Model Analysis of Cellobiose Solubility in Organic Solvents and Water

by

Joseph Heng

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Chemical Engineering

by

_____

May 2020

APPROVED:

_____

Professor Michael T. Timko, Thesis Advisor

_____

Professor William M. Clark, Committee Member

_____

Professor Christopher R. Lambert, Committee Member

_____

Professor Stephen J. Kmiotek, Committee Member

Abstract

The solubility of cellobiose in 18 organic liquids and water was measured at 20°C. Hydrogen bond acceptors were the most effective solvents. Three models were analyzed to evaluate their accuracy and to understand factors that affect cellobiose solubility: Hansen solubility parameters (HSP), linear free energy relationship (LFER), and UNIQUAC functional-group activity coefficients (UNIFAC). The HSP of cellobiose were determined and the model was able to distinguish between most good and poor solvents, however, proved to be occasionally unreliable due to a false negative. The LFER model produced an empirical equation involving contributions from solvent molar refraction, polarizability, acidity, basicity, and molar volume, which predicted cellobiose solubilities to within ±2 log units. LFER indicated that good solvents were highly polarizable and had low molar volume, which was consistent with the good solvents found for cellobiose. A modified version of UNIFAC that includes an association term (A-UNIFAC) predicted the solubility of cellobiose in water and alcohols to within ±0.6 log units, indicating that A-UNIFAC can be used to predict the solubility of cellobiose and other carbohydrates provided additional data to extend the model to solvents other than water and alcohols.

Acknowledgements

Table of Contents

List of Figures and Tables

1. Introduction

Lignocellulosic biomass is an abundant renewable resource that can be used as a source of carbon to produce liquid fuels and organic materials [1,2]. Cellulose, the major component of lignocellulosic biomass, is a polysaccharide composed of chains of glucose units linked by *β*-1,4-glycosidic bonds. Degradation of cellulose is desired for energy extraction and production of useful products for chemical applications [2-4]. However, cellulose is resistant to degradation and, therefore, severe conditions are required to break down cellulose. The conversion of cellulose has been studied for decades to find ways to improve product yields and reduce costs [2,4].

The selection of a solvent to act as a reaction medium for the conversion of cellulose is one of the most important steps to consider. Dissolving cellulose into the liquid phase allows for homogeneous reactions that are easier to control and less energy intensive than heterogeneous reactions [5]. Additionally, solvents can participate in the conversion of sugars and, therefore, careful solvent selection can control reaction pathways to obtain desired products [6]. However, common solvents are ineffective at dissolving cellulose because the size and crystallinity of cellulose affect its solubility [5,7]. Several effective solvent systems for cellulose have been known for many years, but finding new solvent systems for novel applications, especially those that do not degrade cellulose, remains an important topic to pursue.

Solubility data is the first step in finding new solvents because it provides an insight on solute-solvent molecular interactions. Organic solvents, despite their weak ability to dissolve cellulose alone [8,9], remain an important class of solvents to study because they are still used in a variety of solvent systems to dissolve cellulose [10-11]. Understanding how different functional groups and properties of various organic solvents are related to their ability to dissolve cellulose can lead to an understanding of the molecular interactions involving cellulose. In order to minimize

the size and crystallinity factors on cellulose solubility, this study will focus on cellobiose, a disaccharide composed of two glucose units linked by the same β-1,4-glycosidic bond. Cellobiose is often considered the smallest unit of cellulose [5,7,12] and, therefore, can act as a representative model for cellulose that is not affected by size or crystallinity factors. Furthermore, Jóndóttir et al. [13] reported different values for the solubility of cellobiose in water compared to the isomers maltose and trehalose. These isomers contain different glycosidic bonds, indicating that this glycosidic bond is an important factor to consider for solubility. As a result, cellobiose may be more representative of cellulose than glucose despite glucose being a smaller unit of cellulose than cellobiose.

There are limited solubility data of carbohydrates available in the literature and existing data are often not in agreement due to factors such as different equilibration times and different procedures [14]. In order to address the lack of reliable solubility data, predictive models can be used to estimate the solubility of carbohydrates. However, it is difficult to construct an accurate model due to the variability in literature data. Several models have been proposed and specifically fitted to available literature data. Kononenko and Herstein [14] used a linear relation with respect to temperature with two parameters fitted to their own solubility data of sucrose in non-aqueous solvents. More recently, activity coefficient models such as the universal quasi-chemical (UNIQUAC) and UNIQUAC functional-group activity coefficients (UNIFAC) models have been used to describe carbohydrate solubility, primarily in water, alcohols, and water-alcohol mixtures. Various modifications to these models have been proposed specifically for carbohydrates [15-20], but use parameters available only for a limited number of functional groups and require more experimental data for higher accuracy. Useful models can also help identify factors that affect a

solvent's ability to dissolve cellobiose and will be helpful in finding additional solvents that can be used for cellobiose, similar carbohydrates, or new cellulose solvent systems.

