *Prog Lipid Res*. 2004;43(2):105-33.

24. Charman WN, Porter CJ, Mithani S, Dressman JB. Physiochemical and physiological mechanisms for the effects of food on drug absorption: the role of lipids and pH. *J Pharm Sci.* 1997;86(3):269-82.

25. O'Driscoll CM, Griffin BT. Biopharmaceutical challenges associated with drugs with low aqueous solubility-the potential impact of lipid-based formulations. *Adv Drug Deliv Rev.* 2008;60(6):617-24.

26. Williams HD, Trevaskis NL, Yeap YY, Anby MU, Pouton CW, Porter CJ. Lipid-based formulations and drug supersaturation: harnessing the unique benefits of the lipid digestion/absorption pathway. *Pharm Res*. 2013;30(12):2976-92.

27. Anby MU, Williams HD, McIntosh M, Benameur H, Edwards GA, Pouton CW, et al. Lipid digestion as a trigger for supersaturation: evaluation of the impact of supersaturation stabilization on the in vitro and in vivo performance of self-emulsifying drug delivery systems. *Mol Pharm*. 2012;9(7):2063-79.

28. Constantinides PP, Wasan KM. Lipid formulation strategies for enhancing intestinal transport and absorption of P-glycoprotein (P-gp) substrate drugs: in vitro/in vivo case studies. *J Pharm Sci*. 2007;96(2):235-48.

29. Patel JP, Brocks DR. The effect of oral lipids and circulating lipoproteins on the metabolism of drugs. *Expert Opin Drug Metab Toxicol*. 2009;5(11):1385-98.

30. O'Driscoll CM. Lipid-based formulations for intestinal lymphatic delivery. *Eur J Pharm Sci.* 2002;15(5):405-15.

31. Charman WNA, Stella VJ. Effects of lipid class and lipid vehicle volume on the intestinal lymphatic transport of DDT. *Int. J. Pharm.* 1986;33(1):165-72.

32. Ilie AR, Griffin BT, Kolakovic R, Vertzoni M, Kuentz M, Holm R. Supersaturated lipid-based drug delivery systems - exploring impact of lipid composition type and drug properties on supersaturability and physical stability. *Drug Dev Ind Pharm*. 2020;46(3):356-64.

33. Koehl NJ, Henze LJ, Kuentz M, Holm R, Griffin BT. Supersaturated Lipid-Based Formulations to Enhance the Oral Bioavailability of Venetoclax. *Pharmaceutics.* 2020;12(6).

34. Thomas N, Holm R, Müllertz A, Rades T. In vitro and in vivo performance of novel supersaturated self-nanoemulsifying drug delivery systems (super-SNEDDS). *J Control Release*. 2012;160(1):25-32.

35. Kuentz M, Holm R, Elder DP. Methodology of oral formulation selection in the pharmaceutical industry. *Eur. J. Pharm. Sci*. 2016;87:136-63.

36. Hauss DJ. Oral lipid-based formulations. *Adv Drug Deliv Rev*. 2007;59(7):667-76.

37. Chiou WL, Riegelman S. Preparation and dissolution characteristics of several fast-release solid dispersions of griseofulvin. *J Pharm Sci.* 1969;58(12):1505-10.

38. Sekiguchi K, Obi N. Studies on Absorption of Eutectic Mixture. I. A Comparison of the Behavior of Eutectic Mixture of Sulfathiazole and that of Ordinary Sulfathiazole in Man. *Chem. Pharm. Bull*. 1961;9(11):866-72.

39. Vo CL-N, Park C, Lee B-J. Current trends and future perspectives of solid dispersions containing poorly water-soluble drugs. *Eur. J. Pharm. Biopharm*. 2013;85(3, Part B):799-813.

40. Vasconcelos T, Sarmento B, Costa P. Solid dispersions as strategy to improve oral bioavailability of poor water-soluble drugs. *Drug Discovery Today*. 2007;12(23):1068-75.

41. Huang Y, Dai W-G. Fundamental aspects of solid dispersion technology for poorly soluble drugs. *Acta Pharmaceutica Sinica B*. 2014;4(1):18-25.

42. Jermain SV, Brough C, Williams RO, 3rd. Amorphous solid dispersions and nanocrystal technologies for poorly water-soluble drug delivery - An update. *Int J Pharm*. 2018;535(1-2):379-92.

43. Baghel S, Cathcart H, O'Reilly NJ. Polymeric Amorphous Solid Dispersions: A Review of Amorphization, Crystallization, Stabilization, Solid-State Characterization, and Aqueous Solubilization of Biopharmaceutical Classification System Class II Drugs. *J Pharm Sci*. 2016;105(9):2527-44.

44. Edueng K, Mahlin D, Bergström CAS. The Need for Restructuring the Disordered Science of Amorphous Drug Formulations. *Pharm Res.* 2017;34(9):1754-72.

45. Jog R, Kumar S, Shen J, Jugade N, Tan DC, Gokhale R, et al. Formulation design and evaluation of amorphous ABT-102 nanoparticles. *Int J Pharm*. 2016;498(1-2):153-69.

46. Baird JA, Van Eerdenbrugh B, Taylor LS. A classification system to assess the crystallization tendency of organic molecules from undercooled melts. *J Pharm Sci*. 2010;99(9):3787-806.

47. Jachowicz R. Dissolution rates of partially water-soluble drugs from solid dispersion systems. I. Prednisolone. *Int. J. Pharm*. 1987;35(1):1-5.

48. Zane P, Gieschen H, Kersten E, Mathias N, Ollier C, Johansson P, 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;142:222-31.

49. Lennernäs H, Aarons L, Augustijns P, Beato S, Bolger M, Box K, et al. Oral biopharmaceutics tools - time for a new initiative - an introduction to the IMI project OrBiTo. *Eur J Pharm Sci*. 2014;57:292-9.

50. Abrahamsson B, McAllister M, Augustijns P, Zane P, Butler J, Holm R, et al. Six years of progress in the oral biopharmaceutics area – A summary from the IMI OrBiTo project. Eur. J. Pharm. Biopharm. 2020;152:236-47.

51. Griffin BT, Kuentz M, Vertzoni M, Kostewicz ES, Fei Y, Faisal W, et al. 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. *Eur J Pharm Biopharm*. 2014;86(3):427-37.

52. O'Dwyer PJ, Box KJ, Koehl NJ, Bennett-Lenane H, Reppas C, Holm R, et al. Novel Biphasic Lipolysis Method To Predict in Vivo Performance of Lipid-Based Formulations. *Mol. Pharm*. 2020;17(9):3342-52.

53. Andreas CJ, Rosenberger J, Butler J, Augustijns P, McAllister M, Abrahamsson B, et al. Introduction to the OrBiTo decision tree to select the most appropriate in vitro methodology for release testing of solid oral dosage forms during development. *Eur J Pharm Biopharm*. 2018;130:207-13.

54. Bergström CA, Holm R, Jørgensen SA, Andersson SB, Artursson P, Beato S, et al. Early pharmaceutical profiling to predict oral drug absorption: current status and unmet needs. *Eur J Pharm Sci*. 2014;57:173-99.

55. Butler J, Hens B, Vertzoni M, Brouwers J, Berben P, Dressman J, et al. In vitro models for the prediction of in vivo performance of oral dosage forms: Recent progress from partnership through the IMI OrBiTo collaboration. *Eur J Pharm Biopharm.* 2019;136:70-83.

56. Dressman JB, Amidon GL, Reppas C, Shah VP. Dissolution testing as a prognostic tool for oral drug absorption: immediate release dosage forms. *Pharm Res.* 1998;15(1):11-22.

