Pharm Res (2013) 30:3225–3237
DOI 10.1007/s11095-013-1083-7

RESEARCH PAPER

# Computational Prediction of Drug Solubility in Lipid Based Formulation Excipients

Linda C. Persson • Christopher J. H. Porter • William N. Charman • Christel A. S. Bergström

Received: 3 January 2013 / Accepted: 12 May 2013 / Published online: 15 June 2013


## ABSTRACT

***Purpose*** To investigate if drug solubility in pharmaceutical excipients used in lipid based formulations (LBFs) can be predicted from physicochemical properties.

***Methods*** Solubility was measured for 30 structurally diverse drug molecules in soybean oil (SBO, long-chain triglyceride; $TG_{LC}$), Captex355 (medium-chain triglyceride; $TG_{MC}$), polysorbate 80 (PS80; surfactant) and PEG400 co-solvent and used as responses during PLS model development. Melting point and calculated molecular descriptors were used as variables and the PLS models were validated with test sets and permutation tests.

***Results*** Solvation capacity of SBO and Captex355 was equal on a mol per mol scale ($R^2 = 0.98$). A strong correlation was also found between PS80 and PEG400 ($R^2 = 0.85$), identifying the significant contribution of the ethoxylation for the solvation capacity of PS80. *In silico* models based on calculated descriptors were successfully developed for drug solubility in SBO ($R^2 = 0.81$, $Q^2 = 0.76$) and Captex355 ($R^2 = 0.84$, $Q^2 = 0.80$). However, solubility in PS80 and PEG400 were not possible to quantitatively predict from molecular structure.

***Conclusion*** Solubility measured in one excipient can be used to predict solubility in another, herein exemplified with $TG_{MC}$ *versus* $TG_{LC}$, and PS80 *versus* PEG400. We also show, for the first time, that solubility in $TG_{MC}$ and $TG_{LC}$ can be predicted from rapidly calculated molecular descriptors.

**KEY WORDS** computational prediction · lipid based formulation · loading capacity · molecular properties · solubility

L. C. Persson • C. A. S. Bergström (✉)
Department of Pharmacy
Drug Optimization and Pharmaceutical Profiling Platform
Uppsala University, Uppsala Biomedical Center
P.O. Box 580, 751 23 Uppsala, Sweden
e-mail: christel.bergstrom@farmaci.uu.se

C. J. H. Porter • W. N. Charman • C. A. S. Bergström
Drug Delivery, Disposition and Dynamics
Monash Institute of Pharmaceutical Sciences, Monash University,
381 Royal Parade, Parkville, Victoria 3052, Australia

## INTRODUCTION

For many new drug candidates, poor aqueous solubility is a significant barrier to effective drug development. Typically, poorly soluble compounds have erratic absorption from the gastrointestinal (GI) tract (1) and 70–90% of all discovery compounds have been estimated to have solubility-limited absorption (2,3). Erratic absorption raises safety concerns as irreproducible responses may lead to adverse effects or lack of therapeutic effect. This is often revealed late in development, in the worst case during clinical studies, resulting in costly and late project termination. For poorly soluble drugs the dosage form used to deliver the drug plays a critical role in improving absorption. Commonly, the formulation is selected by screening a number of standard formulations and, hence, requires compound synthesis (4–8). This process is time, labour and cost intensive. The overarching hypothesis that underpins the work in our laboratory is that significant improvements in time and efficiency could be accomplished by the development of computational tools to forecast the utility of different formulation strategies. The importance of this hypothesis is supported by recent work in which physiology-based pharmacokinetics were used in an attempt to predict drug exposure. In this case, only 23% of the drugs were predicted correctly after oral administration (compared to 69% of the i.v. drugs), and this was in part, attributed to poor prediction of formulation performance *in vivo* (9).

In response to the increased number of highly lipophilic, poorly water soluble compounds identified during lead optimization, interest in lipid based formulations (LBFs) as a solution to low solubility has increased (2,10). In contrast to conventional oral formulations such as tablets, LBFs usually present the drug to the stomach in a solubilized state. Further, LBFs maintain a supersaturated state in the intestinal fluid and hence, increase the concentration in the GI tract and facilitate GI absorption (4,11). Typically, LBFs consist of oil, surfactant, co-surfactant and water-soluble organic solvents in various proportions depending on the molecular properties of the drug and the purpose with the delivery (*i.e.* oral,

transdermal or injectable formulation). The lipid formulation classification system (LFCS), as proposed by Pouton (12,13), serves as a tool to classify and compare LBFs with regard to composition. Although this scheme has rationalized the design of and characterization needed for LBFs (14,15) the optimization of these formulations remains a complex, iterative and labor-intensive task.

One of the most important conditions for a successful lipid formulation is adequate drug solubility in the lipid system used (2,16). Efforts to facilitate computational prediction of this property are therefore warranted and are expected to increase the throughput and lower the costs of lipid formulation development (2). As described by Anderson *et al.* (17), ideal solubility theory and regular solution theory fail to give an accurate prediction of drug solubility in polar organic lipid solutes and solvents, due to the absence of molecular interactions in the calculations. Similarly, Thi *et al.* (16) examined the solubility of ten compounds in ten LBF excipients, but were unable to find a clear link between the physicochemical properties of the drugs investigated and solubility in the excipients. Recent advances in computational technologies, however, have allowed the development of more complex *in silico* simulations and models, in which molecular structure, physiochemical properties and specific solute-solvent interactions may be taken into account. For example, Rane *et al.* (18,19) have provided an improved understanding of drug solubility in mono- and triglycerides by using molecular dynamics simulations of mixtures of tricaprylin and 1-mono-caprylin. Through these simulations, the authors found that drug solubility in such systems was dependent on the inevitable presence of water and whether the drug resides in the lipid or water phase, or at the lipid-water interface. Similar methodologies have also been applied to other types of lipidic systems (20–22).

To this point, no models exist that accurately predict *de novo* drug solubility in the excipients commonly included in LBFs. Therefore, other means to rationalize the formulation design have been established. Calculations of solubility in formulations based on phase diagrams or more simplified calculations based on solubility determination in the single excipient have recently shown promising results (23,24). In the latter, the amount dissolved in a formulation is calculated based on the fraction of each excipient and the solubility measured in that particular excipient, and the sum of these values provide the maximum solubility in the formulation. Although these studies yield promising results they have so far investigated only a few compounds, and more importantly, still require extensive experimental work.

In aqueous systems, solubility has successfully been predicted from physiochemical properties and molecular structures using the general solubility theory (25) and more advanced computational methods such as partial least squares (PLS) models (26,27). In the current study we have therefore investigated whether similar approaches are feasible for the prediction of drug solubility in excipients commonly used in LBFs. To this end, the solubility of 30 structurally diverse, poorly water soluble drug molecules has been measured in four exemplar excipients; soybean oil (SBO; a long chain triglyceride ($TG_{LC}$)), Captex355 (a medium chain triglyceride ($TG_{MC}$)), polysorbate 80 (PS80, a surfactant) and polyethylene glycol 400 (PEG400, a co-solvent). These excipients were selected to provide examples of the major classes of excipients used in the LFCS. Thus; SBO and Captex355 are commonly included in LFCS formulations I–III and PS80 and PEG400 in LFCS class III–IV (12,13). The measured solubility data was subsequently analyzed together with calculated and measured physicochemical properties using multivariate data analysis in order to develop predictive computational models and an improved understanding of solubility in these systems.

## METHODS

### Dataset Selection and Characteristics

A dataset of 30 structurally diverse compounds were selected for this study (Table I and Fig. 1). Compounds with a calculated logP greater than 2 were selected to focus on those with poor aqueous solubility for which LBFs typically improve bioavailability, minimize interindividual differences in absorption and reduce food effects (11,28,29). All drug compounds were purchased from SigmaAldrich (USA) except acitretin (Ontario chemicals Inc., Canada), candesartan and candesartan cilexetil (Angene ltd, China), danazol (Coral drugs IVT, India), fenofibric acid (Labratoreo chimico internazionale, Italy), halofantrine (SmithKline Beecham Pharmaceuticals, India) and itraconazole (Lee Pharma Ltd, India). Felodipine was a gift from AstraZeneca (Mölndal, Sweden).