In this study, the solubilities of cellobiose in 18 organic solvents and water were evaluated at room temperature (20°C). The predictive abilities of three commonly used solubility models were also evaluated: Hansen solubility parameters (HSP), linear free energy relationship (LFER), and UNIFAC. Each model has a slightly different approach to predict solubility, so each model can provide their own insight on important factors that affect cellobiose solubility. These models can be used to predict the dissolution ability of other solvents for cellobiose as well as other carbohydrates.

2. Solubility Models

2.1. Hansen Solubility Parameters (HSP)

Hansen's solubility theory utilizes solubility parameters to provide an analysis on the solubility of a solute in a solvent or the miscibility of liquids. Based on the principle "like dissolves like," two components with similar solubility parameters are expected to be miscible. These solubility parameters, sometimes called the cohesion energy parameters, are derived from the energy of vaporization, which measures the total cohesive energy that hold molecules together in the liquid phase. Hansen's solubility theory builds on the Hildebrand solubility parameter (δ_t) which is defined in Eq. (1), where E is the total cohesion energy and V is the molar volume.

$$\delta_t = (E/V)^{1/2} \tag{1}$$

The total cohesion energy is calculated from the molar heat of vaporization (ΔH_v) in Eq. (2), where R is the gas constant and T is the absolute temperature.

$$E = \Delta H_v - RT \tag{2}$$

The single Hildebrand solubility parameter is not descriptive enough to always accurately predict miscibility because it cannot account for the association between molecules that arises from polar and hydrogen-bonding interactions [21]. In order to attempt to solve the issues of the Hildebrand solubility parameter, Hansen divides the total cohesion energy into three factors to describe the three major types of interactions: London dispersion, polar (dipole-dipole), and hydrogen-bonding interactions, represented by the subscripts D, P, and H, respectively.

$$E = E_D + E_P + E_H \tag{3}$$

Combining Eq. (3) with Hildebrand's solubility parameter in Eq. (1) gives Hansen's three solubility parameters based on the three types of interactions.

$$\delta_t^2 = \delta_D^2 + \delta_P^2 + \delta_H^2 \tag{4}$$

Solubility parameters have been widely used for many applications and are, therefore, available in the literature for many common compounds. For compounds with unknown solubility parameters, experimental procedures and calculations are reliable for obtaining these unknown parameters. The calculation for δ_D is detailed by Blanks and Prausnitz [22]. δ_P is calculated by a simplified version of the Böttcher Equation, shown in Eq. (5), where μ is the dipole moment in Debye and V is the molar volume [21].

$$\delta_P = \frac{37.4\mu}{V^{\frac{1}{2}}} \left[MPa^{\frac{1}{2}} \right] \tag{5}$$

δ_H can either be calculated by subtracting the dispersion and polar energies from the total energy of vaporization or by using group-contribution methods. One group-contribution method that can be used for all three of Hansen's solubility parameters has been developed by Stefanis and Panayiotou [23]. Alternatively, each solubility parameter can be calculated using statistical mechanics as outlined by Hansen [21].

In order to describe the "like dissolves like" principle, the "distance" between the solubility parameters of a solvent (1) and a solute (2), R_a, can be calculated using Eq. (6), where the constant "4" is a geometric mean that differentiates between atomic and molecular interactions. This constant was found experimentally, but is theoretically predicted by the Prigogine corresponding states theory of polymer solutions [21].

$$R_a = [4(\delta_{D1} - \delta_{D2})^2 + (\delta_{P1} - \delta_{P2})^2 + (\delta_{H1} - \delta_{H2})^2]^{1/2} \qquad (6)$$