57. Markopoulos C, Andreas CJ, Vertzoni M, Dressman J, Reppas C. In-vitro simulation of luminal conditions for evaluation of performance of oral drug products: Choosing the appropriate test media. *Eur J Pharm Biopharm*. 2015;93:173-82.

58. Artursson P. Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorptive (Caco-2) cells. *J Pharm Sci.* 1990;79(6):476-82.

59. Wolk O, Agbaria R, Dahan A. Provisional in-silico biopharmaceutics classification (BCS) to guide oral drug product development. *Drug Des Devel Ther.* 2014;8:1563-75.

60. Dressman JB, Thelen K, Willmann S. An update on computational oral absorption simulation. *Expert Opin Drug Metab Toxicol*. 2011;7(11):1345-64.

61. O'Shea JP, Faisal W, Ruane-O'Hora T, Devine KJ, Kostewicz ES, O'Driscoll CM, et al. Lipidic dispersion to reduce food dependent oral bioavailability of fenofibrate: In vitro, in vivo and in silico assessments. *Eur J Pharm.* Biopharm. 2015;96:207-16.

62. Henze LJ, Koehl NJ, O'Shea JP, Kostewicz ES, Holm R, Griffin BT. The pig as a preclinical model for predicting oral bioavailability and in vivo performance of pharmaceutical oral dosage forms: a PEARRL review. *J Pharm Pharmacol.* 2019;71(4):581-602.

63. Hatton GB, Yadav V, Basit AW, Merchant HA. Animal Farm: Considerations in Animal Gastrointestinal Physiology and Relevance to Drug Delivery in Humans. *J Pharm Sci.* 2015;104(9):2747-76.

64. Amidon GL, Lennernas H, Shah VP, Crison JR. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm Res*. 1995;12(3):413-20.

65. Wu CY, Benet LZ. Predicting drug disposition via application of BCS: transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm Res.* 2005;22(1):11-23.

66. Zaki NM, Artursson P, Bergstrom CA. A modified physiological BCS for prediction of intestinal absorption in drug discovery. *Mol Pharm*. 2010;7(5):1478-87.

67. Rosenberger J, Butler J, Dressman J. A Refined Developability Classification System. *J Pharm Sci*. 2018;107(8):2020-32.

68. Kuentz M, Bergström CAS. Synergistic Computational Modeling Approaches as Team Players in the Game of Solubility Predictions. *J Pharm Sci*. 2021;110(1):22-34.

69. Martinez MN, Amidon GL. A mechanistic approach to understanding the factors affecting drug absorption: a review of fundamentals. *J Clin Pharmacol.* 2002;42(6):620-43.

70. Shultz MD. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. J. Med. Chem. 2019;62(4):1701-14.

71. Vistoli G, Pedretti A, Testa B. Assessing drug-likeness-what are we missing? *Drug Discov Today*. 2008;13(7-8):285-94.

72. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 1997;23(1):3-25.

73. Navia MA, Chaturvedi PR. Design principles for orally bioavailable drugs. *Drug Discovery Today*. 1996;1(5):179-89.

74. Kristl A, Tukker JJ. Negative correlation of n-octanol/water partition coefficient and transport of some guanine derivatives through rat jejunum in vitro. *Pharm Res.* 1998;15(3):499-501.

75. Artursson P, Karlsson J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun*. 1991;175(3):880-5.

76. Abraham MH, Chadha HS, Whiting GS, Mitchell RC. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the delta log P parameter of seiler. *J Pharm Sci*. 1994;83(8):1085-100.

77. Burton PS, Goodwin JT, Vidmar TJ, Amore BM. Predicting drug absorption: how nature made it a difficult problem. *J Pharmacol Exp Ther*. 2002;303(3):889-95.

78. Leeson PD, Davis AM. Time-Related Differences in the Physical Property Profiles of Oral Drugs. *J. Med. Chem*. 2004;47(25):6338-48.

79. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov*. 2007;6:881.

80. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem*. 2002;45(12):2615-23.

81. Ritchie TJ, Macdonald SJ. The impact of aromatic ring count on compound developability--are too many aromatic rings a liability in drug design? *Drug Discov Today.* 2009;14(21-22):1011-20.

82. Lobell M, Hendrix M, Hinzen B, Keldenich J, Meier H, Schmeck C, et al. In silico ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem.* 2006;1(11):1229-36.

83. Congreve M, Carr R, Murray C, Jhoti H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today*. 2003;8(19):876-7.

84. Rabinow BE. Nanosuspensions in drug delivery. *Nat. Rev. Drug Discov*. 2004;3(9):785-96.

85. Ku MS. Use of the Biopharmaceutical Classification System in early drug development. *AAPS J*. 2008;10(1):208-212.

86. Kuentz M, Imanidis G. In silico prediction of the solubility advantage for amorphous drugs - Are there property-based rules for drug discovery and early pharmaceutical development? *Eur J Pharm Sci*. 2013;48(3):554-62.

87. Kuentz M. Lipid-based formulations for oral delivery of lipophilic drugs. *Drug Discovery Today:Technologies.* 2012;9(2):97-104.

88. Williams HD, Sassene P, Kleberg K, Calderone M, Igonin A, Jule E, et al. Toward the establishment of standardized in vitro tests for lipid-based formulations, part 4: proposing a new lipid formulation performance classification system. *J Pharm Sci*. 2014;103(8):2441-55.

89. Törnqvist E, Annas A, Granath B, Jalkesten E, Cotgreave I, Öberg M. Strategic focus on 3R principles reveals major reductions in the use of animals in pharmaceutical toxicity testing. *PloS One*. 2014;9(7):e101638.

90. Benet LZ, Hosey CM, Ursu O, Oprea TI. BDDCS, the Rule of 5 and drugability. *Adv Drug Deliv Rev*. 2016;101:89-98.

91. Benet LZ, Broccatelli F, Oprea TI. BDDCS applied to over 900 drugs. *Aaps j*. 2011;13(4):519-47.

92. O'Shea JP, Holm R, O'Driscoll CM, Griffin BT. Food for thought: formulating away the food effect - a PEARRL review. *J Pharm Pharmacol*. 2019;71(4):510-35.

93. Williams HD, Ford L, Igonin A, Shan Z, Botti P, Morgen MM, et al. Unlocking the full potential of lipid-based formulations using lipophilic salt/ionic liquid forms. *Adv Drug Deliv Rev*. 2019;142:75-90.

94. Mullertz 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 Pharm Pharmacol.* 2010;62(11):1622-36.

95. 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):1743-58.

96. Zhang J, Han R, Chen W, Zhang W, Li Y, Ji Y, et al. Analysis of the Literature and Patents on Solid Dispersions from 1980 to 2015. *Molecules.* 2018;23(7).

97. Kawabata Y, Wada K, Nakatani M, Yamada S, Onoue S. Formulation design for poorly water-soluble drugs based on biopharmaceutics classification system: Basic approaches and practical applications. *Int. J. Pharm.* 2011;420(1):1-10.

98. Wyttenbach N, Kuentz M. Glass-forming ability of compounds in marketed amorphous drug products. *Eur J Pharm Biopharm*. 2017;112:204-8.

99. Strickley RG. Currently Marketed Oral Lipid-Based Dosage Forms. In: Hauss DJ, editor. Oral Lipid-Based Formulations Enhancing the Bioavailability of Poorly Water-Soluble Drugs. 170. *Drugs and Pharmaceutical Sciences*. Taylor & Francis Group; 2007. p. 1-32.

100. Chavan RB, Rathi S, Jyothi VGSS, Shastri NR. Cellulose based polymers in development of amorphous solid dispersions. *Asian J. Pharm. Sci.* 2019;14(3):248-64.