SBO, PS80 and PEG400 were purchased from SigmaAldrich (USA). The representative fatty acid composition of SBO was found to be linoleic acid 51% (Mw 280.45 g/mol), oleic acid 25% (Mw 282.46 g/mol), palmitic acid 10% (Mw 256.42 g/mol), linolenic acid 7% (Mw 278.43 g/mol) and stearic acid 5% (Mw 284.48 g/mol) (30), resulting in an average fatty acid Mw of 273.0 g/mol. Captex355 was purchased from Abitec (Janesville,WI). Captex355 is described in the product specification to comprise caprylic acid 54.8% (Mw 144.21 g/mol), capric acid 44.5% (Mw 172.26 g/mol) and lauric acid 0.5% (Mw 200.52 g/mol) providing an average fatty acid equivalent Mw of 156.7 g/mol. Acetonitrile (analytical grade) was purchased from Ajax Chemicals (Australia).

**Table I** Physicochemical Properties of Investigated Compounds[a]

| Compound | Mw (Da) | logP | Tm (°C) | PSA (Å$^2$) | Number of | | |
|---|---|---|---|---|---|---|---|
| | | | | | Nitrogens | Double bonds | Rotable bonds |
| Acitretin | 326.5 | 5.6 | 221 | 46.5 | 0 | 5 | 6 |
| Bezafibrate | 361.9 | 3.8 | 185 | 75.6 | 1 | 2 | 7 |
| Candesartan | 440.5 | 4.6 | 178 | 118.8 | 6 | 1 | 7 |
| Candesartan cilexetil | 610.7 | 7.4 | 167 | 143.3 | 6 | 2 | 13 |
| Cinnarizine | 368.6 | 5.5 | 119 | 6.5 | 2 | 1 | 6 |
| Clotrimazole | 344.9 | 5.2 | 142 | 17.8 | 2 | 0 | 4 |
| Danazol | 337.5 | 4.9 | 227 | 46.3 | 1 | 1 | 0 |
| Diflunisal | 250.2 | 3.1 | 213 | 57.5 | 0 | 1 | 2 |
| Disulfiram | 296.6 | 4.6 | 67 | 121.3 | 2 | 2 | 7 |
| Ethinylestradiol | 296.4 | 4.9 | 183 | 40.5 | 0 | 0 | 0 |
| Felodipine | 384.3 | 3.6 | 143 | 64.6 | 1 | 4 | 6 |
| Fenbendazole | 299.4 | 3.8 | 226 | 92.3 | 3 | 1 | 4 |
| Fenofibrate | 360.9 | 5.1 | 79 | 52.6 | 0 | 2 | 7 |
| Fenofibric acid | 318.8 | 4.1 | 184 | 63.6 | 0 | 2 | 5 |
| Glibenclamide | 494.1 | 4.1 | 174 | 122.0 | 3 | 4 | 8 |
| Halofantrine | 500.5 | 8.2 | 77 | 23.5 | 1 | 0 | 10 |
| Haloperidol | 375.9 | 3.9 | 151 | 40.5 | 1 | 1 | 6 |
| Indomethacin | 357.8 | 4.2 | 160 | 68.5 | 1 | 2 | 4 |
| Itraconazole | 705.7 | 6.5 | 166 | 104.7 | 8 | 2 | 11 |
| Ivermectin | 875.2 | 4.7 | 150 | 170.1 | 0 | 5 | 8 |
| Levothyroxine | 776.9 | 4.6 | 235 | 92.8 | 1 | 1 | 5 |
| Niclosamide | 327.1 | 3.6 | 231 | 95.2 | 2 | 3 | 3 |
| Noscapine | 413.5 | 3.0 | 175 | 75.7 | 1 | 1 | 4 |
| Perphenazine | 404.0 | 4.2 | 94 | 59.9 | 3 | 0 | 6 |
| Praziquantel | 312.5 | 2.7 | 139 | 40.6 | 2 | 2 | 1 |
| Progesterone | 314.5 | 3.6 | 128 | 34.1 | 0 | 3 | 1 |
| Saquinavir | 670.9 | 3.9 | nd | 166.8 | 6 | 4 | 13 |
| Sulfasalazine | 398.4 | 2.0 | 255 | 146.2 | 4 | 7 | 6 |
| Tolfenamic acid | 261.7 | 4.1 | 213 | 49.3 | 1 | 1 | 3 |
| Toltrazuril | 425.4 | 6.1 | 192 | 111.4 | 3 | 3 | 4 |

[a] All physiochemical properties were calculated with DragonX 1.4 (Talete, Italy) except melting point (Tm) which was experimentally determined with differential scanning calorimetry (see the Methods section). Tm of levothyroxine was taken from Merck index (35). The logP column displays the calculated AlogP from DragonX 1.4 (Talete, Italy)

## Solid State Characterization

The melting temperature (Tm), heat of fusion ($\Delta H_f$), entropy of fusion ($\Delta S_f$), crystallinity and purity were determined for each compound by differential scanning calorimetry (DSC). Thermograms were recorded with a DSC6200, Seiko, Japan coupled to an automatic cooling system. A sample of 1–3 mg was placed in a sealed and pierced aluminium pan (TA Instruments, Delaware), heated from room temperature to approximately 30°C above their expected Tm at a rate of 10°C/min and purged with nitrogen gas at a flow rate of 80 mL/min. For saquinavir thermal analyses were also performed at lower (2°C/min) and higher (50°C/min) flow rates. For candesartan, levothyroxine and saquinavir the solid state transformation was further observed in a capillary melting point apparatus (Electrothermal, England).

## Solubility Measurement

The solubility studies were performed based on previously described standard procedures (31). An excess amount of drug was added to triplicate glass vials containing each of the four excipients, and the vials were vortexed thoroughly and placed in a 37°C incubator for the period of the solubility study. Vials were tightly sealed and vortexed periodically to keep the drug suspended and were sampled at 24, 48 and 72 h, or longer if

