Comparison with quantitative and qualitative assessments of cosolvent models for predicting the solubility of various drug and drug-like compounds in aqueous cosolvent systems.

Abstract

The performance of six cosolvent solubility models were assessed for 12 aqueous cosolvent systems of 11 drug and drug-like solutes which were collected from literature. The cosolvents selected for this study were: 2-propanol, glycerol, and polyethylene glycol 400 (PEG-400). From the error calculations, the models were ranked as following in order of increase error/inaccuracy: Jouyban-Acree, GSM, UNIQUAC, NRTL, log-linear, and predictive log-linear. The actual comparative performance was found to be more complex, with the Jouyban-Acree and GSM models frequently overfitting to the experimental data. A qualitative review of the modelling process conducted in this study is then given to provide readers with guidance for carrying out their own cosolvency modelling activities. This summarises as follows: the Jouyban-Acre and GSM models being the most generally applicable for most use-cases, with the added work required to operate UNIQUAC and NRTL often failing to yield any benefits, and finally that the log-linear models are best suited to scenarios with limited data availability or a priori applications where accuracy is not a priority.

1 Introduction

Cosolvency is the process of adding one or more solvents to a solution to control (typically increase) the solubility of a particular solute [1]. This technique has applications in many fields, from industrial crystallisation processes to pharmaceutical vehicle design [2]. For several decades the effects of cosolvency have been known, studied, and subsequently modelled. Early work began with Yalkowsky and Roseman publishing their findings and developing the log-linear model for predicting solubility [3]. In the following decades further research has been carried out to develop more solubility models, notably by A. Jouyban and W.E. Acree with the Jouyban-Acree model [4]. Other models have been developed to yield more accurate predictions whilst reducing the amount of data required to produce them [2][5][6]. Subsequently with a great number of cosolvency models available, multiple studies and reviews have been carried out to compare the different solubility models [2][6][7][8][9].

The work included in this study is being conducted as part of a research group interested in transdermal permeation modelling. Solubility is one of the key aspects of modelling in this area [10], hence this study has been carried out to investigate and compare different solubility models to report back with holistic modelling recommendations. As such, the models and systems included in this study will be narrower in scope compared to existing cosolvent model comparisons, however, there will be a particular focus on conveying the modelling methods used and to discuss the real decisions and considerations a newcomer to solubility modelling will have when selecting a model for their activities.

To assess the models, a mixture of drug and drug-like solutes in three common cosolvents used with water have been selected (2-proanol, glycerol, and PEG-400), totaling to 12 systems in

all, which represent typical systems that may be encountered with pharmaceutical solubility modelling.

Six commonly used cosolvency models have then been implemented to predict the solubility measurements of each system: the log-linear model, predictive log-linear, Jouyban-Acree model, General Single Model, NRTL model, and UNIQUAC model. These predictions were then assessed using statistical methods to compare their accuracy. All of the modelling work in this study was carried out using MATLAB 2020b, however, the methods outlined in this paper can also be followed using free alternatives such as R or Python.

2 Data Selection, Models, and Methods

2.1 Solvent Data Collection

A key consideration for the solubility data selection was the relevancy of each system to the skin research group's applications. This meant that the systems needed to include a drug or drug-like compound in a cosolvent system which is similar to the primary components commonly used in pharmaceutical vehicle design, as well being measured at a temperature that are not too far off skin surface temperature which is around 30ºC [10].

Water + cosolvent mixtures were selected because extensive literature on aqueous cosolvency data and subsequent modelling has been published to build upon. A. Jouyban's '*Handbook of Solubility Data for Pharmaceuticals'* – glycerol, 2-Propanol, and PEG-400 are identified as commonly used aqueous solubility enhancers used in pharmaceutical applications. Glycerol and 2-propanol were selected for their relevancy for the research group's current work, and PEG-400 was selected to contrast glycerol and 2-propanol in terms of molecular weight and hydrophobicity.

All collected solubility data included in this study were only used if each of the solubility measurements were taken at least in duplicate. Also, for certain model compatibilities; that the solute-free solvent fractions were reported for each measurement, and the data itself included neat solvent and cosolvent solubility measurements. The selected 12 solubility systems and corresponding system temperature and number of measured data points can be found in Table 1.

For the activity coefficient models included in this study, each of the solubility data-sets were converted into a series of ternary mole-fractions of solvent, cosolvent, and solute. The exact procedure to do this conversion was dependent on how the solubility data and solvent fractions were reported in their sources. To provide an example of such a procedure, the calculation steps for the Phenytoin and PEG-400 system was done is will be explained. In the original paper [11] the fractions of water and PEG-400 were reported as solute free volume fractions, with solubility measured as [g/L]. The basis of all the calculation procedures was to interpret the solubility measurements in a unit volume or unit mass, so for this example 1 liter, and calculate a figurate number of moles from there.

1) The molar volume $\bar{v}$ [mol/L] for both water and PEG-400 was calculated by dividing the density ρ [kg/m^3] of each solvent at standard conditions (25°C & 1 atm) by the molar mass m_w [g/mol].
2) The amount of moles in one unit volume (or mass for other systems) of solvents, which in this example is 1 liter, is calculated by multiplying the volume fraction of each solvent by its respective molar volume.

3) Next, the solute solubility would be converted to moles using a similar approach. A measurement of [g/L] would be divided by the molar mass of the solute to determine the moles of solute in the unit volume of solution.
4) With the number of moles for solvent, cosolvent, and solute obtained, the mole fractions of each as a ternary system could easily be calculated from: $x_i = \frac{n_i}{n_p+n_c+n_s}$

Table 1 Solute and Cosolvent of each studied system, with the number of data points (N) and the temperature of each system (T).

System	Solute	Cosolvent	N	T [C]	Reference
1	Sodium Diclofenac	2-Propanol	13	25	[17]
2	Mesalazine	2-Propanol	11	25	[28]
3	Alanine	2-Propanol	8	25	[29][a]
4	Cefazolin Sodium Pentahydrate	2-Propanol	11	25	[30][a]
5	Isoleucine (L)	2-Propanol	9	25	[29][a]
6	Phenyl Alanine (L)	2-Propanol	11	25	[29][a]
7	Acetaminophen	PEG-400	11	30	[31][a]
8	Benzoyl Peroxide	PEG-400	11	25	[32][a]
9	Phenytoin	PEG-400	11	25	[11][a]
10	Valdecoxib (S)	PEG-400	6	25	[33][a]
11	Valdecoxib (S)	Glycerol	5	25	[33][a]
12	Rofecoxib (S)	Glycerol	6	25	[34][a]

[a] *Solubility data obtained from referencing A. Jouyban's 'Handbook of Solubility Data for Pharmaceuticals'*

To re-iterate, this is an example calculation procedure and the calculation steps were dependent on how the data was originally reported. For data sets which were sourced from *A.* Jouyban's *'Handbook of Solubility Data for Pharmaceuticals'*, the original published paper for each system was checked to reduce the risk of error in interpreting the presented units of measurements.