101. FDA. Drugs@FDA: FDAApproved Drugs. Drisol. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/appletter/2003/03444slr020ltr.pdf

102. Koehl NJ, Holm R, Kuentz M, Griffin BT. New Insights into Using Lipid Based Suspensions for 'Brick Dust' Molecules: Case Study of Nilotinib. *Pharm. Res*. 2019;36(4):56.

103. FDA. Drugs@FDA: FDAApproved Drugs. Cesamet. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2006/018677s011lbl.pdf

104. LaFountaine JS, McGinity JW, Williams RO, 3rd. Challenges and Strategies in Thermal Processing of Amorphous Solid Dispersions: A Review. *AAPS PharmSciTech.* 2016;17(1):43-55.

105. FDA. Drugs@FDA: FDAApproved Drugs. Rezulin. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/1999/20720s12lbl.pdf

106. Gilead Sciences. Press Releases: Gilead Sciences Announces Fourth Quarter and Full Year 2015 Financial Results. Available at: https://www.gilead.com/news-and-press/press-room/press-releases/2016/2/gilead-sciences-announces-fourth-quarter-and-full-year-2015-financial-results2016

107. FDA. Drugs@FDA: FDAApproved Drugs. Kaletra. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2016/021251s052_021906s046lbl.pdf

108. Administration FFaD. Drugs@FDA: FDA-Approved Drugs. Norvir. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/209512lbl.pdf

109. Chemburkar SR, Bauer J, Deming K, Spiwek H, Patel K, Morris J, et al. Dealing with the Impact of Ritonavir Polymorphs on the Late Stages of Bulk Drug Process Development. *Org. Process Res. Dev*. 2000;4(5):413-7.

110. Bauer J, Spanton S, Henry R, Quick J, Dziki W, Porter W, et al. Ritonavir: An Extraordinary Example of Conformational Polymorphism. *Pharm. Res.* 2001;18(6):859-66.

111. Vieth M, Siegel MG, Higgs RE, Watson IA, Robertson DH, Savin KA, et al. Characteristic Physical Properties and Structural Fragments of Marketed Oral Drugs. J. Med. Chem. 2004;47(1):224-32.

112. Stratton CF, Newman DJ, Tan DS. Cheminformatic comparison of approved drugs from natural product versus synthetic origins. *Bioorg. Med. Chem. Lett*. 2015;25(21):4802-7.

113. Mahlin D, Bergstrom CA. Early drug development predictions of glass-forming ability and physical stability of drugs. *Eur J Pharm Sci*. 2013;49(2):323-32.

114. Persson LC, Porter CJ, Charman WN, Bergstrom CA. Computational prediction of drug solubility in lipid based formulation excipients. *Pharm Res*. 2013;30(12):3225-37.

115. Alskar LC, Porter CJ, Bergstrom CA. Tools for Early Prediction of Drug Loading in Lipid-Based Formulations. *Mol Pharm*. 2016;13(1):251-61.

116. Goke K, Bunjes H. Drug solubility in lipid nanocarriers: Influence of lipid matrix and available interfacial area. *Int J Pharm*. 2017;529(1-2):617-28.

117. Humphrey MJ. Interface between drug discovery, ADME, and pharmaceutical development. *Bull. Tech*. Gattfosse2005. p. 65-74.

118. Wyttenbach N, Kirchmeyer W, Alsenz J, Kuentz M. Theoretical Considerations of the Prigogine-Defay Ratio with Regard to the Glass-Forming Ability of Drugs from Undercooled Melts. *Mol Pharm*. 2016;13(1):241-50.

119. Fagerberg JH, Karlsson E, Ulander J, Hanisch G, Bergström CA. Computational prediction of drug solubility in fasted simulated and aspirated human intestinal fluid. *Pharm Res*. 2015 Feb;32(2):578-89

120. Trevaskis NL, Charman WN, Porter CJ. Lipid-based delivery systems and intestinal lymphatic drug transport: a mechanistic update. *Adv Drug Deliv Rev*. 2008;60(6):702-16.

121. Branchu S, Rogueda PG, Plumb AP, Cook WG. A decision-support tool for the formulation of orally active, poorly soluble compounds. *Eur. J. Pharm. Sci.* 2007;32(2):128-39.

122. Upton RA, Buskin JN, Williams RL, Holford NHG, Riegelman S. Negligible excretion of unchanged ketoprofen, naproxen, and probenecid in urine. *J. Pharm. Sci.* 1980;69(11):1254-7.

123. Nurzyńska K, Booth J, Roberts CJ, McCabe J, Dryden I, Fischer PM. Long-Term Amorphous Drug Stability Predictions Using Easily Calculated, Predicted, and Measured Parameters. *Mol. Pharm*. 2015;12(9):3389-98.

124. Alhalaweh A, Alzghoul A, Kaialy W, Mahlin D, Bergström CAS. Computational Predictions of Glass-Forming Ability and Crystallization Tendency of Drug Molecules. *Mol. Pharm.* 2014;11(9):3123-32.

125. Kothari K, Ragoonanan V, Suryanarayanan R. The Role of Drug–Polymer Hydrogen Bonding Interactions on the Molecular Mobility and Physical Stability of Nifedipine Solid Dispersions. *Mol. Pharm*. 2015;12(1):162-70.

126. Taylor LS, Zografi G. Spectroscopic characterization of interactions between PVP and indomethacin in amorphous molecular dispersions. *Pharm Res*. 1997;14(12):1691-8.

127. Matsson P, Doak BC, Over B, Kihlberg J. Cell permeability beyond the rule of 5. *Adv. Drug Deliv. Rev*. 2016;101:42-61.

128. Larsen AT, Holm R, Mullertz A. Solution or suspension - Does it matter for lipid based systems? In vivo studies of chase dosing lipid vehicles with aqueous suspensions of a poorly soluble drug. *Eur J Pharm Biopharm*. 2017;117:308-14.

129. Williams HD, Sahbaz Y, Ford L, Nguyen TH, Scammells PJ, Porter CJ. Ionic liquids provide unique opportunities for oral drug delivery: structure optimization and in vivo evidence of utility. *Chem Commun (Camb)*. 2014;50(14):1688-90.

130. Fong SY, Bauer-Brandl A, Brandl M. Oral bioavailability enhancement through supersaturation: an update and meta-analysis. *Expert Opin Drug Deliv*. 2017;14(3):403-26.

131. Mehta CH, Narayan R, Nayak UY. Computational modeling for formulation design. *Drug Discovery Today*. 2019;24(3):781-8.

132. Ouyang D, Smith SC. Introduction to Computational Pharmaceutics in Computational Pharmaceutics 2015.

133. Damiati SA. Digital Pharmaceutical Sciences. *AAPS PharmSciTech*. 2020;21(6):206.

134. Kalyane D, Sanap G, Paul D, Shenoy S, Anup N, Polaka S, et al. Chapter 3 - Artificial intelligence in the pharmaceutical sector: current scene and future prospect. In: Tekade RK. The Future of Pharmaceutical Product Development and Research: Academic Press; 2020. p. 73-107.

135. Williams AM, Liu Y, Regner KR, Jotterand F, Liu P, Liang M. Artificial intelligence, physiological genomics, and precision medicine. *Physiol Genomics*. 2018;50(4):237-43.

136. Kim TH, Shin S, Shin BS. Model-based drug development: application of modeling and simulation in drug development. *Int. J. Pharm. Investig*. 2018;48(4):431-41.

137. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021;26(1):80-93.

138. Certara. The Benefits of Modeling and Simulation in Drug Development. 2016. Available at: https://www.certara.com/app/uploads/2020/06/WP_BenefitsOfBiosimulation-1-1.pdf

139. Mahajan R, Gupta K. Food and drug administration's critical path initiative and innovations in drug development paradigm: Challenges, progress, and controversies. *J Pharm Bioallied Sci.* 2010;2(4):307-13.