A smaller distance between the solubility parameters of two components means the interaction forces between the components are more similar and, therefore, expected to have higher miscibility. The solubility analysis can be visually represented by plotting each solvent as a point in a 3-D plot of δ_D vs. δ_P vs. δ_H. A sphere with radius R_o can be constructed centered at the solute of interest in order to predict the ability of a solvent to dissolve the solute. Good solvents will lie within the HSP sphere and poor solvents will lie outside the HSP sphere. The HSP sphere allows solubility to be predicted through the relative energy difference (*RED*), calculated as the ratio of the distance between the HSP of the solvent and solute to the HSP sphere radius. If $RED \leq 1$, then the solvent lies inside the sphere and has high affinity for the solute. If $RED > 1$, then the solvent lies outside of the sphere and has lower affinity for the solute.

$$RED = \frac{R_a}{R_o} \qquad (7)$$

Interestingly, good solvent mixtures can be formed by mixing two bad solvents. HSP for solvent mixtures are calculated using Eq. (8), where ϕ is the volume fraction of each solvent in the mixture and the subscript *i* represents the dispersion, polar, and hydrogen-bonding interactions [24].

$$\delta_i = \phi_1 \delta_{i1} + \phi_2 \delta_{i2} \qquad (8)$$

Solvent mixtures will not be considered for this study, however, it is worth mentioning the idea because solvent mixtures are common for dissolving cellulose.

2.2. Linear Free Energy Relationship (LFER)

The LFER model, or linear solvation energy relationship (LSER), uses five descriptors to model a solvation property. It views the process of transferring a solute into a solvent as a three-step process. First, a cavity of the correct size and shape for the solute within the solvent must form. This requires breaking the intermolecular bonds between the solvent molecules, which is always thermodynamically unfavorable. Next, a solute molecule fills the cavity without interacting with the solvent. Finally, the solute-solvent interactions activate and "charge" the solute. This is a simplified view of the solvation process that does not explicitly take the entropy of solution or free volume effects into account, although they may be implicitly included within the cavity term because they vary with solute size [25].

The most recent LFER model [26] for solids and liquids is given in Eq. (9), where SP is any property of a series of solvents or solutes for a compound of interest and E, S, A, B, and V are the descriptors for the series of solvents or solutes. The coefficients e, s, a, b, and v and the system constant c are determined by multiple linear regression.

$$SP = c + eE + sS + aA + bB + vV \tag{9}$$

The excess molar refraction E is a measure of the polarizable electrons in a molecule. It is calculated as the difference between the molar refraction of the compound and the molar refraction of an alkane of the same volume, which is represented in Eq. (10) where MR_x is the molar refraction of the compound and V is McGowan's characteristic molar volume [27].

$$E = MR_x - 2.83195V + 0.52553 \tag{10}$$

MR_x is defined in Eq. (11), where η is the refractive index of the compound as a pure liquid at 20°C [26].

$$MR_x = 10 \left[\frac{(\eta^2 - 1)}{\eta^2 + 2}\right] V \tag{11}$$

S is the polarizability/dipolarity descriptor, A is the overall hydrogen bond acidity, and B is the overall hydrogen bond basicity. S, A, and B are usually obtained from experimental chromatography data and water-solvent partition coefficients [28]. The V descriptor is McGowan's characteristic molar volume in $cm^3mol^{-1}/100$, which is a measure of the cavity formation in the solvation process. It is very simple to calculate V for compounds with known molecular structures; the calculation is described by Abraham and McGowan [27]. LFER descriptors for many common compounds are available in the literature and group-contribution methods, such as the one by Platts et al. [29], are also available for estimations of unknown descriptor values.

The regressed coefficients represent the complimentary effect of the compound of interest. e is the tendency of the compound to interact with solvents or solutes through π- or n- electron pairs. This value is often positive, but can be negative for phases containing fluorine atoms [29,30]. s is the compound's tendency to interact with polarizable/dipolar solvents or solutes. a is the compound's hydrogen bond basicity because acidic compounds interact with basic compounds. Similarly, b is the hydrogen bond acidity of the compound. v describes the cavity effects during the solvation process. The sign and magnitude of each regressed coefficient reflects the importance of its corresponding descriptor. A larger coefficient value represents greater influence of its descriptor to the solvation property while a zero or near-zero coefficient value represents minimal influence of its descriptor to the solvation property.