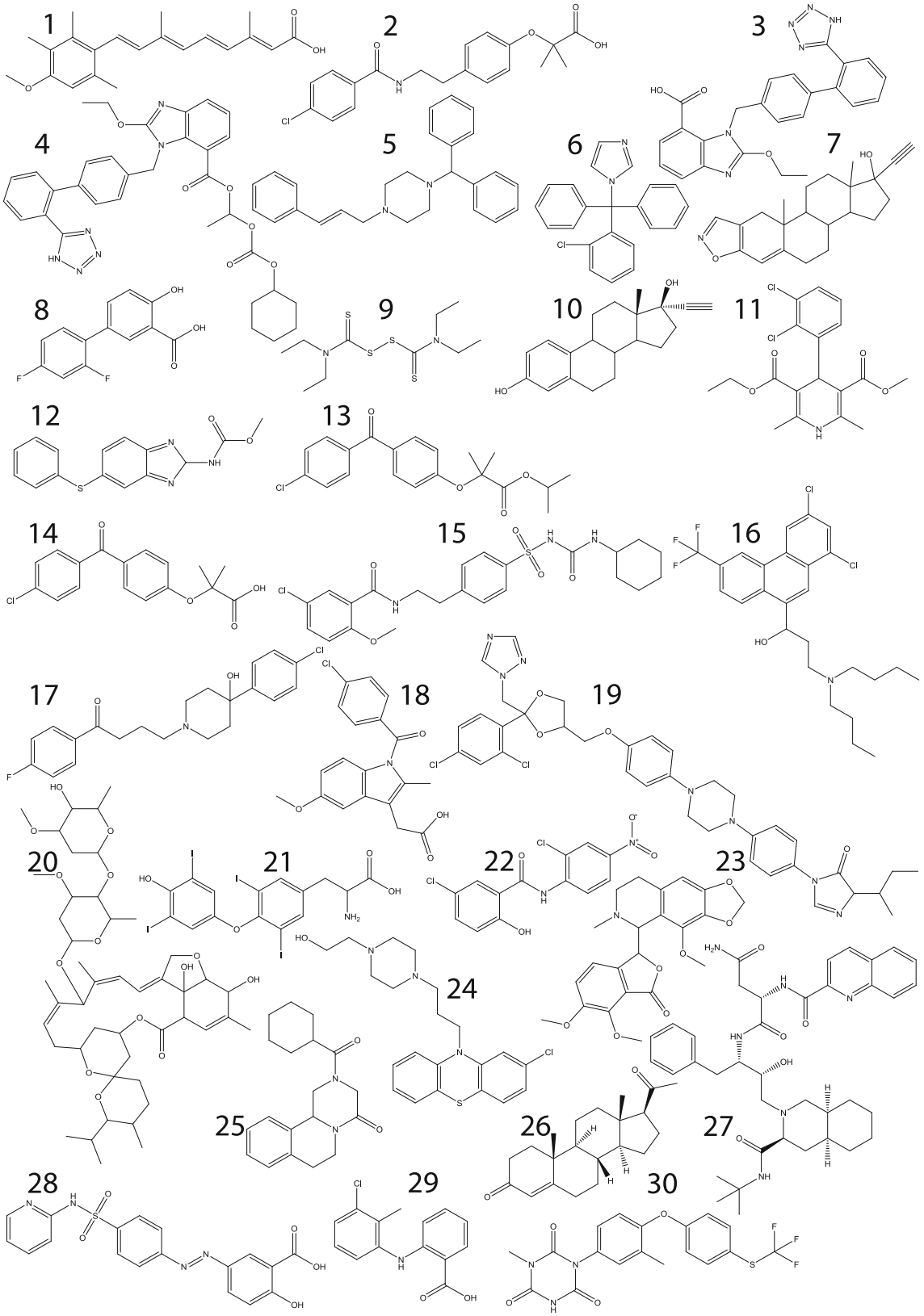


**Fig. 1** Molecular structures of the compounds studied. 1. Acitretin; 2. Bezafibrate; 3. Candesartan; 4. Candesartan cilexetil; 5 Cinnarizine; 6. Clotrimazole; 7. Danazol; 8. Diflunisal; 9. Disulfiram; 10. Ethinylestradiol; 11. Felodipine; 12. Fenbendazole; 13. Fenofibrate; 14. Fenofibric acid; 15. Glibenclamide; 16. Halofantrine; 17. Haloperidol; 18. Indomethacin; 19. Itraconazole; 20. Ivermectin; 21. Levothyroxine; 22. Niclosamide; 23. Noscapine; 24. Perphenazine; 25. Praziquantel; 26. Progesterone; 27. Saquinavir; 28. Sulfasalazine; 29. Tolfenamic acid; 30. Toltrazuril.

required to reach equilibrium. Prior to sampling, the vials were centrifuged at 37°C, 2,800 g for 30 min in a temperature controlled centrifuge (Eppendorf centrifuge 5804R). Approximately 2 drops (20–30 mg) of the supernatant were transferred into tared 5 ml volumetric flask and diluted with 66% v/v chloroform in methanol for the SBO samples or 7% v/v chloroform in methanol for all other samples. Highly concentrated samples were further diluted prior to the analysis if needed to allow quantification using the standard curve established. Drug concentrations were subsequently determined by reverse phase HPLC (Waters 2795 alliance HT, Waters 2489 UV/visible detector), using a Phenomenex $C_{18}$ Gemini 5 μm column (3.0×150 mm). The compounds were analyzed with suitable mobile phases at a flow rate of 1 ml/min using compound specific wavelengths. Equilibrium solubility was determined as the value when the solubility between two consecutive samples points (24 h time difference) differed by less than 10%.

To address the impact of water sorption in ethoxylated excipients four compounds (itraconazol, candesartan, danazol and indomethacin), were chosen as model drugs for complementary solubility studies performed under dry conditions. The solubility of the four drugs ranged from low to high in PEG400 and represents neutral, basic and acidic compounds. The same protocol as for the solubility studies in ambient conditions were followed, but to obtain an anhydrous milieu samples were placed in a desiccator containing phosphopentaoxide and only sampled once after 72 h or 96 h to minimize water sorption. The relative humidity in the desiccator was monitored throughout the experiment and maintained ≤ 1.8%. Anhydrous PEG400 could not be purchased but a freshly opened container of PEG400 declared to contain <0.5% w/w water was used. Later the water content in both PEG400 containers was determined by Karl Fischer titration in room temperature and under ambient conditions (KF Coulometer 831, Metrohm). The PEG400 samples were first dissolved in a 1:3 proportion with anhydrous methanol (Hydranal-Methanol-Dry). Approximately 2 ml of pre-diluted sample were injected to the reaction vessel. The titrants used were Hydranal-Coulomat AG (anodic compartment) and Hydranal-Coulomat CG (cathodic compartment). The drift was recorded to <2 ug/ml and all determinations were performed in triplicates.

## Statistics and Model Development

Experimentally determined solubility values are reported as mean ± standard deviation ($n \geq 3$). A two tailed *t*-test (assuming equal variance) was performed in excel (Microsoft Office Professional Plus 2010) to assess whether a significant difference in the solubility values determined under ambient or dry conditions existed. Linear regressions were also performed in excel, for standard curves and simple correlations $R^2$ (coefficient of determination) was used to validate the goodness of fit.

To investigate how molecular features and physiochemical properties influence solubility and whether solubility in excipients commonly used in LBFs can be predicted from molecular properties multivariate data analysis (Simca v13, Umetrics, Sweden) was performed. Corina 3.0 (Molecular networks, Erlangen, Germany) was used to convert SMILES strings into three-dimensional structures, which then were used as input for calculation of molecular descriptors with DragonX 1.4 (Talete, Italy). The dataset was sorted into training and test set. Strong outliers identified in the DModX plot of the principal component analysis (PCA) of the dataset were excluded from the training set, and instead placed in the test set, as these otherwise may distort the model development. The responses used were the logarithm of the solubility in SBO, Captex355, PS80 and PEG400 presented as mol compound/mol excipient. The average molar mass used were 273 g/mol for SBO, 156.7 g/mol for Captex355,1310 g/mol for PS80 and 400 g/mol for PEG400. Compounds with qualitative solubility values, *e.g.* compounds for which solubility was not determined to better accuracy than 'smaller than'/'greater than' values, were excluded from the model development. For the remaining compounds PCA was applied to randomly select the training and the test set with the criterion that the training set should cover the chemical space of the test set (Fig. 2). Partial least squares projection to latent structures (PLS) was then used to identify trends, to predict quantitative response values and to understand differences between the different excipients studied. The PLS model development followed the standard steps described in previous publications from our group (26,32,33). Firstly, all descriptors were de-identified, mean centered and scaled to unity of variance followed by removal of skewed descriptors. After the initial steps the matrix submitted for variable selection consisted of 725 variables. The variable selection was performed in order to decrease complexity, increase interpretability and robustness (*i.e.* reduce noise) of the model and to identify the key molecular properties of highest importance for excipient solubility. In the next step, all variables, except the 100 found to be most important for the response, were excluded based on the variable of importance (VIP) graph. Thereafter additional variables were removed; those removed were identified as having low importance for the response and/or having information duplicated by other variables and hence, positioned in the same area in the loading plot. The variable selection was monitored by $R^2$ and the leave-one-out cross-validated by $R^2$ ($Q^2$) using 7 cross-validation groups. If the exclusion of variables did not affect, or resulted in an increase in the $Q^2$, the variables were excluded permanently from the model. The accuracy of the PLS models was validated by root-mean square error of the estimate (RMSEE) calculations and permutation tests (100 iterations). The final models were validated with test sets.

## A General Solubility Equation for Lipids

Multiple linear regression (MLR) was performed in excel (Microsoft Office Professional Plus 2010) to investigate

correlations between solubility in excipients and selected variables, with the aim of achieving an easily interpretable equation similar to the general solubility equation (GSE) described by Yalkowsky and coworkers for aqueous systems (25). A number of easily calculated and transparent variables that were found important during the PLS model development was included in the MLR analysis, these being the number of nitrogens (nN), number of double bonds (nDB), polar surface area (PSA) and number of hydrogen acceptors (nHacc). In addition, melting point (Tm) was included since this has been found important in previous studies of homologous series of compounds (34). Moreover, the $\Delta S_f$ and the value calculated from $\Delta S_f$(Tm-25)/1364 were investigated as a substitute to the melting point since these have been shown by Yalkowsky *et al.* to be closely related to the solubility of liquid or crystalline organic nonelectrolytes in aqueous systems (25). Lipophilicity (AlogP descriptor obtained from Dragon 1.4) was included to reveal the significance and contribution of this physicochemical property to lipid solubility. MLR was performed using the same training and test sets as those used in the PLS model development to allow complete comparison between the resulting predictions.

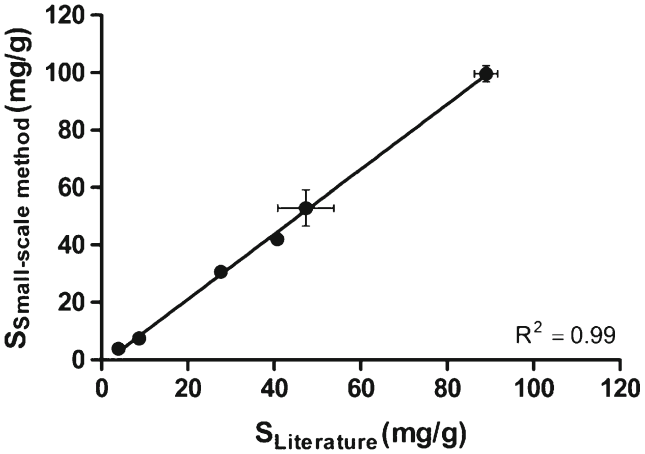


**Fig. 3** Comparison between the established small-scale method and literature values using larger volumes of excipients. The downscaled method was in excellent agreement with literature data and still allowed at least six consecutive samples to be taken for slowly dissolving compounds. Literature data were taken from Kaukonen *et al.* (37) and the following compounds are included in the validation: cinnarizine, danazol and halofantrine solubility in SBO and Captex355.