140. EMA. Regulatory Science Strategy. Big Data. 2021. Available at: https://www.ema.europa.eu/en/about-us/how-we-work/big-data

141. Pesqueira A, Sousa MJ, Rocha Á. Big Data Skills Sustainable Development in Healthcare and Pharmaceuticals. *J. Med. Syst*. 2020;44(11):197.

142. Elragal A, Klischewski R. Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. *Journal of Big Data*. 2017;4(1):19.

143. Marr B. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Forbes. 2018. Available at: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=3e35b3960ba9

144. Szlezák N, Evers M, Wang J, Pérez L. The role of big data and advanced analytics in drug discovery, development, and commercialization. *Clin Pharmacol Ther*. 2014;95(5):492-5.

145. Joshi N. Big Data in the Pharmaceutical Industry. BBN Times. 2019. Available at: https://www.bbntimes.com/technology/big-data-in-the-pharmaceutical-industry

146. Cattell J, Chilukuri S, Levy M, popup Oi. How Big Data Can Revolutionize Pharmaceutical R&D. 2013. Available at: https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d

147. Zhang L, Pfister M, Meibohm B. Concepts and challenges in quantitative pharmacology and model-based drug development. *Aaps j.* 2008;10(4):552-9.

148. Manhart K. Artificial Intelligence Modelling: Data Driven and Theory Driven Approaches. Troitzsch K.G. MU, Gilbert G.N., Doran J.E. (eds), editor: Springer; 1996.

149. Kouskoulis G, Spyropoulou I, Antoniou C. Pedestrian simulation: Theoretical models vs. data driven techniques. *Int. J. Transp. Sci. Technol*. 2018;7(4):241-53.

150. Ghasemi F, Mehridehnavi A, Pérez-Garrido A, Pérez-Sánchez H. Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. *Drug Discov Today*. 2018;23(10):1784-90.

151. Provost F, Fawcett T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*. 2013;1(1):51-9.

152. Karpatne A, Atluri G, Faghmous J. Theory-guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Trans Knowl Data Eng*. 2017;29:2318-31.

153. Cheng L, Wong H. Food Effects on Oral Drug Absorption: Application of Physiologically-Based Pharmacokinetic Modeling as a Predictive Tool. *Pharmaceutics*. 2020;12(7).

154. Winiwarter S, Bonham NM, Ax F, Hallberg A, Lennernäs H, Karlén A. Correlation of Human Jejunal Permeability (in Vivo) of Drugs with Experimentally and Theoretically Derived Parameters. A Multivariate Data Analysis Approach. J. Med. Chem. 1998;41(25):4939-49.

155. Norinder U, Bergström CA. Prediction of ADMET Properties. *ChemMedChem*. 2006;1(9):920-37.

156. Zhao Q, Ye Z, Su Y, Ouyang D. Predicting complexation performance between cyclodextrins and guest molecules by integrated machine learning and molecular modeling techniques. *Acta Pharm Sin B*. 2019;9(6):1241-52.

157. Hossain S, Kabedev A, Parrow A, Bergström CAS, Larsson P. Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning. *Eur J Pharm Biopharm*. 2019;137:46-55.

158. Yang Y, Ye Z, Su Y, Zhao Q, Li X, Ouyang D. Deep learning for in vitro prediction of pharmaceutical formulations. *Acta Pharmaceutica Sinica B*. 2019;9(1):177-85.

159. Alskar LC, Keemink J, Johannesson J, Porter CJH, Bergstrom CAS. Impact of Drug Physicochemical Properties on Lipolysis-Triggered Drug Supersaturation

and Precipitation from Lipid-Based Formulations. *Mol Pharm*. 2018;15(10):4733-44.

160. Alsenz J, Kuentz M. From Quantum Chemistry to Prediction of Drug Solubility in Glycerides. *Mol. Pharm.* 2019;16(11):4661-9.

161. Sacchetti M, Nejati E. Prediction of drug solubility in lipid mixtures from the individual ingredients. *AAPS PharmSciTech*. 2012;13(4):1103-9.

162. Brinkmann J, Exner L, Verevkin SP, Luebbert C, Sadowski G. PC-SAFT Modeling of Phase Equilibria Relevant for Lipid-Based Drug Delivery Systems. *J. Chem. Eng. Data.* 2021, 66, 3, 1280–1289.

163. Brinkmann J, Exner L, Luebbert C, Sadowski G. In-Silico Screening of Lipid-Based Drug Delivery Systems. *Pharm. Res*. 2020;37(12):249.

164. Benson SP, Pleiss J. Molecular dynamics simulations of self-emulsifying drug-delivery systems (SEDDS): influence of excipients on droplet nanostructure and drug localization. *Langmuir.* 2014;30(28):8471-80.

165. Warren DB, King D, Benameur H, Pouton CW, Chalmers DK. Glyceride lipid formulations: molecular dynamics modeling of phase behavior during dispersion and molecular interactions between drugs and excipients. *Pharm Res.* 2013;30(12):3238-53.

166. Birru WA, Warren DB, Han S, Benameur H, Porter CJ, Pouton CW, et al. Computational Models of the Gastrointestinal Environment. 2. Phase Behavior and Drug Solubilization Capacity of a Type I Lipid-Based Drug Formulation after Digestion. *Mol Pharm*. 2017;14(3):580-92.

167. DeBoyace K, Wildfong PLD. The Application of Modeling and Prediction to the Formation and Stability of Amorphous Solid Dispersions. *J Pharm Sci.* 2018;107(1):57-74.

168. Han R, Xiong H, Ye Z, Yang Y, Huang T, Jing Q, et al. Predicting physical stability of solid dispersions by machine learning techniques. *J Control Release*. 2019;311-312:16-25.

169. Gao H, Wang W, Dong J, Ye Z, Ouyang D. An integrated computational methodology with data-driven machine learning, molecular modeling and PBPK modeling to accelerate solid dispersion formulation design. *Eur J Pharm Biopharm*. 2021;158:336-46.

170. Mendyk A, Pacławski A, Szafraniec-Szczęsny J, Antosik A, Jamróz W, Paluch M, et al. Data-Driven Modeling of the Bicalutamide Dissolution from Powder Systems. *AAPS PharmSciTech*. 2020;21(3):111.

171. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 62019. p. 94-8.

172. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*. 1996;49(11):1225-31.

173. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*. 2020;408:189-215.

174. Marsland S. Machine Learning: An Algorithmic Perspective. 2 ed. Chapman & Hall/CRC2015.

175. Li M, Wang J. An Empirical Comparison of Multiple Linear Regression and Artificial Neural Network for Concrete Dam Deformation Modelling. *Math. Probl. Eng.* 2019;2019:7620948.

176. Panagou EZ, Mohareb FR, Argyri AA, Bessant CM, Nychas G-JE. A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints. *Food Microbiol*. 2011;28(4):782-90.

177. Mahani M, ShaikhGhomi H. Comparison of multiple linear regression, partial least squares and artificial neural network for quantitative structure retention relationships of some polycyclic aromatic hydrocarbons. *Anal. Methods.* 2012;4(10):3381-5.

178. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374(2065):20150202.

179. Ringnér M. What is principal component analysis? *Nat Biotechnol.* 2008;26(3):303-4.

180. Glassey J. Multivariate data analysis for advancing the interpretation of bioprocess measurement and monitoring data. *Adv Biochem Eng Biotechnol.* 2013;132:167-91.

181. Singh KP, Ojha P, Malik A, Jain G. Partial least squares and artificial neural networks modeling for predicting chlorophenol removal from aqueous solution. *Chemom. Intell. Lab. Syst.* 2009;99(2):150-60.