2.3. Solid-Liquid Equilibrium (SLE) and UNIFAC

The solubility of a solid solute in a liquid solvent is derived from Eq. (12), which describes the equilibrium between the solid and liquid phases of the solute.

$$f_2^S = f_2^L = x_2 \gamma_2 f_2^o \tag{12}$$

f_2^S is the fugacity of the solute (component 2) in the solid phase, f_2^L is the fugacity of the solute in the liquid phase, x_2 is the solubility of the solute in the solvent in mole fraction, γ_2 is the liquid phase activity coefficient, and f_2^o is the standard state fugacity to which the activity coefficient refers, which is normally taken as the pure subcooled liquid of the solute at the solution temperature and at the saturation pressure because it can be accurately calculated in those conditions. For solids, a thermodynamic cycle involving a phase change through the triple point temperature may be used to define the ratio of f_2^o to f_2^S; the step-by-step derivation is detailed by Prausnitz et al. [31]. The derived relationship, shown in Eq. (13), is used to calculate the solubility of a solute in any given solvent, which depends only on the solute properties (Δh_{fus}, T_t, and Δc_p) and not the solvent properties. The ideal solubility can be calculated by setting the activity coefficient to unity.

$$\ln \frac{f_2^o}{f_2^S} = \ln \frac{1}{x_2 \gamma_2} = \frac{\Delta h_{fus}}{RT_t}\left(\frac{T_t}{T} - 1\right) - \frac{\Delta c_p}{R}\left(\frac{T_t}{T} - 1\right) + \frac{\Delta c_p}{R}\ln \frac{T_t}{T} \tag{13}$$

Eq. (13) is frequently simplified in two ways. First, the melting temperature (T_m) is often used in place of the triple point temperature (T_t) because there is little difference between T_m and T_t for most substances and the enthalpy of fusion (Δh_{fus}) data is more commonly available at T_m. Second, the two latter terms are often neglected because they tend to cancel each other out, especially if the solution temperature T is close to T_m or T_t [31]. In this study, T_m will be used in place of T_t, but no terms will be neglected because each term has a significant contribution to the solubility calculation for cellobiose.

Cellobiose is expected to form non-ideal solutions with many solvents due to dissimilar molecular structures, therefore, activity coefficients (which describe intermolecular forces between the solute and solvent) are required to calculate the actual solubility. If the activity coefficient is greater than unity, then dispersion forces dominate and the actual solubility is less than ideal. If the activity coefficient is less than unity, then polar forces play an important role and the actual solubility is greater than ideal [31]. UNIFAC is a widely used, relatively simple, and predictive method for activity coefficient estimations. It combines equations from the UNIQUAC activity coefficient model with a group-contribution method. In the UNIFAC model, the activity coefficient of each component in a solution consists of a combinatorial part (γ_i^C), describing the differences in size and shape of the molecules in the solution, and a residual part (γ_i^R), describing the intermolecular forces occurring in the mixture. The details of calculating each part of the activity coefficient is described by Fredenslund et al. [32] and outlined in Appendix B.

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \tag{14}$$

UNIFAC considers physical properties of each compound as a sum of contributions by its functional groups. Interaction parameters between every pair of groups in a mixture are evaluated from experimental phase equilibrium data, which can then be used to predict activity coefficients for mixtures with unknown phase equilibrium data if given the functional groups within the mixture. As a result, the accuracy of UNIFAC depends on the accuracy and availability of the interaction parameters. This study uses interaction parameters available from the Dortmund Data Bank [33-38].

Modified UNIFAC models have been proposed to improve the model's accuracy, such as the modified UNIFAC (Dortmund) [39] and modified UNIFAC (Lyngby) [40] models. The more recent modified UNIFAC (NIST) [41] was evaluated in this study because a more complete set of

interaction parameters was available. The modified UNIFAC (NIST) model uses the same equations as the modified UNIFAC (Dortmund) model with interaction parameters derived from the NIST database. These two models use a slightly different volume fraction calculation for the combinatorial part and include a temperature dependence to the interaction parameters for the residual part.

$$\phi_i' = \frac{r_i^{3/4}}{\sum_j x_j r_j^{3/4}} \tag{15}$$

$$\ln \gamma_i^C = 1 - \phi_i' + \ln \phi_i' - 5q_i \left(1 - \frac{\phi_i}{\theta_i} + \ln \frac{\phi_i}{\theta_i}\right) \tag{16}$$

$$\psi_{nm} = \exp(-(a_{nm} + b_{nm}T + c_{nm}T^2)/T) \tag{17}$$

Other modifications have been proposed specifically to model the solubility of carbohydrates, primarily in water and alcohols. One of these modified model is A-UNIFAC, proposed by Mengarelli et al. [42] and adapted by Ferreira et al. [17], which adds an association part to the activity coefficient that takes hydrogen bonding into account and attempts to solve self- and cross-association problems in mixtures.