## RESULTS

### Solid State Characterization

The DSC measurements generated sharp peaks in the thermograms when the compounds melted and the Tm (Table I) correlated well with literature data for the stable polymorphs. However, the thermograms for candesartan and levothyroxine were unclear and therefore these compounds were further studied in a capillary melting point apparatus (Electrothermal, England). The Tm of candesartan was confirmed by this approach to be 178°C. Levothyroxine melted quickly above 200°C and the Tm could not be accurately determined with this method and hence, the literature value of 235°C as reported in Merck Index (35) was used in the modelling. Saquinavir did not produce a sharp peak in the thermogram despite repeated measurements using different heating rates. Through step-wise analysis of saquinavir, in

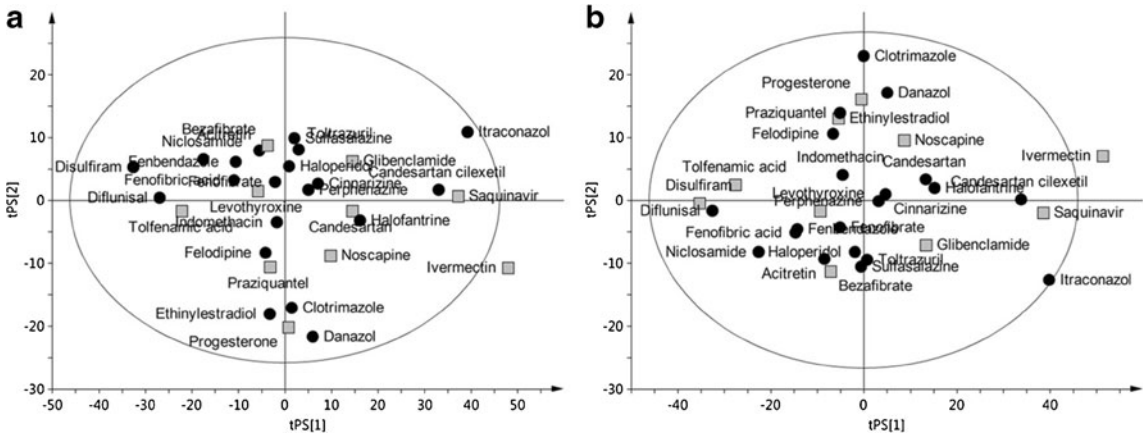


**Fig. 2** Principal component analysis (PCA) of chemical space covered by the studied compounds. PCA was used to select training and test sets for the solubility model development. Principal components 1 and 2 are shown explaining 52% of the chemical variation observed in the dataset. **a** The training (*black dots*) and test (*grey squares*) sets used for SBO and Captex355 model development. **b** The training (*black dots*) and test (*grey squares*) sets used for PS80 and PEG400 model development. The training set was mainly kept constant in all models, but smaller changes had to be made due to the slight difference in compounds that were quantitatively determined in the four different vehicles.

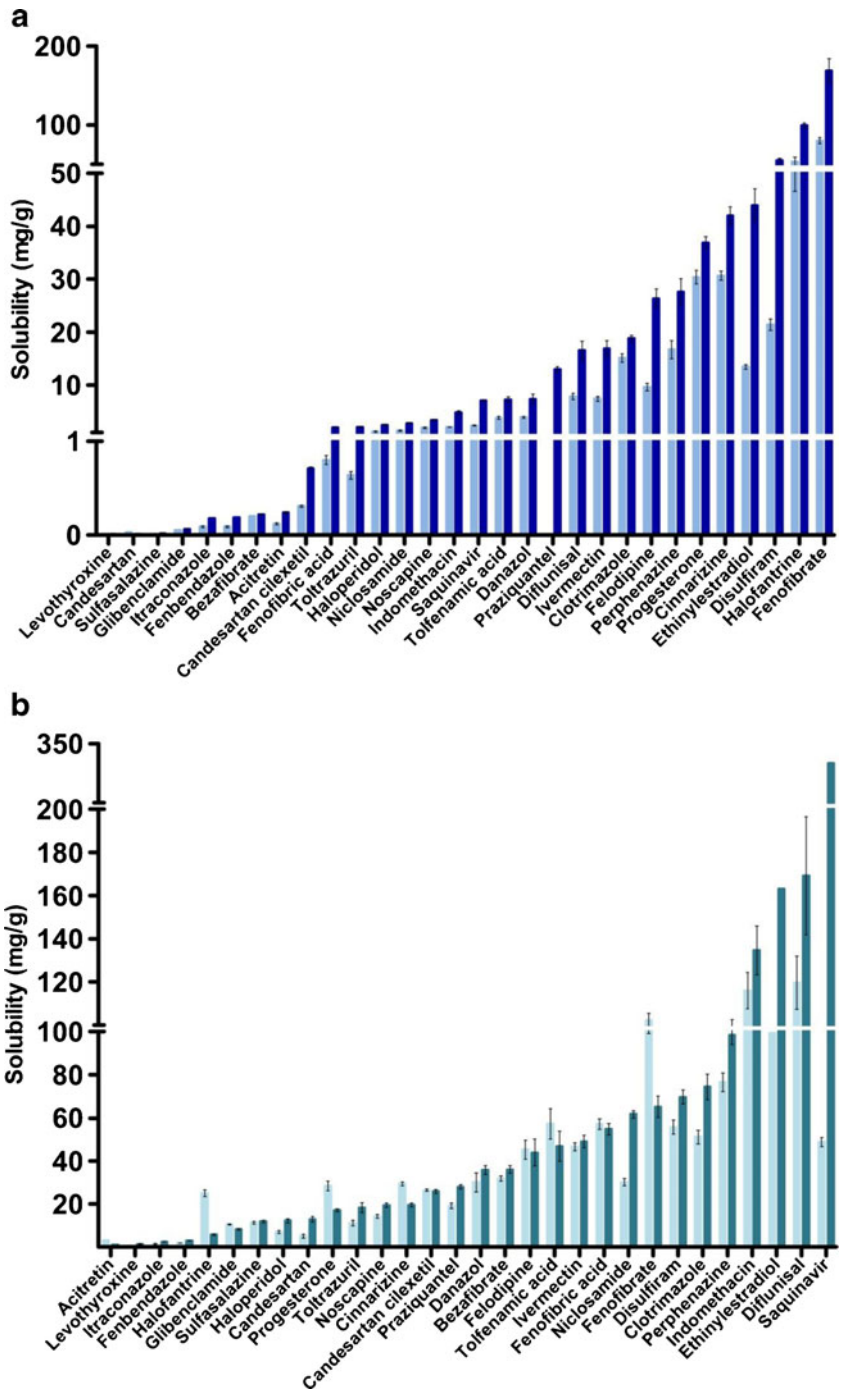


**Fig. 4** Measured solubility in four different excipients. (**a**) Solubility in soybean oil (SBO; *light blue*) and Captex355 (*dark blue*). In SBO bezafibrate, candesartan, glibenclamide and levothyroxine were determined qualitatively due to the solubility being less than the limit of detection of the HPLC method used. Praziquantel could not be detected due to interfering peaks with the SBO itself. In Captex355 the solubility in levothyroxine was determined qualitatively. (**b**) Solubility in polysorbate 80 (PS80; *dark blue*) and polyethylene glycol 400 (PEG400; *light blue*). In both excipients ethinylestradiol was determined qualitatively due to the high solubility and limited amount of compound available. In PEG400 also saquinavir was determined qualitatively for the same reason. All solubility values are plotted as mean ± standard deviation.

which several samples were run but terminated at different temperatures in the interval of 150–300°C, it was suggested by visual examination of the samples that the material gradually transformed and likely decomposed as visualised by colour change. The gradual solid state transformation was also observed in the capillary melting point apparatus where solid saquinavir slowly liquefied starting at ~100°C. In conclusion, saquinavir does not appear to have a sharp melting point and at high temperatures decomposes. As such a Tm could not be determined.