2.2 Cosolvency Models

The models chosen for this study were selected based on their frequent use in solubility studies in the last 20 years, indicating that these models would be six of the first models someone would consider from choosing between. Additionally, the selection reflects a variety of theoretical approaches to solubility modelling, as a well as a range of applications based on the data available to the modeler. All of the models, with the exception of the predictive log-linear model, are used in a correlative manner.

2.2.1 Log-Linear Model

The log-linear model calculates solubility by assuming that the log of molar solubility ($\log_{10} X_M$) of a solute has a linear relationship between the logs of the solute's solubility in neat primary solvent (X_P) and cosolvent (X_C) as weighted by the solvent-free fractions of

primary solvent (f_P) and cosolvent (f_C) [2]. The solvent & cosolvent fraction terms of this model can be interpreted as: weight, volume, or molar fractions, in this study they have been used depending on how the data is originally presented for each system.

$$\log_{10} X_M = f_P \log_{10} X_P + f_C \log_{10} X_C \qquad (1)$$

This model is favored for its simplicity and ease of use [2][6]. Only two data-points are required for use which do not require any regression. This suits the log-linear model to a-priori activities, or for cases where limited solubility data is available.

2.2.2 Predictive Log-Linear Model

This extended form of the log-linear model replaces the solubility in neat cosolvent with a solubilization term (σ) calculated from the solute octanol/water partition coefficient ($\log_{10} K_{OW}$) and cosolvent parameters calculated from regression (s and t) [1]. The other terms in the model have the same meaning as with the ideal mixing log-linear model.

$$\log_{10} X_M = \log_{10} X_S + \sigma f_C \qquad (2)$$

$$\sigma = s \cdot \log_{10} K_{OW} + t \qquad (3)$$

The cosolvent parameters of s and t for many cosolvents are widely available in literature. Furthermore, the solute's partition coefficient can be taken from experimental data or be calculated by software.

2.2.3 Jouyban-Acree Model

The Jouyban-Acree model is a popular correlative cosolvency model which introduces additional correlative terms based on thermodynamic two-body and three-body interactions [4]. In this study the most simplest form of the model with isothermal correlative parameters (S_i) has been used.

$$\ln X_M = f_P \ln X_P + f_C \ln X_C + f_P f_C \sum_{i=0}^{2} S_i (f_P - f_C)^i \qquad (4)$$

In this equation the remaining terms have the same meaning as with the previous models. This model is more suitable for systems which deviate from ideal mixing rules for solubility, and when various data points are available, however there is still a reliance on neat solvent data points.

2.2.4 General Single Model

The general single model is a polynomial power series (presented below alongside its general form) that is derived from the Jouyban-Acree and Excess Energies cosolvency models [2][12]. The model is applied by regressing solubility data to obtain a series of correlative terms of $B_i^{(i)}$.

$$\ln X_M = \sum_{i=0}^{m=4} B_i \cdot f_p^i = B_0 + f_p B_1 + f_p^2 B_2 + f_p^3 B_3 + f_p^4 B_4 \qquad (5)$$

The biggest advantage of the GSM over the Jouyban-Acree model is that it does not require neat solvent/cosolvent data, this is useful for circumstances where a full set of solubility data points is not available.

2.2.5 NRTL Model

The non-random two liquid or NRTL model proposed by Prausnitz and Renon in 1968 derived from local composition theory that predicts the activity coefficients of components in miscible and partially immiscible systems [13]. In this study, the NRTL model is paired with the equation for real-solubility from activity coefficient shown below [6].

$$\ln X_M = -\ln \gamma_s - \left[\frac{\Delta_{fus}H}{R} \cdot \left(\frac{1}{T_{fus}} - \frac{1}{T}\right)\right] \quad (6)$$

In this equation mole fraction solubility X_M is calculated from: the solute enthalpy and temperature of fusion – $\Delta_{fus}H$ [J/mol] and T_{fus} [K], the temperature of the solution T [K], the universal gas constant R [J/mol K], and the solute's activity coefficient γ_s.

The NRTL model itself is represented by the following equation:

$$\ln \gamma_i = \frac{\sum_j \tau_{ji} G_{ji} x_j}{\sum_k G_{ki} x_k} + \sum_j \frac{x_j G_{ij}}{G_{ki} x_k}\left(\tau_{ij} - \frac{\sum_n x_n \tau_{nj} G_{nj}}{\sum_k G_{kj} x_k}\right) \quad (7)$$

In this equation the mole fraction of each component is represented by x_j. The parameter G_{ij} is calculated from the expression: $G_{ij} = \exp(-\alpha_{ij} \cdot \tau_{ij})$, α_{ij} is the non-randomness parameter, and τ_{ij} ($\tau_{ii} = \tau_{jj} = 0$) is the binary interaction parameter. Where τ_{ij} for each component pair of i & j is calculated from correlating temperature dependent energy interaction parameters (g_{ij}) or from a series of co-efficient parameters. In this study τ_{ij} was directly correlated from experimental data.

2.2.6 UNIQUAC Model

For the Universal Quasi-Chemical Theory model (or UNIQUAC model) the same equation was used to calculate mole fraction solubility from the solute activity coefficient as with the NRTL model. The UNIQUAC model calculates the activity coefficient of each component γ_i from the contribution of combinational γ_i^C and residual γ_i^R activities [6] [14].

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \quad (8)$$

The combinational term accounts for the effects the difference in the shape of each component in the mixture, and the residual term is a correction for the intermolecular forces in the mixture.