182. Ruiz-Perez D, Guan H, Madhivanan P, Mathee K, Narasimhan G. So you think you can PLS-DA? *BMC Bioinformatics*. 2020;21(Suppl 1):2.

183. Kausar N, Belhaouari Samir B, Abdullah A, Ahmad I, Hussain M, editors. A Review of Classification Approaches Using Support Vector Machine in Intrusion Detection. Informatics Engineering and Information Science; 2011 2011//; Berlin, Heidelberg: Springer Berlin Heidelberg.

184. Kavzoglu T, Colkesen I. A kernel functions analysis for support vector machines for land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* 2009;11(5):352-9.

185. 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.

186. Lou C. Artificial Neural Networks: Their Training Process and Applications: Whitman College; 2019.

187. Farzad A, Mashayekhi H, Hassanpour H. A comparative performance analysis of different activation functions in LSTM networks for classification. *Neural. Comput. Appl.* 2019;31(7):2507-21.

188. Walch K. The Increasing Use of AI In The Pharmaceutical Industry. Forbes. 2020. Available at: https://www.forbes.com/sites/cognitiveworld/2020/12/26/the-increasing-use-of-ai-in-the-pharmaceutical-industry/?sh=2a33b72e4c01

189. Welling PG. Effects of food on drug absorption. *Annu Rev Nutr.* 1996;16:383-415.

190. Koziolek M, Alcaro S, Augustijns P, Basit AW, Grimm M, Hens B, et al. The mechanisms of pharmacokinetic food-drug interactions – A perspective from the UNGAP group. *Eur. J. Pharm. Sci.* 2019;134:31-59.

191. Fleisher D, Li C, Zhou Y, Pao LH, Karim A. Drug, meal and formulation interactions influencing drug absorption after oral administration. Clinical implications. *Clin Pharmacokinet.* 1999;36(3):233-54.

192. FDA. Food-effect bioavailability and fed bioequivalence studies: guidance for industry. 2002. Available at: https://www.fda.gov/files/drugs/published/Food-Effect-Bioavailability-and-Fed-Bioequivalence-Studies.pdf

193. Farha M, Masson E, Tomkinson H, Mugundu G. Food Effect Study Design With Oral Drugs: Lessons Learned From Recently Approved Drugs in Oncology. *J Clin Pharmacol.* 2019;59(4):463-71.

194. Varum FJ, Hatton GB, Basit AW. Food, physiology and drug delivery. *Int J Pharm.* 2013;457(2):446-60.

195. Yasuji T, Kondo H, Sako K. The effect of food on the oral bioavailability of drugs: a review of current developments and pharmaceutical technologies for pharmacokinetic control. *Ther Deliv.* 2012;3(1):81-90.

196. Yan J-H. Food Effect on Oral Bioavailability: Old and New Questions. *Clin. Pharmacol.* 2017;6(4):323-30.

197. Zhuang X, Lu C. PBPK modeling and simulation in drug research and development. *Acta Pharmaceutica Sinica B.* 2016;6(5):430-40.

198. Toothaker RD, Welling PG. The Effect of Food on Drug Bioavailability. *Annu. Rev. Pharmacol. Toxicol.* 1980;20(1):173-99.

199. Koziolek M, Schneider F, Grimm M, Modeβ C, Seekamp A, Roustom T, et al. Intragastric pH and pressure profiles after intake of the high-caloric, high-fat meal as used for food effect studies. *J Control Release*. 2015;220(Pt A):71-8.

200. Zhang T, Wells E. A Review of Current Methods for Food Effect Prediction During Drug Development. *Curr. Pharmacol. Rep.* 2020;6(5):267-79.

201. Lentz KA. Current methods for predicting human food effect. *Aaps j.* 2008;10(2):282-8.

202. Christiansen ML, Müllertz A, Garmer M, Kristensen J, Jacobsen J, Abrahamsson B, et al. Evaluation of the Use of Göttingen Minipigs to Predict Food Effects on the Oral Absorption of Drugs in Humans. *J. Pharm. Sci.* 2015;104(1):135-43.

203. Kawai Y, Fujii Y, Tabata F, Ito J, Metsugi Y, Kameda A, et al. Profiling and trend analysis of food effects on oral drug absorption considering micelle interaction and solubilization by bile micelles. *Drug Metab Pharmacokinet.* 2011;26(2):180-91.

204. Borbás E, Kádár S, Tsinman K, Tsinman O, Csicsák D, Takács-Novák K, et al. Prediction of Bioequivalence and Food Effect Using Flux- and Solubility-Based Methods. *Mol Pharm.* 2019;16(10):4121-30.

205. Shono Y, Jantratid E, Janssen N, Kesisoglou F, Mao Y, Vertzoni M, et al. Prediction of food effects on the absorption of celecoxib based on biorelevant dissolution testing coupled with physiologically based pharmacokinetic modeling. *Eur. J. Pharm. Biopharm*. 2009;73(1):107-14.

206. Gamsiz ED, Ashtikar M, Crison J, Woltosz W, Bolger MB, Carrier RL. Predicting the effect of fed-state intestinal contents on drug dissolution. *Pharm Res.* 2010;27(12):2646-56.

207. Henze LJ, Koehl NJ, O'Shea JP, Holm R, Vertzoni M, Griffin BT. Toward the establishment of a standardized pre-clinical porcine model to predict food effects – Case studies on fenofibrate and paracetamol. *Int. J. Pharm. 2019*;1:100017.

208. Bennett-Lenane H, Jørgensen JR, Koehl NJ, Henze LJ, O'Shea JP, Müllertz A, et al. Exploring porcine gastric and intestinal fluids using microscopic and solubility estimates: Impact of placebo self-emulsifying drug delivery system administration to inform bio-predictive in vitro tools. *Eur J Pharm Sci.* 2021;161:105778.

209. Heimbach T, Xia B, Lin TH, He H. Case studies for practical food effect assessments across BCS/BDDCS class compounds using in silico, in vitro, and preclinical in vivo data. *Aaps j.* 2013;15(1):143-58.

210. Custodio JM, Wu CY, Benet LZ. Predicting drug disposition, absorption/elimination/transporter interplay and the role of food on drug absorption. *Adv Drug Deliv Rev*. 2008;60(6):717-33.

211. Riedmaier AE, DeMent K, Huckle J, Bransford P, Stillhart C, Lloyd R, et al. Use of Physiologically Based Pharmacokinetic (PBPK) Modeling for Predicting Drug-Food Interactions: an Industry Perspective. *The AAPS Journal*. 2020;22(6):123.

212. Kesisoglou F. Can PBPK Modeling Streamline Food Effect Assessments? *J Clin Pharmacol.* 2020;60 Suppl 1:S98-s104.

213. Singh BN. A quantitative approach to probe the dependence and correlation of food-effect with aqueous solubility, dose/solubility ratio, and partition coefficient (Log P) for orally active drugs administered as immediate-release formulations. *Drug Dev. Res*. 2005;65(2):55-75.

214. Gu CH, Li H, Levons J, Lentz K, Gandhi RB, Raghavan K, et al. Predicting effect of food on extent of drug absorption based on physicochemical properties. *Pharm Res*. 2007;24(6):1118-30.

215. Omachi F, Kaneko M, Iijima R, Watanabe M, Itagaki F. Relationship between the effects of food on the pharmacokinetics of oral antineoplastic drugs and their physicochemical properties. *J Pharm Health Care Sci*. 2019;5:26.

216. Aksu B, Paradkar A, de Matas M, Ozer O, Güneri T, York P. Quality by design approach: application of artificial intelligence techniques of tablets manufactured by direct compression. *AAPS PharmSciTech*. 2012;13(4):1138-46.