## Solubility in Excipients

Thirty poorly soluble drugs were selected and measured for solubility in four commonly used excipients of LBFs. The dataset was selected to be as diverse in chemical properties

as possible but still suitable for development as LBFs. The dataset had the following physicochemical properties: lipophilicity (reflected by the calculated octanol/water partition coefficient, $AlogP_{oct}$) 2.0 to 8.2, molecular weight 250.2 g/mol to 875.2 g/mol, polar surface are (PSA) 6.5–171.1 (Table I) and the compounds were selected to be exemplar structures of acids, bases and non-ionizable compounds. During the course of the work the standard solubility method was down scaled and the initial amount of excipient used (2 g) was lowered to 1 g to reduce the amount of drug needed. This downsizing resulted in the same solubility values as those measured in the larger scale (Fig. 3). The determined equilibrium solubilities ranged from <0.01 mg/g to 79.9 mg/g in SBO (Fig. 4a), <0.01 mg/g to 168.8 mg/g in Captex355 (Fig. 4a), 0.7 mg/g to >119.7 mg/g in PS80 (Fig. 4b), and 1.1 mg/g to >300.0 mg/g in PEG400 (Fig. 4b).

For this dataset it was observed that the solubility in Captex355 frequently appeared to be twice as high as the solubility in SBO when reported in mg/g (Fig. 4a). However, when the solubilities were converted to mol solubility the correlation became close to linear ($R^2$ of 0.98) (Fig. 5a), which demonstrates equal solvation capacity for the two excipients. Interestingly, the correlation between solvation capacity of PS80 and PEG400 also proved to be strong ($R^2$ of 0.85) (Fig. 5b).

## Prediction of Solubility in Excipients

The model development was performed in three steps and the results are presented in Tables II and III. Firstly, PLS models were developed for all excipients using calculated molecular descriptors. This resulted in excellent predictions for SBO ($R^2$ of 0.81, $Q^2$ of 0.76; Fig. 6a) and Captex355 ($R^2$ of 0.84, $Q^2$ of 0.80; Fig. 7a) based on only a few calculated descriptors. For PS80 and PEG400 only qualitative models could be developed resulting in $R^2<0.62$ for both vehicles (data not shown). In the next step we therefore included the experimentally determined Tm to analyze if this property would strengthen, in particular, the predictions obtained for PS80 and PEG400. Interestingly, inclusion of Tm did not improve the prediction of solubility in PS80 and PEG400 and was in fact identified as a variable of minor importance and excluded early during model development. However, inclusion of Tm did improve the solubility predictions for SBO ($R^2$ of 0.90, $Q^2$ of 0.83; Fig. 6b) and Captex355 ($R^2$ of 0.88, $Q^2$ of 0.83; Fig. 7b).

The finding that Tm together with just a few other molecular descriptors was sufficient for the prediction of solubility in SBO and Captex355 led to an investigation whether a General Solubility Equation applicable to lipids ($GSE_{Lipid}$) was possible based on the current dataset. This resulted in the following equations based on Tm, number of nitrogen atoms (nN) and number of double bonds (nDB):

$$\log S_{SBO} = -0.19-0.01Tm-0.26nN-0.21nDB \quad (1)$$

with an F value of 28.3 and p value of $1.24\times10^{-6}$, the results are presented in Fig. 6c.

$$\log S_{Captex355} = -0.15-0.01Tm-0.27nN-0.18nDB \quad (2)$$

with an F value of 26.7 and p value of $1.81\times10^{-6}$, the results are presented in Fig. 7c.

The results from the multilinear regression did not improve by exchanging Tm with $\Delta S_f$ alone but the value obtained after correcting for the Tm (*i.e.* $\Delta S_f$ (Tm-25)/1364) gave slightly better accuracy in the predictions. The latter resulted in $R^2$ of

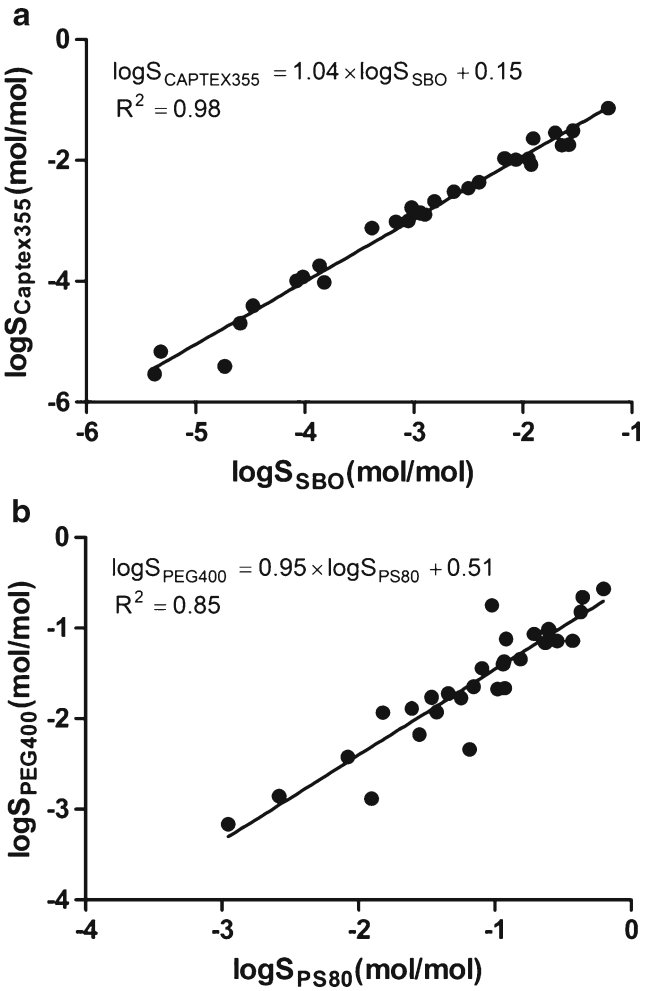


**Fig. 5** Relation between solubility in different excipients. (**a**) A strong correlation was observed between the solubility in SBO and Captex355 resulting in an $R^2$ of 0.98. (**b**) A strong relationship between the solubility determined in PS80 and PEG400 was also identified resulting in an $R^2$ of 0.85. The latter indicates that the ethoxylation is the main determinant for the solvation capacity of PS80.

**Table II** Results from the Model Development of Drug Solubility in SBO

| | $R^2$ | $Q^2$ | $RMSE_{Tr}$ | $RMSE_{Te}$ | Variables |
|---|---|---|---|---|---|
| PLS descriptors | 0.81 | 0.76 | 0.52 | 0.25 ($n=5$) | TPSA(NO), DECC, MOR21v, MATS6m, DP06 |
| PLS descriptors + Tm | 0.90 | 0.83 | 0.44 | 0.35 ($n=4$) | TPSA(NO), Tm, DECC, TI2, MATS6m, MOR21v |
| MLR | 0.84 | – | 0.45 | 0.48 ($n=4$) | Tm, nN, nDB |

The following abbreviations are used: training set (Tr), test set (Te), partial least square projection to latent structures (PLS), multilinear regression (MLR), total polar surface area of N and O atoms (TPSA(NO)), eccentric (DECC), the 3D MoRSE signal 21 weigthed for van der Waals volume (MOR21v), the Moran autocorrelation lag 6 weighted by atomic masses (MATS6m), Randic global molecular 3D profile number 6 (DP06), melting point (Tm), the second Mohar index (TI2), number of nitrogens (nN) and number of double bonds (nDB)

0.94 and $RMSE_{Te}$ of 0.75 in SBO and $R^2$ of 0.94 and $RMSE_{Te}$ of 0.85 in Captex355.

## DISCUSSION

Drug solubility in lipid based formulations is a critical determinant of utility as it defines the maximum possible dose that can be administered, without resorting to two-phase suspension formulations. Measurement of drug solubility, however, is both compound and time dependent and the development of LBFs would therefore be more cost and time efficient if solubility in the formulation could be predicted from molecular structure or based on minimal experimental effort. Recently it was proposed that the maximal drug concentration that could be obtained in a formulation is possible to calculate by summing the amounts that can be dissolved in each excipient fraction in the final formulation (24). Even though this process requires experimental determination of drug solubility in each excipient, the total number of solubility measurements that must be undertaken is reduced since the potential drug loading of a wide range of excipient combinations can subsequently be forecasted.