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R + \ln \gamma_i^A \tag{18}$$

The combinatorial and residual parts are calculated the same way as the original UNIFAC model by Fredenslund et al. [32]. The calculation for the association part is detailed by Ferreira et al. [17] and outlined in Appendix B. Additionally, new groups were defined to better represent the carbohydrate compounds and distinguish differences between isomers: the pyranose ring (PYR), the furanose ring (FUR), the hydroxyl group attached to the ring (OH_{ring}), and the osidic bond (-O-). Interaction parameters between these new groups and the groups already defined for alcohols and water (CH_3, CH_2, OH, H_2O, CH_3OH) have been estimated by Ferreira et al. [17] from regressing experimental phase equilibrium data of glucose, sucrose, and fructose. These interaction

parameters were revised by Montañés et al. [43] to fit a larger database of solubility data and those revised parameters were used in this study. These modifications allow A-UNIFAC to predict carbohydrate solubility in alcohols and water more accurately than the original UNIFAC.

There are other proposed modifications to the UNIFAC model for carbohydrates, such as the S-UNIFAC [20] that defines different groups to describe the sugar molecules, the Bio-UNIFAC [18] that is based on an asymmetric activity coefficient model, the P&M UNIFAC [19] that modifies the calculation for the combinatorial part, and the modified UNIFAC [16] that modifies the calculation for both the combinatorial and residual parts. These models were not used in this study because A-UNIFAC is the most recently published model and showed promising results when Gong et al. [44] applied it to sugars in aqueous solutions.

3. Materials and Methods

Materials: Cellobiose >98% was obtained from Sigma Aldrich. Ethanol ACS grade was obtained from Fisher Scientific, ethylbenzene >99% was obtained from Fluka, aniline >98% was obtained from Tokyo Chemical Industry, and the remaining solvents (all >98%) were obtained from Sigma Aldrich. DI water was used. All the materials were used as delivered; all solvents were delivered in the liquid state.

Solubility Measurements: 2 mL of solvent and cellobiose in excess of the expected saturation concentration were added to a tube with an agitator operating at approximately 200 rpm and placed in a water bath at room temperature ($20 \pm 1°C$). The tube was covered with Parafilm and was then allowed to reach equilibrium over 24 hours. The temperature of the water bath was monitored with a thermocouple throughout the mixing time. The solution was centrifuged at 1500 rpm for 30 minutes. The liquid portion of the solution containing the saturated amount of cellobiose was recovered, filtered, and then analyzed using an Agilent 1200 series high-performance liquid

chromatography (HPLC) with a 5 mM sulfuric acid solution as the mobile phase. Solubility results were averaged from at least two runs. Standard deviation reached up to ~10% for solvents that dissolved at least 0.1 g L^{-1} of cellobiose and up to 45% for solvents that dissolved less than 0.1 g L^{-1}.

Solubility Calculations: Cellobiose concentrations obtained from HPLC were calibrated to units of g L^{-1}. It is more useful to express solubility in mole fraction (mol cellobiose/mol solution), especially for the UNIFAC model. The conversion from g L^{-1} to mole fraction requires the density of the solution (ρ_{soln}), which was unfortunately not measured. For low cellobiose solubilities, the density of solution can be assumed to be nearly equivalent to the density of the pure solvent ($\rho_{solvent}$). For higher solubilities, this assumption is not valid. Montañés et al. [43] reports the difference between the density of solution and density of pure solvent for carbohydrate-alcohol mixtures to be nearly 11% for a solubility of approximately 230 g L^{-1}. Additionally, Taylor [45] reports a 16% difference between the density of cellobiose-water solution and the density of pure water at 0.03 mole fraction. In this study, densities of cellobiose solutions were estimated as a weighted average of the pure solvent density and pure cellobiose density (ρ_c). Cellobiose density is predicted to be approximately 1.8 g mL^{-1} by the ACD/Labs Percepta Platform. The cellobiose weight factor was estimated using the cellobiose concentration (C_C).

$$\rho_{soln} = \frac{C_C}{\rho_{solvent}+C_C}\rho_C + \left(1 - \frac{C_C}{\rho_{solvent}+C_C}\right)\rho_{solvent} \tag{19}$$

Estimated densities of cellobiose solutions using Eq. (19) matched the previously mentioned trends: the difference between the density of cellobiose-dimethyl sulfoxide (DMSO) solution (220 g L^{-1}) and the density of pure DMSO was nearly 11%. Therefore, this estimation was considered sufficient for this study.