217. Damiati SA, Martini LG, Smith NW, Lawrence MJ, Barlow DJ. Application of machine learning in prediction of hydrotrope-enhanced solubilisation of indomethacin. *Int. J. Pharm.* 2017;530(1):99-106.

218. Manda A, Walker RB, Khamanga SMM. An Artificial Neural Network Approach to Predict the Effects of Formulation and Process Variables on Prednisone Release from a Multipartite System. *Pharmaceutics.* 2019;11(3).

219. Zhang Y, Evans JRG, Yang S. Exploring Correlations Between Properties Using Artificial Neural Networks. *Metall Mater Trans A Phys Metall Mater Sci.* 2020;51(1):58-75.

220. Alhalaweh A, Alzghoul A, Mahlin D, Bergström CAS. Physical stability of drugs after storage above and below the glass transition temperature: Relationship to glass-forming ability. *Int J Pharm*. 2015;495(1):312-7.

221. Ding X, Rose JP, Van Gelder J. Developability assessment of clinical drug products with maximum absorbable doses. *Int J Pharm*. 2012;427(2):260-9.

222. 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 Nov;109(11):3248-3261.

223. Bisgin H, Bera T, Ding H, Semey HG, Wu L, Liu Z, et al. Comparing SVM and ANN based Machine Learning Methods for Species Identification of Food Contaminating Beetles. *Scientific Reports*. 2018;8(1):6532.

224. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics.* 2020;21(1):6.

225. Alshalif SA, Ibrahim N, Herawan T, editors. Artificial Neural Network with Hyperbolic Tangent Activation Function to Improve the Accuracy of COCOMO II Model. *Recent Advances on Soft Computing and Data Mining*; 2017; Cham: Springer International Publishing.

226. Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks.* 1993;6(4):525-33.

227. Milton KA, DJ N. The predictability of food effects on bioavailability (BA). *Eur J Pharm Sci.* 1996;4(Suppl. 1):S94.

228. Ren J. ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *Knowledge-Based Systems*. 2012;26:144-53.

229. Jones HM, Gardner IB, Watson KJ. Modelling and PBPK simulation in drug discovery. *Aaps j.* 2009;11(1):155-66.

230. Marasanapalle VP, Boinpally RR, Zhu H, Grill A, Tang F. Correlation between the systemic clearance of drugs and their food effects in humans. *Drug Dev Ind Pharm.* 2011;37(11):1311-7.

231. Takagi T, Ramachandran C, Bermejo M, Yamashita S, Yu LX, Amidon GL. A provisional biopharmaceutical classification of the top 200 oral drug products in the United States, Great Britain, Spain, and Japan. *Mol Pharm.* 2006;3(6):631-43.

232. Pajouhesh H, Lenz GR. Medicinal chemical properties of successful central nervous system drugs. *NeuroRx.* 2005;2(4):541-53.

233. Palm K, Stenberg P, Luthman K, Artursson P. Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. *Pharm. Res.* 1997;14(5):568-71.

234. Paterson DA, Conradi RA, Hilgers AR, Vidmar TJ, Burton PS. A Non-aqueous Partitioning System for Predicting the Oral Absorption Potential of Peptides. Quantitative Structure-Activity Relationships. 1994;13(1):4-10.

235. 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 Jan;110(1):164-175

236. Chen XQ, Gudmundsson OS, Hageman MJ. Application of lipid-based formulations in drug discovery. *J Med Chem*. 2012;55(18):7945-56.

237. Cao Y, Marra M, Anderson BD. Predictive relationships for the effects of triglyceride ester concentration and water uptake on solubility and partitioning of small molecules into lipid vehicles. *J Pharm Sci*. 2004;93(11):2768-79.

238. Gautschi N, Bergstrom CA, Kuentz M. Rapid determination of drug solubilization versus supersaturation in natural and digested lipids. *Int J Pharm.* 2016;513(1-2):164-74.

239. Khan J, Rades T, Boyd B. The Precipitation Behavior of Poorly Water-Soluble Drugs with an Emphasis on the Digestion of Lipid Based Formulations. *Pharm Res.* 2016;33(3):548-62.

240. Bevernage J, Brouwers J, Annaert P, Augustijns P. Drug precipitation-permeation interplay: supersaturation in an absorptive environment. *Eur J Pharm Biopharm.* 2012;82(2):424-8.

241. Klein S. The use of biorelevant dissolution media to forecast the in vivo performance of a drug. *Aaps j.* 2010;12(3):397-406.

242. Schwebel HJ, van Hoogevest P, Leigh ML, Kuentz M. The apparent solubilizing capacity of simulated intestinal fluids for poorly water-soluble drugs. *Pharm Dev Technol.* 2011;16(3):278-86.

243. Fagerberg JH, Tsinman O, Sun N, Tsinman K, Avdeef A, Bergstrom CA. Dissolution rate and apparent solubility of poorly soluble drugs in biorelevant dissolution media. *Mol Pharm.* 2010;7(5):1419-30.

244. Van den Bergh A, Van Hemelryck S, Bevernage J, Van Peer A, Brewster M, Mackie C, et al. Preclinical Bioavailability Strategy for Decisions on Clinical Drug Formulation Development: An In Depth Analysis. *Mol Pharm.* 2018;15(7):2633-45.

245. Klumpp L, Leigh M, Dressman J. Dissolution behavior of various drugs in different FaSSIF versions. *Eur. J. Pharm. Sci.*. 2020;142:105138.

246. Augustijns P, Wuyts B, Hens B, Annaert P, Butler J, Brouwers J. A review of drug solubility in human intestinal fluids: implications for the prediction of oral absorption. *Eur J Pharm Sci.* 2014;57:322-32.

247. Wiest J, Saedtler M, Bottcher B, Grune M, Reggane M, Galli B, et al. Geometrical and Structural Dynamics of Imatinib within Biorelevant Colloids. *Mol Pharm.* 2018;15(10):4470-80.

248. Bhattacharjee S. DLS and zeta potential - What they are and what they are not? *J Control Release*. 2016;235:337-51.

249. Fagerberg JH, Bergstrom CA. Intestinal solubility and absorption of poorly water soluble compounds: predictions, challenges and solutions. *Ther Deliv.* 2015;6(8):935-59.

250. Forina M, Lanteri S, Oliveros MCC, Millan CP. Selection of useful predictors in multivariate calibration. *Anal. Bioanal. Chem.* 2004;380(3):397-418.

251. Hosey CM, Benet LZ. Predicting the extent of metabolism using in vitro permeability rate measurements and in silico permeability rate predictions. *Mol Pharm.* 2015;12(5):1456-66.

252. Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Applying computational modeling to drug discovery and development. *Drug Discovery Today*. 2006;11(17):806-11.

253. Bernkop-Schnürch A, Jalil A. Do drug release studies from SEDDS make any sense? *J Control Release*. 2018;271:55-9.

254. Tolls J, van Dijk J, Verbruggen EJM, Hermens JLM, Loeprecht B, Schüürmann G. Aqueous Solubility−Molecular Size Relationships: A Mechanistic Case Study Using C10- to C19-Alkanes. *J. Phys. Chem.* 2002;106(11):2760-5.

255. Wassvik CM, Holmén AG, Draheim R, Artursson P, Bergström CAS. Molecular Characteristics for Solid-State Limited Solubility. *J. Med. Chem.* 2008;51(10):3035-9.

256. Kier LB, Hall LH. An Electrotopological-State Index for Atoms in Molecules. *Pharm. Res.* 1990;7(8):801-7.

257. Clulow AJ, Parrow A, Hawley A, Khan J, Pham AC, Larsson P, et al. Characterization of Solubilizing Nanoaggregates Present in Different Versions of Simulated Intestinal Fluid. *J Phys Chem B*. 2017;121(48):10869-81.