In the current work we sought to extend this concept further either by predicting solubility *de novo* from drug physicochemical properties, or to predicting drug solubility in one excipient from data obtained in another. With respect to the latter suggestion, during the experimental study we found a strong correlation between the solvation capacities of a wide range of drugs in SBO and Captex355 ($R^2$ of 0.98). We propose that this correlation can be used to calculate the maximum solubility in *e.g.* SBO from determinations in Captex355 and hence reduce the experimental screening efforts. This has additional practical ramifications as the measurement of drug solubility in the $TG_{MC}$ is somewhat easier than it is in the more viscous $TG_{LC}$. This trend, *i.e.* that solubility, when measured in mg/g, decrease with increasing fatty acid chain length has previously been observed for a small number of compounds (36). Further, a close relationship between the solvation capacities of $TG_{LC}$ and $TG_{MC}$ has been suggested based on data obtained for a limited compound set of five lipophilic drugs (37). The current work has confirmed these initial trends and expanded the relationship to a significantly larger and more structurally diverse dataset. The equal solvation capacity of $TG_{LC}$ and $TG_{MC}$ has previously been proposed to reflect the equal concentrations of ester function per mol of the vehicle, rather than, for example, the length of the fatty acid constituents of the glycerides (38,39). A strong correlation was also obtained between the solvation capacities of PS80 and PEG400 ($R^2$ of 0.85), but not between either of the glycerides and either of the ethoxylated materials. This indicates that the solubility determinants for PS80 and PEG400 are related, but different to that of the triglycerides and infers a specific contribution of ethoxylation to the solvation capacity of PS80 and PEG400. The experimental data reported here therefore suggests that drug solubility in one triglyceride may be predicted from another, and that solubility in an ethoxylated excipient may be predicted from solubility in *e.g.* PEG400. This finding has the potential to

**Table III** Results from the Model Development of Drug Solubility in Captex355

| | $R^2$ | $Q^2$ | $RMSE_{Tr}$ | $RMSE_{Te}$ | Variables |
|---|---|---|---|---|---|
| PLS descriptors | 0.84 | 0.80 | 0.47 | 0.73 ($n=9$) | TPSA(NO), ICR, JGI6, Mor21v |
| PLS descriptors + Tm | 0.88 | 0.83 | 0.41 | 0.75 ($n=8$) | TPSA(NO),Tm, Mor18m, nN, GATS7m |
| MLR | 0.83 | – | 0.45 | 0.99 ($n=8$) | Tm, nN, nDB |

The following abbreviations are used: training set (Tr), test set (Te), partial least square projection to latent structures (PLS), multilinear regression (MLR), total polar surface area of N and O atoms (TPSA(NO)), radial centric information index (ICR), mean topological charge index of order 6 (JGI6), the 3D MoRSE signal 18 weigthed for atomic masses (MOR18m), number of nitrogens (nN),Geary autocorrelation lag 7 weighted by atomic masses (GATS7m) and number of double bonds (nDB)

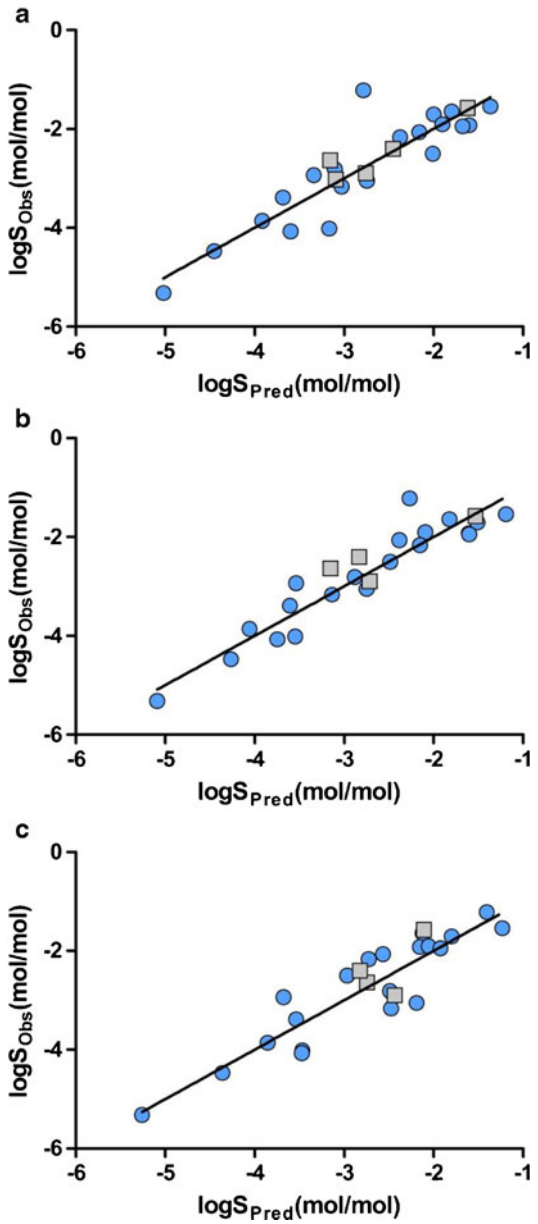


**Fig. 6** PLS and MLR models obtained for SBO solubility. (**a**) PLS model obtained based on molecular descriptors only. The following calculated descriptors, TPSA(NO), DECC, MOR21v, MATS6m and DPO6 resulted in $R^2$ of 0.81 and $RMSE_{te}$ of 0.25. (**b**) Including the experimentally determined melting point (Tm) into the variable selection produced a PLS model with $R^2$ of 0.90 and $RMSE_{te}$ of 0.35. In the final model, six variables TPSA(NO), Tm, DECC, TI2, MATS6m and MOR21v were included. (**c**) MLR based on Tm, number of nitrogens and number of double bonds resulted in $R^2$ of 0.84 and $RMSE_{te}$ of 0.48. In figures (**a**–**c**) the training set is shown as light blue circles and test set as grey squares.

reduce the number of solubility experiments that must be performed to profile drug solubility across a series of non-aqueous vehicles.

The main aim of the current work was to investigate if drug solubility in excipients commonly used in LBFs could be predicted solely from physicochemical properties, and in particular whether calculated molecular descriptors could be used for this purpose. Our modeling efforts resulted in excellent models for SBO and Captex355, and suggest that for simple triglyceride lipids this may indeed be true. In contrast, we were less successful in obtaining models for PS80 and PEG400. The reasons for the differences in behavior of these excipients are unknown and are likely to be multifactorial. We speculated that two of these reasons were the relatively small solubility interval of the 30 compounds in PS80 and PEG400 and the potential for water to have a greater complicating effect on solubility prediction in highly ethoxylated excipients. The solubility interval for the compound set in PS80 and PEG400 was ~100 times (*i.e.* 2 $\log_{10}$ units) whereas the same series of compounds had >10,000 times (*i.e.* 4 $\log_{10}$ units) difference in solubility in SBO and Captex355. The smaller solubility interval makes the modeling more demanding since the large chemical diversity of the compounds is not reflected by a large variation in PEG400 and PS80 solubility. The potential for water to complicate solubility assessment was well known to us in initiating the project, but in the first instance the excipients were not used in their dry state. The main reason for this was that more valuable predictive data was expected to be obtained for drug solubility in 'off the shelf' excipients, since these materials typically are used under ambient (non-dry) conditions during the formulation stage. Additionally, the practical utility of the data obtained would also be limited since water will be present during formulation processing, capsule filling *etc*. However, after measuring the water content in PEG400 and also performing solubility studies in PEG400 under dry conditions it became clear that the water sorption in this study was minimal and therefore unlikely to have had a significant effect on the solubility values determined. The PEG400 used in the original study contained 0.32% w/w water compared to the freshly opened container which contained 0.11% w/w water. Solubility determinations of itraconazol, danazol and indomethacin confirmed the first results, no significant difference were seen between ambient or dry conditions. The solubility of candesartan was of the same magnitude for both determinations, 13.8± 1.5 mg/g in ambient milieu compared to 16.0±0.6 mg/g in dry milieu ($p$=0.042). Candesartan was one of the last drugs to be determined for solubility in our compound series and hence is an exemplar of a 'worst case' effect of the water sorption in the ethoxylated vehicles. Based on the measurements we argue that the difficulties in developing quantitative models for PEG400 and PS80 were not results of the water sorption.

Only a few molecular descriptors were needed to successfully predict drug solubility in the triglycerides. For SBO, these were PSA, eccentric (DECC), the 3D MoRSE signal 21 weighted for van der Waals volume (MOR21v), the Moran autocorrelation lag 6 weighted by atomic masses (MATS6m) and Randic global molecular 3D profile number 6 (DP06).