The combinational terms is calculated from the following equations:

$$\ln \gamma_i^C = \ln\left(\frac{\Phi_i}{x_1}\right) + \frac{z}{2} \cdot \ln\left(\frac{\theta_i}{\Phi_i}\right) + l_i - \frac{\Phi_i}{x_i} \cdot \sum_{j=1}^{m} x_j l_j \qquad \textbf{(9)}$$

$$l_i = \frac{Z}{2}(r_i - q_i) - (r_i - 1) \qquad \textbf{(10)}$$

For the residual term:

$$\ln\ \gamma_i^{\mathrm{R}} = q_i \left(1 - \ln\left(\sum_{j=1}^{m} \theta_i \tau_{ji}\right) - \sum_{j=1}^{m} \frac{\theta_i \tau_{ij}}{\sum_k^m\ \theta_k \tau_{kj}}\right) \qquad \textbf{(11)}$$

In equation 9 the parameter z is the molecular co-ordination number, which was set to 10. The parameter τ_{ij} is again the interaction parameter between molecules, is typically calculated from correlated temperature independent parameters of u_{ij} where $\ln \tau_{ij} = \exp\left(u_{ij}/RT\right)$. In this study, the τ_{ij} parameters were directly correlated from experimental data, which will be explained in the methodology.

Where θ_i and Φ_i are the terms for relative molecular surface area and relative molecular volume respectively, calculated from the molecular surface area and volume terms of r_i and q_i. With the UNIQUAC method r_i and q_i are calculated from the contribution of molecular groups in each component with R_g and Q_g which occur in each molecule v times. The values for R_g and Q_g can be obtained from direct calculation (such as using the Bondi method [15]), or from using existing published values.

$$\theta_i = \frac{q_i \cdot x_i}{\sum_j^m q_j \cdot x_j} \qquad \textbf{(12)} \qquad\qquad \Phi_i = \frac{r_i \cdot x_i}{\sum_j^m r_j \cdot x_j} \qquad \textbf{(13)}$$

$$q_i = \sum_g v_g^i \cdot Q_g \qquad \textbf{(14)} \qquad\qquad r_i = \sum_g v_g^i \cdot R_g \qquad \textbf{(15)}$$

2.3 Model Implementation

For this work all solubility models were implemented using MATLAB 2020b. Five of the six models (excluding UNIQUAC) were incorporated into a solubility modelling tool using the MATLAB app designer framework.

As discussed, this body of work is being carried out with the objective of applying one or more of the studied solubility models into transdermal mass transfer models. Such skin penetration models assume an isothermal nature at the skin surface. Therefore any selected solubility model for this purpose would not be used to predict solubility across a temperature range. This has impacted the model implementation by only selecting the isothermal forms of the included

models, and/or correlating the models exclusively on isothermal solubility data to obtain temperature dependent correlative parameters.

2.3.1 Predictive Log-Linear Model

The cosolvent parameters of s and t for the cosolvents included in this study were sourced from *'Handbook of Solubility Data for Pharmaceuticals' by* A. Jouyban. The solute partition co-efficient values for $\log_{10}K_{OW}$ were obtained from the Physical Sciences Data-Science Service (PSDS) *ACD/I-LABS* online database [16]. The database presents two values and the accuracy of each value, and the $\log_{10}K_{OW}$ value with the greatest accuracy was selected for use.

2.3.2 Jouyban-Acree model

To determine the S_i parameters of the Jouyban-Acree model the linear regression procedure outlined by A. Jouyban [17] was used. This involved regressing $\ln X_M - f_P \ln X_P + f_C \ln X_C$ against: $f_P f_C$, $f_P f_C(f_P - f_C)$, and $f_P f_C(f_P - f_C)^2$ for each experimental data point. The regression was performed using the MATLAB *regress* function.

2.3.3 General Single Model

The GSM $B_i^{(i)}$parameters were also obtained using the MATLAB *regress* function. For this model, the term $\ln X_M$ was correlated against: $0_{N,1}$ (N = the number of experimental data points), f_p, f_p^2, f_p^3, and f_p^4.

2.3.4 NRTL Model

Typically when correlating the NRTL model one of the several temperature independent & dependent parameters used to calculate the τ_{ij} interaction parameter are determined. However, for this study, with isothermal applications in mind. The values of τ_{ij} were determined directly for all combination of pairs for each solubility data set. As a result, the τ_{ij} values obtained from the method used in this study can only be used reliably to predict solubility data for the same solubility system under the exact same conditions. These values cannot assumed to be accurate binary interaction parameters (BIPs) for a single pair applied to a different system (e.g. solvent and cosolvent in vapor liquid equilibrium). Further study and correlation involving multiple solubility data sets would be required to determine more traditional and widely applicable BIP values for SLE systems. Additionally, by determining the values of τ_{ij} directly, as opposed to calculating any additional sub-parameters, accuracy of parameter estimation is increased by reducing the number of parameters.

None the less, to determine τ_{ij} values for each system a non-linear optimisation procedure to determine the BIPs for LLE systems was modified for this application of solubility determination [18]. The modified optimisation procedure applied is defined as follows:

$$\text{Objective Function:} \qquad F = \sum_{h=1}^{N} \left(\frac{X_{h,j}^{exp} - X_{h,j}^{pred}}{X_{m,h}^{exp}} \right)^2 \qquad \textbf{(16)}$$

$$s.t.$$

$$J_{NC} \cdot I_{NC} \cdot \tau = 0 \qquad \textbf{(17)}$$

The objective function represents the relative error between each experimental data point and the predicted solubility using the current set of τ_{ij} values generated by the optimisation solver. The term N refers to the number of experimental data points in the current solubility data set being studied. Relative error was used because during the development of model implementation it was found that this procedure yielded more accurate results for the example set of data used for testing. The single constraint ensures that the diagonal values of τ_{ij} are equal to zero. The optimisation solver employed was the *fmincon* function of MATLAB 2020b using default optimisation (*optimset)* options, all remaining constraints required by *fmincon* were set to null in the MATLAB code.

Due to the mole fraction solubility being the output of this NRTL model application, but also being an input of the NRTL model itself, the problem becomes implicit in nature. A typical approach may be to use a loop of calculations to re-estimate the input value of the unknown mole fraction. However, in this study a simpler approach was taken where the optimisation script used a single-step estimation with no feedback by using values of $X_M = x_{solute} = 0$ to calculate $X_{h,j}^{pred}$. This meant that the estimated parameters from this procedure should be used with a single-step implicit calculation with starting values of $X_M = 0$ to calculate solubility.

The non-randomness parameter α_{ij} was determined using a qualitative and manual approach. In this study a value of α_{ij} was arbitrarily chosen between 0.1 and 0.5 (with 0.1 intervals), this was because the limited number of data points for parameter estimation caused high sensitivity in finding a minimum of the objective function. Quite often the value of α_{ij} used was the only value between the allowed range which successfully found an accurate solution. This sensitivity was more prevalent in certain systems, and will be discussed further in the results section.