258. Wickham M, Garrood M, Leney J, Wilson PD, Fillery-Travis A. Modification of a phospholipid stabilized emulsion interface by bile salt: effect on pancreatic lipase activity. *J Lipid Res.* 1998;39(3):623-32.

259. Fatouros DG, Bergenstahl B, Mullertz A. Morphological observations on a lipid-based drug delivery system during in vitro digestion. *Eur J Pharm Sci.* 2007;31(2):85-94.

260. Christensen J, Schultz K, Mollgaard B, Kristensen HG, Mullertz A. Solubilisation of poorly water-soluble drugs during in vitro lipolysis of medium- and long-chain triacylglycerols. *Eur J Pharm Sci.* 2004;23(3):287-96.

261. Mishra M. Vegetable Oil-based Formulations for Controlled Drug Delivery. Handbook of Encapsulation and Controlled Release: CRC Press; 2015. p. 1383.

262. Williams HD, Sassene P, Kleberg K, Calderone M, Igonin A, Jule E, et al. Toward the establishment of standardized in vitro tests for lipid-based formulations, part 3: understanding supersaturation versus precipitation potential during the in vitro digestion of type I, II, IIIA, IIIB and IV lipid-based formulations. *Pharm Res*. 2013;30(12):3059-76.

263. Kuentz M, Holm R, Elder DP. Methodology of oral formulation selection in the pharmaceutical industry. *Eur J Pharm Sci.* 2016;87:136-63.

264. Rane SS, Anderson BD. What determines drug solubility in lipid vehicles: is it predictable? *Adv Drug Deliv Rev.* 2008;60(6):638-56.

265. Niederquell A, Wyttenbach N, Kuentz M. New prediction methods for solubility parameters based on molecular sigma profiles using pharmaceutical materials. *Int J Pharm.* 2018;546(1-2):137-44.

266. Wyttenbach N, Niederquell A, Kuentz M. Machine Estimation of Drug Melting Properties and Influence on Solubility Prediction. *Mol. Pharm.* 2020;17(7):2660-71.

267. Farizawani AG, Puteh M, Marina Y, Rivaie A. A review of artificial neural network learning rule based on multiple variant of conjugate gradient approaches. *Journal of Physics: Conference Series*. 2020;1529:022040.

268. Barmpalexis P, Karagianni A, Nikolakakis I, Kachrimanis K. Artificial neural networks (ANNs) and partial least squares (PLS) regression in the quantitative analysis of cocrystal formulations by Raman and ATR-FTIR spectroscopy. *J. Pharm. Biomed. Anal.* 2018;158:214-24.

269. Galata DL, Farkas A, Könyves Z, Mészáros LA, Szabó E, Csontos I, et al. Fast, Spectroscopy-Based Prediction of In Vitro Dissolution Profile of Extended Release Tablets Using Artificial Neural Networks. *Pharmaceutics*. 2019;11(8).

270. Djuris J, Cirin-Varadjan S, Aleksic I, Djuris M, Cvijic S, Ibric S. Application of Machine-Learning Algorithms for Better Understanding of Tableting Properties of Lactose Co-Processed with Lipid Excipients. *Pharmaceutics*. 2021;13(5).

271. Tosca EM, Bartolucci R, Magni P. Application of Artificial Neural Networks to Predict the Intrinsic Solubility of Drug-Like Molecules. *Pharmaceutics.* 2021;13(7).

272. Van Hauwermeiren D, Stock M, De Beer T, Nopens I. Predicting Pharmaceutical Particle Size Distributions Using Kernel Mean Embedding. *Pharmaceutics.* 2020;12(3).

273. Michaelsen MH, Wasan KM, Sivak O, Müllertz A, Rades T. The Effect of Digestion and Drug Load on Halofantrine Absorption from Self-nanoemulsifying Drug Delivery System (SNEDDS). *Aaps j.* 2016;18(1):180-6.

274. Thomas N, Holm R, Garmer M, Karlsson JJ, Müllertz A, Rades T. Supersaturated self-nanoemulsifying drug delivery systems (Super-SNEDDS) enhance the bioavailability of the poorly water-soluble drug simvastatin in dogs. *Aaps j.* 2013;15(1):219-27.

275. Blaabjerg LI, Lindenberg E, Löbmann K, Grohganz H, Rades T. Is there a correlation between the glass forming ability of a drug and its supersaturation propensity? *Int J Pharm.* 2018;538(1-2):243-9.

276. Palmelund H, Madsen CM, Plum J, Müllertz A, Rades T. Studying the Propensity of Compounds to Supersaturate: A Practical and Broadly Applicable Approach. *J Pharm Sci.* 2016;105(10):3021-9.

277. Ilie A-R, Griffin BT, Vertzoni M, Kuentz M, Cuyckens F, Wuyts K, et al. Toward simplified oral lipid-based drug delivery using mono-/di-glycerides as single component excipients. *Drug Dev. Ind. Pharm*. 2020;46(12):2051-60.

278. Baghel S, Cathcart H, Redington W, O'Reilly NJ. An investigation into the crystallization tendency/kinetics of amorphous active pharmaceutical ingredients: A case study with dipyridamole and cinnarizine. *Eur J Pharm Biopharm*. 2016;104:59-71.

279. Alhalaweh A, Alzghoul A, Bergström CAS. Molecular Drivers of Crystallization Kinetics for Drugs in Supersaturated Aqueous Solutions. *J. Pharm. Sci.* 2019;108(1):252-9.

280. Kirkham MB. Chapter 3 - Structure and Properties of Water. In: Kirkham MB, editor. Principles of Soil and Plant Water Relations (Second Edition). Boston: Academic Press; 2014. p. 27-40.

281. Fradera X, Solà M. Second-order atomic Fukui indices from the electron-pair density in the framework of the atoms in molecules theory. *J Comput Chem*. 2004;25(3):439-46.

282. Fukui K, Yonezawa T, Shingu H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys*. 1952;20(4):772-25.

283. Teleki A, Nylander O, Bergström CAS. Intrinsic Dissolution Rate Profiling of Poorly Water-Soluble Compounds in Biorelevant Dissolution Media. *Pharmaceutics.* 2020;12(6).

284. Geidl S, Bouchal T, Raček T, Svobodová Vařeková R, Hejret V, Křenek A, et al. High-quality and universal empirical atomic charges for chemoinformatics applications. *J Cheminformatics*. 2015;7(1):59.

285. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*. 1980;36(22):3219-28.

286. Clarysse S, Brouwers J, Tack J, Annaert P, Augustijns P. Intestinal drug solubility estimation based on simulated intestinal fluids: comparison with solubility in human intestinal fluids. *Eur J Pharm Sci.* 2011;43(4):260-9.

287. Christiansen ML, Holm R, Abrahamsson B, Jacobsen J, Kristensen J, Andersen JR, et al. Effect of food intake and co-administration of placebo self-nanoemulsifying drug delivery systems on the absorption of cinnarizine in healthy human volunteers. *Eur. J. Pharm. Sci.* 2016;84:77-82.

288. Perlman ME, Murdande SB, Gumkowski MJ, Shah TS, Rodricks CM, Thornton-Manning J, et al. Development of a self-emulsifying formulation that reduces the food effect for torcetrapib. *Int J Pharm*. 2008;351(1-2):15-22.

289. Keemink J, Mårtensson E, Bergström CAS. Lipolysis-Permeation Setup for Simultaneous Study of Digestion and Absorption in Vitro. *Mol Pharm*. 2019;16(3):921-30.