Not all of these descriptors are easy to understand. However, the DECC, MOR21v and DP06 are related to the shape of 2D or 3D structures, whereas MATS6m rather describes the distribution of the atomic masses in a molecule (based on the 2D structure). The model shows that, in general, all these properties negatively impact on lipid solubility. The MOR21v descriptor has negative values for all drugs and hence, the lower MOR21v descriptor the higher the SBO solubility. The MATS6m range from positive to negative values and as such, drugs with negative MATS6m values will have higher SBO solubility than those with positive values. Overall, the descriptors found to be of importance to drug solubility in SBO are related to polarity, shape, size and atomic mass distribution. Several shape factors were weighted for size and atomic mass, complicating interpretation, yet the negative impact of the DECC and DP06 descriptors indicates that the more elongated the molecule is, the lower the solubility. This effect is likely coupled to the need to form larger cavities in the SBO in order to solubilize an oval/elongated molecular structure when compared to that required for more spherical molecules. As expected, the model also shows that the larger the PSA, the lower the solubility in $TG_{LC}$. For Captex, PSA and MOR21v remained important, and in addition to these the radial centric information index (ICR) and the mean topological charge index of order 6 (JGI6) proved important for the drug solubility. ICR reflects eccentricity but at the atom level (rather than DECC which captures this information at the molecule level). JGI6 gives information on the topological charge distribution. JGI6 positively contributes to drug solubility in Captex355, whereas ICR contributes negative. Taken all together the results from the Captex355 model suggests that similar properties (polarity, size and shape) are of importance for solubility in $TG_{MC}$ as in $TG_{LC}$ but also informs that molecules with low topological charge index are poorly solubilized in $TG_{MC}$.

The addition of Tm as a descriptor in the variable matrix resulted in slightly different models with improved accuracy for prediction. Based on our compound series we found that an increase in Tm negatively impacted on the solubility in both SBO and Captex355, highlighting the impact of the crystal lattice also on solubility of drugs in lipids. Tm has been identified previously as an important indicator of drug solubility in lipids using a homologous series of compounds (34), however, in a recent study based on a small dataset ($n=10$) of structurally diverse drug molecules no clear link between Tm and solubility in triglycerides was found (16). Interestingly we did not find Tm to significantly influence drug solubility in PS80 and PEG400. This was surprising as Tm has been suggested previously to be important in attempts to model drug solubility in PEG400 (40,41). The latter studies examined the solubility of a dataset of 92 molecules in PEG400 resulting in solubility differences of more than 50,000-fold. In that study regression analysis of predicted solubility based on

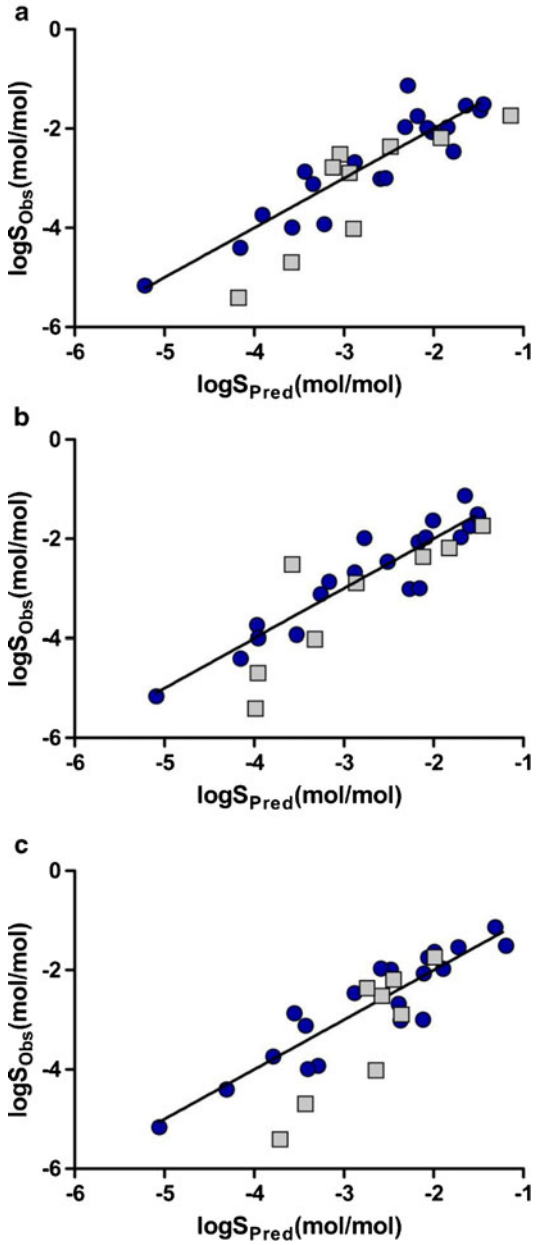


**Fig. 7** PLS and MLR models obtained for Captex355 solubility. (**a**) PLS model obtained based on calculated molecular descriptors only. The following calculated descriptors, TPSA(NO), ICR, JGI6 and MOR21v resulted in $R^2$ of 0.84 and $RMSE_{te}$ of 0.73. (**b**) Including the experimentally determined melting point (Tm) into the variable selection produced a PLS model with $R^2$ of 0.88 and $RMSE_{te}$ of 0.75. In the final model five variables, TPSA(NO), Tm, MOR18 m, nN and GATS7m were included. (**c**) MLR based on Tm, number of nitrogens and number of double bonds resulted in $R^2$ of 0.83 and $RMSE_{te}$ of 0.99. In figures (**a**–**c**) the training set is shown as dark blue circles and test set as grey squares.

Tm *versus* observed PEG400 solubility resulted in models with $R^2$ of 0.71 and $RMSE_{Tr}$ of 0.55. Our PEG400 model gave $R^2$ values of 0.62 and $RMSE_{Tr}$ of 0.44. The rather modest statistics of both of these modeling efforts for drug solubility in PEG400, regardless of the dataset, modeling technique or solubility interval used for the prediction, suggests that

solvents containing high fraction of ethoxylated chains are difficult to predict. We speculate that Molecular Dynamics simulations of ethoxylated excipients *e.g.* PEG400 and PS80 may facilitate improved understanding of the solvation capacity of such excipients.

The finding that Tm improved the predictions of drug solubility in lipids, did allow the development of a $GSE_{Lipid}$ starting with the properties that were found to be most important during the PLS model development. Here we included Tm as a property describing the dissociation of the molecule from the crystal lattice. We also added calculated descriptors that were identified as important during model development and which were believed to reflect the ability of the drug to interact with the lipid. In this step descriptors that were easily calculated either manually (*e.g.* nDB, nHacc, nN) or through standard and non-expensive software (*e.g.* PSA) were included in order to make the $GSE_{Lipid}$ widely applicable to laboratories where more advanced software and calculation programs are not available. This strategy produced good models based on Tm and calculated molecular properties (Tables II and III). Exchanging Tm with $\Delta S_f$ alone did not result in successful solubility predictions but when it was combined with Tm slightly better accuracy was obtained. However, the scope of the MLR was to create a generally applicable equation based on easily calculated and/or measured properties. For this purpose Tm is the preferred variable since this property can be determined with methodologies, *e.g.* capillary melting point apparatus, were $\Delta S_f$ is not generated. Further, when new formulations are explored during *e.g.* generic product development and/or life cycle management, the Tm is easily found for most molecules whereas it is difficult to find *e.g.* literature data on $\Delta S_f$. Several versions of GSE have been published to date but to the best of our knowledge this is the first contribution in which the GSE has been adapted to triglycerides and where the focus has been on predictions across a structurally diverse dataset of drugs.

## CONCLUSION

In this work we have shown that solubility predictions in excipients can be rationalized at several levels. Experimentally the solubility in one vehicle may be possible to predict from the solubility in another, in this study exemplified with the strong correlation observed for SBO and Captex355, and PS80 and PEG400. This reduces the experimental screening efforts and simultaneously reduces the amount of material used during excipient screening. It was also shown, for the first time that drug solubility in triglycerides (SBO and Captex355) may be predicted with high accuracy from calculated molecular descriptors. Molecular properties with strong impact on the resulting solubility in these vehicles were related to PSA, the number of nitrogen atoms, the number of double bonds, eccentricity, topological charge, size and shape. The PLS predictions were further improved when Tm was included as a descriptor, and general solubility equations based on Tm, the number of double bonds and the number of nitrogens resulted in accurate calculations of SBO and Captex355 solubility.

However, solubility in the ethoxylated excipients (PS80 and PEG400) was not possible to quantitatively predict from molecular structure and the inclusion of Tm in the modeling efforts did not improve the predictions. The difficulty to predict drug solubility in PS80 and PEG400 may be associated with the rather small solubility interval obtained in these vehicles, which from a modeling perspective results in data that is more demanding to predict. To obtain predictive models for the ethoxylated excipients other modeling techniques than linear regression approaches *e.g.* neural networks, vector machine or random forest may improve the predictions and expand the mechanistic understanding of such systems.