The code for the NRTL model itself was modified from a script shared on the MATHWORKS online file exchange [19].

2.3.5 UNIQUAC Model

Firstly, parameter estimation of τ_{ij} values for the UNIQUAC model was achieved using a similar approach was taken with the NRTL model. The same method was also applied with using $X_M = 0$ to in a single-step estimation for this problem which is also implicit. In this case, a slightly different optimisation procedure, shown below, was used by modifying the single constraint applied to determining BIPs. This constraint ensures that the diagonal τ_{ij} values were set to equal ones. Once again the *fmincon* solver with default options was used.

$$\text{Objective Function:} \qquad F = \sum_{h=1}^{N} \left(\frac{X_{m,h}^{exp} - X_{m,h}^{pred}}{X_{m,h}^{exp}} \right)^2 \qquad \textbf{(18)}$$

$$s.t.$$

$$J_{NC} \cdot I_{NC} \cdot \tau - 1 = 0 \qquad \textbf{(19)}$$

For each component in a solubility system the UNIQUAC parameters of q_i and r_i need to be determined prior application. In this work the group contribution structure for each molecule was determined using the DDBST Online UNIFAC Group Assignment tool [20]. For sub-components which could be identified by the online tool, values of Q_g and R_g were taken from Table 11-6 published in the *'The Handbook of Chemical Property Estimation Methods'* by Lyman et. al [21].

However, for several systems, the structure of the solute could not be broken down using online group assignment tool. For two solutes a work-around could be used; the q_i and r_i values of Sodium Diclofenac was taken from literature [22], and the Q_g and R_g values for the SO_2 group in Rofecoxib were estimated as being the equivalent of that of a SO_2 gas molecule [23]. Whilst the approach taken for Sodium Diclofenac is certainly more sound, it should be noted that using this method of estimation for Rofecoxib fails to take into account the two additional covalent bonds present in the SO_2 group , but not in the gaseous molecule.

The MATLAB code used in this work was a modified form of an internal research group UNIFAC script which was reduced in the residual term to represent the UNIQUAC model. However, for readers, there are many publicly available scripts of the UNIQUAC model on the internet which may be used directly or adapted into another language.

2.4 Data Analysis Methods

To determine the accuracy of each of the models both the Mean Percentage Difference (MPD) and the Normalized Root Mean Square Difference (NRMSE) were employed. The main advantage of using MPD is that it is clear to understand on average how far-off the predictions from a model will be. However, the division of each error by the measured data point can lead to misleading results for cases where the solubility crosses a value of 1, for this reason NRMSE is also being utilized to make up for this short-fall. Additionally, the error squaring of the NRMSE values exemplifies any large divergence of the solubility predictions from the experimental data set. The NRMSE was normalized by dividing by the range of the experimental data as opposed to the mean, this was because the experimental data points were not uniformly distributed by solvent fractions. By utilizing both methods of error measurement it is possible to analyze the accuracy of predictions with greater fairness and insight. Also, because predictions were produced in two unit sets for some systems, this meant that only normalized error calculations could be used so that the error of each model can be directly compared.

$$\mathrm{MPD} = \frac{100}{N} \cdot \sum_{h=1}^{N} \left| \frac{X_{m,h}^{exp} - X_{m,h}^{pred}}{X_{m,h}^{exp}} \right| \qquad \textbf{(20)}$$

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{h=1}^{N} \left(X_{m,h}^{exp} - X_{m,h}^{pred}\right)^2}{N}} \Big/ \left(X_m^{exp,MAX} - X_m^{exp,MIN}\right) \qquad \textbf{(21)}$$

For the activity coefficient models of NRTL and UNIQUAC, these models require solution mole fractions to determine the solute activity coefficient, which is then used to calculate solubility as a ternary mole fraction. Although all collected data sets were converted into ternary mole fractions for use of these two models, for the remaining four models, the predictions were made using the units as the data was presented in literature. This decision was

made to reduce the impact of any errors introduced in the unit conversion calculation. Also, because predictions were produced in two unit sets for some systems, this meant that only normalized error calculations could be used so that the error of each model can be directly compared.

3 Results

The model parameters for the Jouyban-Acree, GSM, NRTL, and UNIQUAC models which for each system which were determined in this study can be found in the Appendices of this report.

The accuracy of each model was tested by predicting the solubility data sets outlined in Table 1 following the previously discussed methodology. Table 2 shows the calculated MPD and NRMSE errors for each of the models for every system it could be applied to. Table 3 summarises the results by sorting the mean NRMSE for each model from lowest to highest in an order of: Jouyban-Acree, GSM, UNIQUAC, NRTL, log-linear, and predictive log-linear.

The Jouyban-Acree has the lowest average error, followed closely by the GSM at 0.0293 and 0.0296 respectively. In all prediction sets each model produced nearly identical results, this is to be expected since the GSM is derived from a modification to the Jouyban-Acree model. The only difference is that the neat solubilities are predicted using the GSM, but the Jouyban-Acree model directly uses the experimental data which leads to the slightly lower error.

Table 2 The Mean Percentage Difference (MPD [%]) and Normalised Root Mean Square Error (NRMSE [error]) of solubility predictions

Solute	System	Jouyban-Acree		General Single Model		Log-Linear		Predictive Log-Linear		NRTL		UNIQUAC	
		MPD	*NRMSE*	*MPD*	*NRMSE*	*MPD*	*NRMSE*	*MPD*	*NRMSE*	*MPD*	*NRMSE*	*MPD*	*NRMSE*
Sodium Diclofenac	1	13.0%	0.140	13.9%	0.138	64.6%	0.37	7.82E+6 %	879	25.5%	0.162	17%	0.154
Mesalazine	2	2.35%	0.0166	2.54%	0.0165	59.5%	0.59	291%	1.06	2.38%	0.0194	1.82%	0.0162
Alanine	3	1.02%	0.00805	1.12%	0.00820	7.74%	0.0321	2100%[b]	0.251	3.04%	0.0156	2.03%	0.0143
Cefazolin sodium pentahydrate	4	15.4%	0.0339	16.0%	0.0380	214%	0.116	164%	0.0917	10.4%	0.0263	-	-
Isoleucine (L)	5	4.47%	0.0326	4.62%	0.0327	22.4%	0.0607	23.7%	0.0617	5.43%	0.0423	4.43%	0.0287
Phenylalanine (L)	6	5.29%	0.0445	4.83%	0.0427	35.5%	0.215	35.8%	0.195	7.50%	0.0579	2.07%	0.0224
Benzoyl peroxide	7	13.2%	0.0250	13.9%	0.0306	25.8%	0.0221	76.7%	0.313	65.2%	0.320	-	-
Phenytoin	8	3.45%	0.0196	3.69%	0.0182	41.0%	0.221	65.9%	0.402	3.66%	0.474	-	-
Acetaminophen	9	2.25%	0.0276	2.36%	0.0273	39.6%	0.352	34.4%	0.290	44.3%	0.663	-	-
Valdecoxib (S)	10	1.40%	0.00203	1.40%	0.00203	35.4%	0.0985	238%	5.40	8.40E-06%	3.40E-06	-	-
Valdecoxib (S)	11	0.00%	0.000	0.00	0.00	28.2%	0.152	232%	2.43	1.02%	0.00163	-	-
Rofecoxib (S)	12	0.449%	0.00160	0.458%	0.00161	68.3%	0.231	148%	0.614	9.47%	0.0646	0.150%	0.000224
Mean		***5.19%***	***0.0293***	***5.40%***	***0.0297***	***53.5%***	***0.205***	***310%[a]***	***1.01[a]***	***15%***	***0.15***	***4.50%***	***0.0393***