The average absolute deviation (AAD) and average relative deviation (ARD) were used in this study to compare the accuracy of LFER, UNIFAC, and their variations. AAD and ARD were calculated using Eqs. (20) and (21) [44], where S_n is the solubility of cellobiose in log units (mole fraction), superscripts *expt* and *calc* are experimental values and calculated values, respectively, and *NDP* is the number of experimental data points. Log units were used for these calculations in order to weigh each data point more evenly. Lower values for AAD and ARD are desirable; AAD and ARD of 0% would indicate that the model is 100% accurate.

$$AAD = \frac{\sum_n \left| S_n^{expt} - S_n^{calc} \right|}{NDP} \times 100\% \qquad (20)$$

$$ARD = \frac{\sum_n \left| \frac{S_n^{expt} - S_n^{calc}}{S_n^{expt}} \right|}{NDP} \times 100\% \qquad (21)$$

All models were evaluated using the MATLAB (R2019a) software: Gharagheizi [46] presented an improved HSP algorithm, the function *fitlm* was used for the LFER model, and the algorithm for the UNIFAC model was modified from a code developed by Saeed Mardani [47].

4. Results and Discussion

4.1. Cellobiose Solubility

Solvents are often classified into three categories: polar aprotic, polar protic, and nonpolar solvents. These categories differentiate between hydrogen bond acceptors and hydrogen bond donors (and nonpolar solvents), which will facilitate our understanding of factors that affect cellobiose solubility. Multiple solvents in each of the three solvent categories were selected to dissolve cellobiose. DMSO and N,N-dimethylformamide (DMF) were selected because they are often used for cellulose solvent systems [9,11] so they were expected to be good solvents for cellobiose. Morpholine was selected because Kononenko and Herstein [14] reported morpholine

to be a good solvent for sucrose and therefore was expected to also be a good solvent for cellobiose. Aniline was selected for the Hansen solubility analysis, which will be further discussed in the next section. It was also important to test commonly-used solvents, including water, formamide, acetic acid, methanol, and ethanol. As a result, the other alcohols and carboxylic acids were selected to compare differences in size and functional groups to the common solvents. Additionally, dichloroacetic acid was important for the LFER analysis because its set of LFER descriptors were different from all other solvents in this study (high values for both S and A). Finally, three nonpolar solvents were selected to test aliphatic and aromatic structures. The experimental solubility of cellobiose in various organic solvents and water at $20 \pm 1°C$ are listed in Table 1. It is well-known that temperature has a strong effect on solubility, so data was collected for only one temperature because this study will focus primarily on the solvent factors that affect solubility rather than temperature factors.

Table 1. Experimental solubilities and activity coefficients of cellobiose in select organic solvents and water.

Solvent	Solubility (g L^{-1})	Solubility (mole fraction×10^{-3})	γ_2^{sat}
Polar aprotic (hydrogen bond acceptors)			
Dimethyl sulfoxide	220 ± 10	48 ± 3	0.019 ± 0.001
Dimethylformamide	4.6 ± 0.2	1.03 ± 0.07	0.88 ± 0.06
Morpholine	162 ± 3	42 ± 1	0.0216 ± 0.0006
Aniline	0.83 ± 0.08	0.22 ± 0.03	4.1 ± 0.5
Polar protic (hydrogen bond donors)			
Water	177 ± 1	9.78 ± 0.09	0.0924 ± 0.0009
Formamide	42 ± 4	4.9 ± 0.7	0.18 ± 0.03
Acetic acid	0.475 ± 0.009	0.079 ± 0.002	11.4 ± 0.3
Dichloroacetic acid	40 ± 7	10 ± 2	0.09 ± 0.02
Propionic acid	0.04 ± 0.01	0.010 ± 0.004	90 ± 40
Methanol	0.44 ± 0.02	0.052 ± 0.003	17 ± 1
Ethanol	0.14 ± 0.01	0.024 ± 0.004	38 ± 6
1-Propanol	0.0315 ± 0.0004	0.0069 ± 0.0001	131 ± 2
1-Butanol	0.018 ± 0.008	0.005 ± 0.003	200 ± 100
1-Pentanol	0.010 ± 0.002	0.0031 ± 0.0009	290 ± 90
Benzyl alcohol	0.012 ± 0.002	0.0036 ± 0.0007	250 ± 50
Ethylene glycol	8.3 ± 0.4	1.36 ± 0.08	0.67 ± 0.04
Nonpolar			
1-Octene	≤ 0.001	≤ 0.0005	≥ 2000
Ethylbenzene	≤ 0.001	≤ 0.0004	≥ 2000
p-Xylene	≤ 0.001	≤ 0.0004	≥ 2000