## ACKNOWLEDGMENTS AND DISCLOSURES

We thank Dr. Joakim Bjerketorp at the Department of Microbiology, Swedish University of Agricultural Sciences, for kindly helping us with the Karl Fischer titrations and Anna Skogh at Department of Medicinal Chemistry, Uppsala University for lending us the capillary melting point apparatus. We also thank the Swedish Research Council (Grants 621-2008-3777 and 621-2011-2445) and the Australian National Health and Medical Research Council for financial support to the project. C.A.S.B. is grateful to The Swedish Agency for Innovation Systems (Grant 2010–00966) and to Monash Institute of Pharmaceutical Sciences for financially supporting her Marie Curie fellowship at Monash University.



## REFERENCES

1. 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):3–25.
2. Rane SS, Anderson BD. What determines drug solubility in lipid vehicles: is it predictable? Adv Drug Deliv Rev. 2008;60(6):638–56.
3. Benet LZ, Wu C-Y, Custodio JM. Predicting drug absorption and the effects of food on oral bioavailability. Bull Tech Gattefossé. 2006;99:9–16.

4. Zheng W, Jain A, Papoutsakis D, Dannenfelser RM, Panicucci R, Garad S. Selection of oral bioavailability enhancing formulations during drug discovery. Drug Dev Ind Pharm. 2011;00:1–13.
5. Mansky P, Dai WG, Li S, Pollock–Dove C, Daehne K, Dong L, *et al.* Screening method to identify preclinical liquid and semi–solid formulations for low solubility compounds: miniaturization and automation of solvent casting and dissolution testing. J Pharm Sci. 2006;96(6):1548–63.
6. Dai WG, Dong LC, Li S, Pollock-Dove C, Chen J, Mansky P, *et al.* Parallel screening approach to identify solubility-enhancing formulations for improved bioavailability of a poorly water-soluble compound using milligram quantities of material. Int J Pharm. 2007;336(1):1–11.
7. Wyttenbach N, Alsenz J, Grassmann O. Miniaturized assay for solubility and residual solid screening (SORESOS) in early drug development. Pharm Res. 2007;24(5):888–98.
8. 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.
9. Poulin P, Jones R, Jones HM, Gibson CR, Rowland M, Chien JY, *et al.* PHRMA CPCDC initiative on predictive models of human pharmacokinetics, part 5: prediction of plasma concentration–time profiles in human by using the physiologically–based pharmacokinetic modeling approach. J Pharm Sci. 2011;100(10):4127–57.
10. Hauss DJ. Oral lipid-based formulations. Adv Drug Deliv Rev. 2007;59(7):667–76.
11. Porter CJH, Pouton CW, Cuine JF, Charman WN. Enhancing intestinal drug solubilisation using lipid-based delivery systems. Adv Drug Deliv Rev. 2008;60(6):673–91.
12. 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(Supplement 2(0)):S93–8.
13. 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.
14. Williams HD, Anby MU, Sassene P, Kleberg K, Bakala N'Goma JC, Calderone M, *et al.* Toward the establishment of standardized in vitro tests for lipid-based formulations, Part 2: the effect of bile salt concentration and drug loading on the performance of Type I, II, IIIA, IIIB and IV formulations during in vitro digestion. Mol Pharm. 2012;9(11):3286–300.
15. Williams HD, Sassene P, Kleberg K, Bakala-N'Goma JC, Calderone M, Jannin V, *et al.* Toward the establishment of standardized in vitro tests for lipid–based formulations, part 1: method parameterization and comparison of in vitro digestion profiles across a range of representative formulations. J Pharm Sci. 2012;101(9):3360–80.
16. Thi TD, Van Speybroeck M, Barillaro V, Martens J, Annaert P, Augustijns P, *et al.* Formulate-ability of ten compounds with different physicochemical profiles in SMEDDS. Eur J Pharm Sci. 2009;38(5):479–88.
17. Anderson BD, Rytting JH, Higuchi T. Solubility of polar organic solutes in nonaqueous systems: role of specific interactions. J Pharm Sci. 1980;69(6):676–80.
18. Rane S, Cao Y, Anderson B. Quantitative solubility relationships and the effect of water uptake in triglyceride/monoglyceride microemulsions. Pharm Res. 2008;25(5):1158–74.
19. Rane SS, Anderson BD. Molecular dynamics simulations of functional group effects on solvation thermodynamics of model solutes in decane and tricaprylin. Mol Pharm. 2008;5(6):1023–36.
20. Kasimova AO, Pavan GM, Danani A, Mondon K, Cristiani A, Scapozza L, *et al.* Validation of a novel molecular dynamics simulation approach for lipophilic drug incorporation into polymer micelles. The J Phys Chem B. 2012;116(14):4338–45.
21. Warren DB, Chalmers DK, Pouton CW. Structure and dynamics of glyceride lipid formulations, with propylene glycol and water. Mol Pharm. 2009;6(2):604–14.
22. Huynh L, Grant J, Leroux J-C, Delmas P, Allen C. Predicting the solubility of the anti-cancer agent docetaxel in small molecule excipients using computational methods. Pharm Res. 2008; 25(1):147–57.
23. Prajapati HN, Dalrymple DM, Serajuddin ATM. A comparative evaluation of mono-, di-and triglyceride of medium chain fatty acids by lipid/surfactant/water phase diagram, solubility determination and dispersion testing for application in pharmaceutical dosage form development. Pharm Res. 2012;29(1):285–305.
24. Sacchetti M, Nejati E. Prediction of drug solubility in lipid mixtures from the individual ingredients. AAPS PharmSciTech. 2012:1–7.
25. Yalkowsky SH, Valvani SC. Solubility and partitioning I: solubility of nonelectrolytes in water. J Pharm Sci. 1980;69(8):912–22.
26. Bergström CAS, Wassvik CM, Norinder U, Luthman K, Artursson P. Global and local computational models for aqueous solubility prediction of drug-like molecules. J Chem Inf Comput Sci. 2004;44(4):1477–88.
27. Johnson SR, Zheng W. Recent progress in the computational prediction of aqueous solubility and absorption. AAPS J. 2006;8(1):27–40.
28. Persson EM, Nordgren A, Forsell P, Knutson L, Öhgren C, Forssén S, *et al.* Improved understanding of the effect of food on drug absorption and bioavailability for lipophilic compounds using an intestinal pig perfusion model. Eur J Pharm Sci. 2008;34(1):22–9.
29. Bergström CAS, Wassvik CM, Johansson K, Hubatsch I. Poorly soluble marketed drugs display solvation limited solubility. J Med Chem. 2007;50(23):5858–62.
30. Biswas A, Sharma BK, Willett J, Advaryu A, Erhan S, Cheng H. Azide derivatives of soybean oil and fatty esters. J Agric Food Chem. 2008;56(14):5611–6.
31. Kossena GA, Boyd BJ, Porter CJH, Charman WN. Separation and characterization of the colloidal phases produced on digestion of common formulation lipids and assessment of their impact on the apparent solubility of selected poorly water-soluble drugs. J Pharm Sci. 2003;92(3):634–48.
32. Bergström CAS, Charman SA, Nicolazzo JA. Computational prediction of CNS drug exposure based on a novel in vivo dataset. Pharm Res. 2012:1–12.
33. Fagerberg JH, Al-Tikriti Y, Ragnarsson G, Bergström CAS. Ethanol effects on apparent solubility of poorly soluble drugs in simulated intestinal fluid. Mol Pharm. 2012;9(7):1942–52.
34. Larsen DB, Parshad H, Fredholt K, Larsen C. Characteristics of drug substances in oily solutions. Drug release rate, partitioning and solubility. Int J Pharm. 2002;232(1):107–17.
35. Smith A, Heckelman PE, Budavari S, editors. The Merck index: an encyclopedia of chemicals, drugs, and biologicals. New Jersey: Merck and Co., Inc.; 2001.
36. Prajapati HN, Patel DP, Patel NG, Dalrymple DD, Serajuddin A. Effect of difference in fatty acid chain lengths of medium-chain lipids on lipid-surfactant-water phase diagrams and drug solubility. J Excipients Food Chem. 2011;2(3):73–88.
37. Kaukonen AM, Boyd BJ, Porter CJH, Charman WN. Drug solubilization behavior during in vitro digestion of simple triglyceride lipid solution formulations. Pharm Res. 2004;21(2):245–53.
38. Anderson BD, Marra MT. Chemical and related factors controlling lipid solubility. Bull Tech Gattefosse'. 1999;92:11–9.
39. 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.
40. Ghafourian T, Bozorgi AHA. Estimation of drug solubility in water, PEG 400 and their binary mixtures using the molecular structures of solutes. Eur J Pharm Sci. 2010;40(5):430–40.
41. Rytting E, Lentz KA, Chen XQ, Qian F, Venkatesh S. Aqueous and cosolvent solubility data for drug-like organic compounds. AAPS J. 2005;7(1):78–105.