[a] *Error calculations not including system 1 to reduce interference of extremely high error.*

.

For all systems, both of these models were capable of easily matching the curve profile of the experimental data – particularly so for systems 1, 2, and 6 which did not follow a simple logarithmic curve. It is clear that both of these models are very efficient at fitting experimental data, however, with over fitting became an issue. For system 11 with Valdecoxib and Glycerol + Water, only 5 data points were available for correlation, the fitted Jouyban-Acree and GSM models produced a model curve that went exactly through all five data points, resulting in errors of zero. This likely occurred because of the low number of data points occurring in a simple increasing logarithmic pattern reduced the amount of variability between each data point which resulted in a simple regression outcome.

Table 3 Mean MPD and NRMSE errors for each model as sorted by NRMSE lowest to highest.

Model	MPD	NRMSE
Jouyban-Acree	5.19%	0.0293
General Single Model	5.41%	0.0296
UNIQUAC	4.51%	0.0394
NRTL	14.8%	0.154
Log-Linear	53.5%	0.205
Predictive Log-Linear	310%	1.01

The UNIQUAC and NRTL models were the next two models with median average NRMSE errors of 0.0394 and 0.154 respectively. Unfortunately, it was not possible to use the UNIQUAC model for all of the studied systems, this was because of the incompatibility of the solute's structure with the basic UNIQUAC model. With that said, the significantly lower error of the UNIQUAC model should not be entirely discounted due to the lower sample size, for three out of the six systems modelled (5,6, and 12) UNIQUAC predictions were significantly more accurate than the NRTL predictions. However, due to not requiring any compound structure data, the NRTL model was easily applied to each of the systems included in this study. The main drawback with using the NRTL model encountered during this study was its high sensitivity in parameter estimation. Quite often the user changes in the adjustable α_{ij} parameter was the difference between a correlated model and a non-correlated model. The sensitivity was most prevalent when modelling the PEG-400 + water systems. This is likely because the high molecular weight of PEG-400 compared to water meant uniformly distributed volume/mass fraction data because highly non-uniform towards water mole fractions once the unit conversion was carried out, as a result there was a reduced distribution of data-points which negatively affected parameter estimation. This sensitivity problem, for the relevant systems, was not encountered when correlating the UNIQUAC model. For both UNIQUAC and NRTL models, it is important to remember that the binary interaction parameters which are obtained from parameter estimation can only be reliably applied to the experimental data they were determined from under the same conditions. Further work would be required to produce both generalised BIPs across multiple solutes and solvents for either model in solubility applications, during the literature review of this project no such research could be found. Whilst both the UNIQUAC and NRTL models may have greater errors compared to the Jouyban-Acree and GSM models, it is important to again consider over-fitting of the experimental data. For system 1 with sodium diclofenac and water + 2-propanol both the Jouyban-Acree and GSM predictions directly followed each data point, even as a data point may be above or below the trend in a downwards or upwards slope. Such points were likely due to inaccuracies in the solubility measurements and then manifested themselves in the regressed models. However, the UNIQUAC and NRTL models would not show any sensitivity to this variance in the

experimental data, which may be desirable a quality if the accuracy of the available solubility data of a particular cannot be cross-referenced.

More advanced UNIQUAC models exist to accommodate more complex compounds such as ionic solutes [20] [21], they were excluded from this comparison due to a lack of availability of group contribution parameters for each model. Likewise, extensions and developments to the NRTL model such as eNRTL-SAC and eNRTL overcome the limitations of non-electrolyte support in the NRTL model and improve applicability for cases where solubility experimental may not be available [26] [27]– at least for the systems which have been covered so far by research to publish model parameters.

Expectedly the log-linear and predictive log-linear models had the greatest average NRMSE values of 0.205 and 1.01 respectively. Despite the NRMSE of the log-linear model being in-line with the NRTL model, it has an MPD error almost 4 times greater at 53.5% compared to 14.8%. Table 2 indicates that for seven out of twelve systems the log-linear model had MPD errors of less than 40% (with system 8 just over at 40.1%), this is rather impressive considering that only the two neat solubility data points are required to operate this model. Nine out of twelve solubility data sets followed a logarithmic solubility profile which satisfies the main assumption behind the log-linear model which could contribute to the relatively accurate performance of this model. However, a similar study by [6] et al. on 43 ethanol + water systems showed an average relative deviation (assumed equivalent to MPD) of 48.6% which is only 4.9% less than the 53.5% average MPD obtained for the far fewer systems studied here. As for the predictive log-linear model, there was a much greater variation in prediction errors. For three systems (4, 5, and 9) MPD values less than 40% were obtained, which for these specific applications is impressive considering less experimental data is required. However, for the rest of the systems the predictive log-linear error was 1-4 times greater in magnitude over the rest of models. For systems where the cosolvent solubility is greater than the aqueous solubility, such as system 1 with sodium diclofenac and 2-propanol, extremely inaccurate results were produced. These high errors were to be expected since the predictive log-linear model replaces a solubility data point with the solubilization parameter σ as determined for the partition coefficient $\log_{10}K_{OW}$, therefore if a system has a solute with a $\log_{10}K_{OW}$, then the σ value will exponentially dominate resulting in the extremely high errors.