290. Berthelsen R, Klitgaard M, Rades T, Müllertz A. In vitro digestion models to evaluate lipid based drug delivery systems; present status and current trends. *Adv Drug Deliv Rev.* 2019;142:35-49.

291. Henze LJ, Koehl NJ, O'Shea JP, Kostewicz ES, Holm R, Griffin BT. The pig as a preclinical model for predicting oral bioavailability and in vivo performance of pharmaceutical oral dosage forms: a PEARRL review. *J Pharm Pharmacol.* 2019;71(4):581-602.

292. Dressman JB, Vertzoni M, Goumas K, Reppas C. Estimating drug solubility in the gastrointestinal tract. *Adv Drug Deliv Rev*. 2007;59(7):591-602.

293. Merchant HA, Afonso-Pereira F, Rabbie SC, Youssef SA, Basit AW. Gastrointestinal characterisation and drug solubility determination in animals. *J Pharm Pharmacol.* 2015;67(5):630-9.

294. Henze LJ, Koehl NJ, Jansen R, Holm R, Vertzoni M, Whitfield PD, et al. Development and evaluation of a biorelevant medium simulating porcine gastrointestinal fluids. *Eur. J. Pharm. Biopharm.* 2020;154:116-26.

295. Mullertz A, Reppas C, Psachoulias D, Vertzoni M, Fatouros DG. Structural features of colloidal species in the human fasted upper small intestine. *J Pharm Pharmacol.* 2015;67(4):486-92.

296. Riethorst D, Baatsen P, Remijn C, Mitra A, Tack J, Brouwers J, et al. An In-Depth View into Human Intestinal Fluid Colloids: Intersubject Variability in Relation to Composition. *Mol Pharm*. 2016;13(10):3484-93.

297. Riethorst D, Mols R, Duchateau G, Tack J, Brouwers J, Augustijns P. Characterization of Human Duodenal Fluids in Fasted and Fed State Conditions. *J Pharm Sci*. 2016;105(2):673-81.

298. Tran T, Fatouros DG, Vertzoni M, Reppas C, Müllertz A. Mapping the intermediate digestion phases of human healthy intestinal contents from distal ileum and caecum at fasted and fed state conditions. *J Pharm Pharmacol.* 2017;69(3):265-73.

299. Fatouros DG, Walrand I, Bergenstahl B, Mullertz A. Colloidal structures in media simulating intestinal fed state conditions with and without lipolysis products. *Pharm Res*. 2009;26(2):361-74.

300. Mullertz A, Fatouros DG, Smith JR, Vertzoni M, Reppas C. Insights into intermediate phases of human intestinal fluids visualized by atomic force microscopy and cryo-transmission electron microscopy ex vivo. *Mol Pharm.* 2012;9(2):237-47.

301. Alskär LC, Parrow A, Keemink J, Johansson P, Abrahamsson B, Bergström CAS. Effect of lipids on absorption of carvedilol in dogs: Is coadministration of lipids as efficient as a lipid-based formulation? *J Control Release*. 2019;304:90-100.

302. Al-Gousous J, Tsume Y, Fu M, Salem II, Langguth P. Unpredictable Performance of pH-Dependent Coatings Accentuates the Need for Improved Predictive in Vitro Test Systems. *Mol. Pharm*. 2017;14(12):4209-19.

303. Löbenberg R, Amidon GL. Modern bioavailability, bioequivalence and biopharmaceutics classification system. New scientific approaches to international regulatory standards. *Eur. J. Pharm. Biopharm*. 2000;50(1):3-12.

304. Mudie DM, Murray K, Hoad CL, Pritchard SE, Garnett MC, Amidon GL, et al. Quantification of Gastrointestinal Liquid Volumes and Distribution Following a 240 mL Dose of Water in the Fasted State. *Mol. Pharm.* 2014;11(9):3039-47.

305. Moriguchi I, Hirono S, Liu Q, Nakagome I, Matsushita Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull*. 1992;40(1):127-30.

306. Kostewicz ES, Abrahamsson B, Brewster M, Brouwers J, Butler J, Carlert S, et al. In vitro models for the prediction of in vivo performance of oral dosage forms. *Eur J Pharm Sci.* 2014;57:342-66.

307. Miller DB, Spence JD. Clinical Pharmacokinetics of Fibric Acid Derivatives (Fibrates). *Clin. Pharmacokinet.* 1998;34(2):155-62.

308. Porter CJH, Williams HD, Trevaskis NL. Recent Advances in Lipid-Based Formulation Technology. *Pharm. Res.* 2013;30(12):2971-5.

309. Thomas N, Richter K, Pedersen TB, Holm R, Mullertz A, Rades T. In Vitro Lipolysis Data Does Not Adequately Predict the In Vivo Performance of Lipid-Based Drug Delivery Systems Containing Fenofibrate. *Aaps Journal*. 2014;16(3):539-49.

310. Li P, Ford L, Haque S, McInerney MP, Williams HD, Scammells PJ, et al. Lipophilic Salts and Lipid-Based Formulations: Enhancing the Oral Delivery of Octreotide. *Pharm Res*. 2021;38(6):1125-37.

311. Williams HD, Ford L, Lim S, Han S, Baumann J, Sullivan H, 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):203-16.

312. Schultz HB, Thomas N, Rao S, Prestidge CA. Supersaturated silica-lipid hybrids (super-SLH): An improved solid-state lipid-based oral drug delivery system with enhanced drug loading. *Eur. J. Pharm. Biopharm*.2018;125:13-20.

313. Bala V, Rao S, Li P, Wang S, Prestidge CA. Lipophilic Prodrugs of SN38: Synthesis and in Vitro Characterization toward Oral Chemotherapy. *Mol Pharm.* 2016;13(1):287-94.

314. Ilie AR, Griffin BT, Vertzoni M, Kuentz M, Kolakovic R, Prudic-Paus A, et al. Exploring precipitation inhibitors to improve in vivo absorption of cinnarizine from supersaturated lipid-based drug delivery systems. *Eur J Pharm Sci*. 2020:105691.

315. Selen A, Dickinson PA, Müllertz A, Crison JR, Mistry HB, Cruañes MT, et al. The biopharmaceutics risk assessment roadmap for optimizing clinical drug product performance. *J Pharm Sci.* 2014;103(11):3377-97.

316. Todeschini R, Consonni V. Molecular Descriptors for Chemoinformatics: Volume I & II Wiley-VCH Verlag GmbH & Co. KGaA, 2009.

317. Zhao YH, Abraham MH, Zissimos AM. Determination of McGowan Volumes for Ions and Correlation with van der Waals Volumes. *J Chem Inform Comput Sci.* 2003;43(6):1848-54.

318. Moriguchi I, Hirono S, Nakagome I, Hirano H. Comparison of Reliability of log P Values for Drugs Calculated by Several Methods. *Chem. Pharm. Bull.* 1994;42(4):976-8.

319. Van Eerdenbrugh B, Raina S, Hsieh YL, Augustijns P, Taylor LS. Classification of the crystallization behavior of amorphous active pharmaceutical ingredients in aqueous environments. *Pharm Res.* 2014;31(4):969-82.

320. Kalra A, Luner P, Taylor LS, Byrn SR, Li T. Gaining Thermodynamic Insight From Distinct Glass Formation Kinetics of Structurally Similar Organic Compounds. *J Pharm. Sci.* 2018;107(1):192-202.

321. Mahlin D, Ponnambalam S, Heidarian Höckerfelt M, Bergström CAS. Toward In Silico Prediction of Glass-Forming Ability from Molecular Structure Alone: A Screening Tool in Early Drug Development. *Mol. Pharm*. 2011;8(2):498-506.