From the results of this study only for systems 4, 5, and 10 did the NRTL & UNIQUAC models yield a more accurate prediction than either the Jouyban-Acree or GSM model, and in these cases, the improvement on the latter models was only marginal. Both the NRTL or UNIQUAC models requires more time and (importantly) data to make operational over their other two correlative models. In this study, obtaining the UNIQUAC parameters for certain compounds was not possible within reasonable time, and determining the NRTL BIPs was a highly sensitive procedure for nearly half of the systems studied. As such, it is also important to compare the models qualitatively in terms of their capability and error-based return on investment for the time put into getting them operational.

Before selecting a solubility model, a few questions must first be answered about the nature of the use-case and system being modelled. Primarily these are: 1) How much experimental solubility data is available and what is the quality? 2) What component data is available? 3) What level of accuracy is required of the model? 4) How much time and effort is desired to put into the model?. In Table 4, a series of solubility model guidelines based directly on these questions, ahave been written to help aide readers in their model selection process. By

consulting this table composed from the authors experience and literature reading, the modeler will reduce the time spent exploring their options and finding the models which are suitable for them.

To summarise, if a modeler desires accurate model predictions with minimal effort on their part and has solubility data for their system available – they should use either the Jouyban-Acree or GSM model depending if the neat solvent data is included in their data set. However, if the modeler also requires the activity coefficients for their system (perhaps to model another system property) then they should choose either the NRTL or UNIQUAC model, choosing the former if a component in their system is not supported by UNIQUAC. If the modeler only has the neat primary and cosolvent solubility data, and no intermediate points, they should use the log-linear model. If they only have one solvent solubility data available, and their solute and solvents are supported, they should use the predictive log-linear model, but should expect a strong risk of highly inaccurate results.

Table 4 Decision making table for selecting the most appropriate solubility model for a modeller unfamiliar with solubility modelling.

Consideration	Jouyban-Acree	General Single Model	Log-Linear	Predictive Log-Linear	NRTL	UNIQUAC
1) How much experimental solubility data is available and what is the quality?	Requires multiple solvent mixture data points. Tendency to overfit variations if present. Requires need solvent solubility data.	Requires multiple solvent mixture data points. Tendency to overfit variations if present. Does not requires need solvent solubility data.	Only requires neat solvent solubility data.	Requires only one solvent data point.	Requires multiple solvent mixture data points. Does not overfit to variations in solubility data.	Requires multiple solvent mixture data points. Does not overfit to variations in solubility data.
2) What component data is available?	No component data required.	No component data required.	No component data required.	Only suitable for solvents where s and t have been determined. Requires solute $\log_{10}K_{OW}$ value	Solute temperature and enthalpy of fusion ($\Delta_{fus}H$ and T_{fus})	Solute temperature and enthalpy of fusion ($\Delta_{fus}H$ and T_{fus}) Components must be supported by UNIQUAC for r_i and q_i determination.
3) What level of accuracy is required of the model?	Suitable for high accuracy applications.	Suitable for high accuracy applications.	Adequate accuracy is needed for a system known to follow ideal solubility mixing rules.	Suitable for a priori solubility modelling purposes. Risk of highly inaccurate calculations	Suitable for high accuracy applications where the activity coefficients of system components are required.	Suitable for high accuracy applications where the activity coefficients of system components are required.
4) How much time and effort is desired to put into the model?	Regression can be completed easily completed in a short period of time using a software such a MATLAB.	Regression can be completed easily completed in a short period of time using a software such a MATLAB.	Model can immediately be implemented without correlation.	Model can immediately be implemented without correlation. Some time is needed to collect s, t, and $\log_{10}K_{OW}$	Parameter estimation requires access to optimisation software Parameter estimation requires user input and decision making.	Parameter estimation requires access to optimisation software. User must spend time determining r_i and q_i for each component of system.

4 References

[1] J. W. Millard, F. A. Alvarez-Núñez, and S. H. Yalkowsky, "Solubilization by cosolvents: Establishing useful constants for the log-linear model," *Int. J. Pharm.*, vol. 245, no. 1–2, pp. 153–166, 2002, doi: 10.1016/S0378-5173(02)00334-4.

[2] A. Jouyban, "Review of the cosolvency models for predicting solubility of drugs in water-cosolvent mixtures," *J. Pharm. Pharm. Sci.*, vol. 11, no. 1, pp. 32–58, 2008, doi: 10.18433/J3PP4K.

[3] S. H. Yalkowsky and T. J. Roseman, "Solubilization of drugs by cosolvents," *Tech. solubilization drugs, Marcel Dekker, New York*, vol. 12, pp. 91–134, 1981.

[4] A. Jouyban and W. E. Acree, "Mathematical derivation of the Jouyban-Acree model to represent solute solubility data in mixed solvents at various temperatures," *J. Mol. Liq.*, vol. 256, pp. 541–547, 2018, doi: 10.1016/j.molliq.2018.01.171.

[5] M. Barzegar-Jalali, E. Rahimpour, F. Martinez, and A. Jouyban, "Generally trained models to predict drug solubility in methanol + water mixtures," *J. Mol. Liq.*, vol. 264, pp. 631–644, 2018, doi: 10.1016/j.molliq.2018.05.084.

[6] S. N. Mirheydari, M. Barzegar-Jalali, W. E. Acree, H. Shekaari, A. Shayanfar, and A. Jouyban, "Comparison of the Models for Correlation of Drug Solubility in Ethanol + Water Binary Mixtures," *J. Solution Chem.*, vol. 48, no. 7, pp. 1079–1104, 2019, doi: 10.1007/s10953-019-00897-9.

[7] A. Jouyban-Gharamaleki, L. Valaee, M. Barzegar-Jalali, B. J. Clark, and W. E. Acree, "Comparison of various cosolvency models for calculating solute solubility in water-cosolvent mixtures," *Int. J. Pharm.*, vol. 177, no. 1, pp. 93–101, 1999, doi: 10.1016/S0378-5173(98)00333-0.

[8] W. E. A. J. Jouyban, "In silico prediction of drug solubility in water-ethanol mixtures using Jouyban- Acree model," *J Pharm Pharm. Sci 9(2) May - August 2006*, vol. 9, no. 2, pp. 262–269, 2006.

[9] A. Jouyban, "Review of the cosolvency models for predicting solubility in solvent mixtures: An update," *J. Pharm. Pharm. Sci.*, vol. 22, no. 1, pp. 466–485, 2019, doi: 10.18433/jpps30611.

[10] M. A. Miller and G. B. Kasting, "A spreadsheet-based method for simultaneously estimating the disposition of multiple ingredients applied to skin," *J. Pharm. Sci.*, vol. 104, no. 6, pp. 2047–2055, 2015, doi: 10.1002/jps.24450.

[11] J. T. Rubino and S. H. Yalkowsky, "Cosolvency and Deviations from Log-Linear Solubilization," *Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists*, vol. 4, no. 3. pp. 231–236, 1987, doi: 10.1023/A:1016408211963.

[12] A. Jouyban, F. Martinez, and W. E. Acree, "Correct Derivation of Cosolvency Models and Some Comments on 'solubility of Fenofibrate in Different Binary Solvents: Experimental Data and Results of Thermodynamic Modeling,'" *J. Chem. Eng. Data*,

vol. 62, no. 3, pp. 1153–1156, 2017, doi: 10.1021/acs.jced.6b00722.

[13] R. H. and P. J.M., “Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures,” *AICHE J.*, vol. 14, no. 1, pp. 135–144, 1968.

[14] D. S. Abrams and J. M. Prausnitz, “Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems,” *AIChE J.*, vol. 21, no. 1, pp. 116–128, 1975, doi: 10.1002/aic.690210115.

[15] A. Bondi, “van der Waals Volumes and Radii,” *J. Phys. Chem.*, vol. 68, no. 3, pp. 441–450, 1964.

[16] Physical Sciences Data-Science Service, “ACD/I-LABS.” https://www.psds.ac.uk/.

[17] A. A. Saei, F. Jabbaribar, M. A. A. Fakhree, W. E. Acree, and A. Jouyban, “Solubility of sodium diclofenac in binary water + alcohol solvent mixtures at 25°C,” *J. Drug Deliv. Sci. Technol.*, vol. 18, no. 2, pp. 149–151, 2008, doi: 10.1016/S1773-2247(08)50024-4.

[18] J. M. Sorensen, T. Magnussen, P. Ramussen, and A. Fredenslund, “LIQUID-LIQUID EQUILIBRIUM DATA: THEIR RETRIEVAL, CORRELATION AND PREDICTION,” *Fluid Phase Equilib.*, vol. 3, pp. 47–81, 1979.

[19] Nieck, “Component activities in liquid mixture, using NRTL,” *Mathworks File Exchange*. https://uk.mathworks.com/matlabcentral/fileexchange/48651-component-activities-in-liquid-mixture-using-nrtl.

[20] DDBST, “Online Group Assignment for UNIFAC and PSRK.” http://www.ddbst.com/unifacga.html.

[21] W. J. Lyman, W. F. Reehl, and D. H. Rosenblatt, *The Handbook of Chemical Property Estimation Methods*. 1982.

[22] L. F. Žilnik, A. Jazbinšek, A. Hvala, F. Vrečer, and A. Klamt, “Solubility of sodium diclofenac in different solvents,” *Fluid Phase Equilib.*, vol. 261, no. 1–2, pp. 140–145, 2007, doi: 10.1016/j.fluid.2007.07.020.

[23] R. Haghbakhsh and S. Raeissi, “Modeling vapor-liquid equilibria of mixtures of SO2 and deep eutectic solvents using the CPA-NRTL and CPA-UNIQUAC models,” *J. Mol. Liq.*, vol. 250, pp. 259–268, 2018, doi: 10.1016/j.molliq.2017.11.161.

[24] W. Zhu, B. Liu, H. Hou, L. Dou, and Q. Li, “A modified UNIQUAC model for calculating phase-equilibrium of ionic liquid containing systems,” *Fluid Phase Equilib.*, vol. 517, p. 112619, 2020, doi: 10.1016/j.fluid.2020.112619.

[25] K. Thomsen, M. C. Iliuta, and P. Rasmussen, “Extended UNIQUAC model for correlation and prediction of vapor-liquid-liquid-solid equilibria in aqueous salt systems containing non-electrolytes. Part B. Alcohol (ethanol, propanols, butanols)-water-salt systems,” *Chem. Eng. Sci.*, vol. 59, no. 17, pp. 3631–3647, 2004, doi: 10.1016/j.ces.2004.05.024.

[26] C. C. Chen and Y. Song, “Extension of nonrandom two-liquid segment activity coefficient model for electrolytes,” *Ind. Eng. Chem. Res.*, vol. 44, no. 23, pp. 8909–8921, 2005, doi: 10.1021/ie0503592.

[27] L. D. Simoni, Y. Lin, J. F. Brennecke, and M. A. Stadtherr, “Modeling liquid-liquid equilibrium of ionic liquid systems with NRTL, electrolyte-NRTL, and UNIQUAC,”

Ind. Eng. Chem. Res., vol. 47, no. 1, pp. 256–272, 2008, doi: 10.1021/ie070956j.

[28] E. Mazaher Haji Agha, M. Barzegar-Jalali, K. Adibkia, S. Hemmati, F. Martinez, and A. Jouyban, "Solubility and thermodynamic properties of mesalazine in {2-propanol + water} mixtures at various temperatures," *J. Mol. Liq.*, vol. 301, p. 112474, 2020, doi: 10.1016/j.molliq.2020.112474.

[29] C. J. Orella and D. J. Kirwan, "Correlation of Amino Acid Solubilities in Aqueous Aliphatic Alcohol Solutions," *Ind. Eng. Chem. Res.*, vol. 30, no. 5, pp. 1040–1045, 1991, doi: 10.1021/ie00053a028.

[30] J. Wu and J. Wang, "Solubility of cefazolin sodium pentahydrate in aqueous 2-propanol mixtures," *J. Chem. Eng. Data*, vol. 50, no. 3, pp. 980–982, 2005, doi: 10.1021/je049549u.

[31] T. Nagai and S. Prakongpan, "Solubility of acetaminophen in cosolvents. Chemical and Pharmaceutical Bulletin," *Chem. Pharm. Bull.*, vol. 57, no. 534, pp. 364–370, 1977, [Online]. Available: http://www.mendeley.com/research/geology-volcanic-history-eruptive-style-yakedake-volcano-group-central-japan/.

[32] E. M. Chellquist and W. G. Gorman, "Benzoyl Peroxide Solubility and Stability in Hydric Solvents," *Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists*, vol. 9, no. 10. pp. 1341–1346, 1992, doi: 10.1023/A:1015873805080.

[33] C. Liu, K. G. H. Desai, X. Chen, and X. Tang, "Solubility of valdecoxib in the presence of glycerol, propylene glycol, and poly(ethylene glycol) 400 at (298.15, 303.15, and 308.15) K," *J. Chem. Eng. Data*, vol. 50, no. 5, pp. 1736–1739, 2005, doi: 10.1021/je050190z.

[34] C. Liu, K. G. H. Desai, X. Tang, and X. Chen, "Solubility of rofecoxib in the presence of aqueous solutions of glycerol, propylene glycol, ethanol, Span 20, Tween 80, and sodium lauryl sulfate at (298.15, 303.15, and 308.15) K," *J. Chem. Eng. Data*, vol. 50, no. 6, pp. 2061–2064, 2005, doi: 10.1021/je050276s.

Appendices: Tables of Model Parameters

Table A2 Correlated Jouyban-Acree model parameters

System	Solute	Cosolvent	T [°C]	S_0	S_1	S_2
1	Sodium Diclofenac	2-Propanol	25	5.651	-2.973	15.67
2	Mesalazine	2-Propanol	25	7.517	0.1388	0.6312
3	Alanine	2-Propanol	25	0.6754	-1.075	-4.901
4	Cefazolin Sodium Pentahydrate	2-Propanol	25	-5.001	-8.020	-5.702
5	Isoleucine (L)	2-Propanol	25	3.673	5.248	0.625
6	Phenyl Alanine (L)	2-Propanol	25	5.010	0.2006	-3.096
7	Acetaminophen	PEG-400	30	3.681	1.055	0.3633
8	Benzoyl Peroxide	PEG-400	25	1.935	-2.106	0.9834
9	Phenytoin	PEG-400	25	3.105	4.217	4.942
10	Valdecoxib (S)	PEG-400	25	3.906	2.095	2.226
11	Valdecoxib (S)	Glycerol	25	1.703	0.2298	0.3024
12	Rofecoxib (S)	Glycerol	25	-2.345	-3.094	-2.706

Table A2 Correlated GSM model parameters

System	Solute	Cosolvent	T [°C]	B_0	B_1	B_2	B_3	B_4
1	Sodium Diclofenac	2-Propanol	25	-3.662	20.53	-81.63	130.2	-68.26
2	Mesalazine	2-Propanol	25	-9.843	9.060	-11.56	5.792	-2.660
3	Alanine	2-Propanol	25	-7.286	2.265	28.61	-43.53	20.59
4	Cefazolin Sodium Pentahydrate	2-Propanol	25	-7.358	-13.96	52.94	-56.91	21.29
5	Isoleucine (L)	2-Propanol	25	-7.090	15.53	-24.40	18.80	-4.294
6	Phenyl Alanine (L)	2-Propanol	25	-6.937	7.805	6.897	-19.29	-9.634
7	Acetaminophen	PEG-400	30	0.2496	3.029	-9.952	5.262	-1.499
8	Benzoyl Peroxide	PEG-400	25	3.747	-14.16	10.57	12.93	3.933
9	Phenytoin	PEG-400	25	4.247	4.290	-40.74	47.80	-19.44
10	Valdecoxib (S)	PEG-400	25	0.7590	2.908	-21.43	22.19	-9.005
11	Valdecoxib (S)	Glycerol	25	3.920	-2.764	-0.4984	2.879	-1.210
12	Rofecoxib (S)	Glycerol	25	4.257	-10.29	25.15	27.82	10.82

Table A3 Optimised Temperature Dependent NRTL model parameters

System	Solute (3)	Cosolvent (2)	T [°C]	τ_{12}	τ_{13}	τ_{21}	τ_{23}	τ_{31}	τ_{32}
1	Sodium Diclofenac	2-Propanol	25	-17.23	29.26	-9.891	-0.8549	73.79	5.0273
2	Mesalazine	2-Propanol	25	13.88	0.0064	-1.308	-8.189	-3.733	13.28
3	Alanine	2-Propanol	25	-5.171	5.182	-0.3844	4.664	-2.063	-0.1872
4	Cefazolin Sodium Pentahydrate	2-Propanol	25	-12.93	-1.155	12.93	2.313	22.46	2.725
5	Isoleucine (L)	2-Propanol	25	-2.013	1.850	12.54	7.260	5.663	5.586
6	Phenyl Alanine (L)	2-Propanol	25	-17.39	3.884	1635	10.04	1374	-14.86
7	Acetaminophen	PEG-400	30	5.867	1289	-3.131	2.151	1289	-2.761
8	Benzoyl Peroxide	PEG-400	25	-275.8	14.22	964.5	-4.285	0.3291	8.991
9	Phenytoin	PEG-400	25	16.17	7.411	300.8	201.6	-0.7394	-4.009
10	Valdecoxib (S)	PEG-400	25	695.8	8.860	-85.78	60.12	3.333	590.5
11	Valdecoxib (S)	Glycerol	25	351.3	7.750	6.436	7.288	3.668	166.2
12	Rofecoxib (S)	Glycerol	25	396.0	7.674	-227.5	4.281	0.0542	71.53

1: Water, 2: Cosolvent, 3: Solute

Table A4 Optimised Temperature Dependent UNIQUAC model parameters

System	Solute (3)	Cosolvent (2)	T [°C]	τ_{12}	τ_{13}	τ_{21}	τ_{23}	τ_{31}	τ_{32}
1	Sodium Diclofenac	2-Propanol	25	-0.0191	0.0374	1.603	0.4029	2.337	0.5041
2	Mesalazine	2-Propanol	25	-0.1123	4.120	0.9339	18.04	2.251	-0.0041
3	Alanine	2-Propanol	25	56910	3.008	3.396	0.1247	0.4811	1.607
4	Cefazolin Sodium Pentahydrate	2-Propanol	25	-	-	-	-	-	-
5	Isoleucine (L)	2-Propanol	25	6.401	0.8510	5.617	0.0245	0.3000	1.959
6	Phenyl Alanine (L)	2-Propanol	25	0.4616	0.2789	3.212	0.0438	2.097	2.382
7	Acetaminophen	PEG-400	30	-	-	-	-	-	-
8	Benzoyl Peroxide	PEG-400	25	-	-	-	-	-	-
9	Phenytoin	PEG-400	25	-	-	-	-	-	-
10	Valdecoxib (S)	PEG-400	25	-	-	-	-	-	-
11	Valdecoxib (S)	Glycerol	25	-	-	-	-	-	-
12	Rofecoxib (S)	Glycerol	25	4.405	0.8826	1.566	1.289	0.7342	0.4214

1: Water, 2: Cosolvent, 3: Solute