Based on the trends in Table 1, polar aprotic solvents generally performed better than polar protic and nonpolar solvents with DMSO as notably the best solvent out of the ones tested in this study. Certain polar protic solvents also work quite well at dissolving cellobiose: water and formamide are good polar protic solvents possibly because they also have the ability to accept

hydrogen bonds (making them amphiprotic solvents) while dichloroacetic acid and ethylene glycol are good polar protic solvents likely because they have high polarity. Unsurprisingly, cellobiose is essentially insoluble in nonpolar solvents. Each of the three nonpolar solvents that were evaluated dissolved trace amounts of cellobiose that could not be detected by HPLC, which was enough to suggest that other nonpolar solvents follow a similar trend. The insolubility of cellobiose in nonpolar solvents is also important information to report because they can be useful in certain reactions in which the desired products are soluble in nonpolar solvents [48].

Comparing solubilities between the alcohols in Table 1 suggests that a smaller molecular size correlates with a higher cellobiose solubility. Solubility is greatest for methanol and decreases as the alcohol size increases to 1-pentanol and benzyl alcohol. This correlation is also seen when comparing propionic acid to acetic acid and DMF to formamide. Propionic acid is a slightly larger compound compared to acetic acid as it contains an additional CH_2 group, which is likely the major reason it cannot dissolve cellobiose as well as acetic acid. Similarly, DMF is a slightly larger compound than formamide and dissolves less cellobiose than formamide. However, other factors such as polarity are also important to consider in order to explain differences in solubilities between DMF and formamide.

4.2. HSP Analysis

In order to construct a HSP sphere for cellobiose, each solvent must be classified as a "good" solvent with a score of 1 or a "poor" solvent with a score of 0. Table 2 lists the HSP and score for each solvent. The criteria to be a good solvent is to be able to dissolve more than the calculated ideal solubility of cellobiose at room temperature (i.e. $\gamma_2^{sat} < 1$). DMSO, DMF, morpholine, water, formamide, dichloroacetic acid, and ethylene glycol were able to meet the criteria and received a score of 1, however, water was excluded from the analysis as recommended

by Hansen due to the unpredictability of its behavior [21]. The HSP sphere and the solubility parameters for cellobiose were calculated by a method proposed by Gharagheizi and Angaji [46], which uses the Nelder-Mead method [49] as a minimization algorithm to improve Hansen's fitting procedure. Figure 1 shows the HSP sphere constructed for cellobiose.

Table 2. Hansen solubility parameters and solubility score of select organic solvents. HSP values were obtained from Hansen [21].

Solvent	δ_D (MPa$^{1/2}$)	δ_P (MPa$^{1/2}$)	δ_H (MPa$^{1/2}$)	δ_t (MPa$^{1/2}$)	Score
Dimethyl sulfoxide	18.4	16.4	10.2	26.7	1
Dimethylformamide	17.4	13.7	11.3	24.9	1
Morpholine	18.8	4.9	9.2	21.5	1
Aniline	19.4	5.1	10.2	22.5	0
Formamide	17.2	26.2	19	36.7	1
Acetic acid	14.5	8	13.5	21.4	0
Dichloroacetic acid	18.2	8.1	12.2	23.4	1
Propionic acid	14.7	5.3	12.4	19.9	0
Methanol	15.1	12.3	22.3	29.6	0
Ethanol	15.8	8.8	19.4	26.5	0
1-Propanol	16	6.8	17.4	24.6	0
1-Butanol	16	5.7	15.8	23.2	0
1-Pentanol	15.9	5.9	13.9	21.9	0
Benzyl alcohol	18.4	6.3	13.7	23.8	0
Ethylene glycol	17	11	26	33.0	1
1-Octene	15.3	1	2.4	15.5	0
Ethylbenzene	17.8	0.6	1.4	17.9	0
p-Xylene	17.6	1	3.1	17.9	0
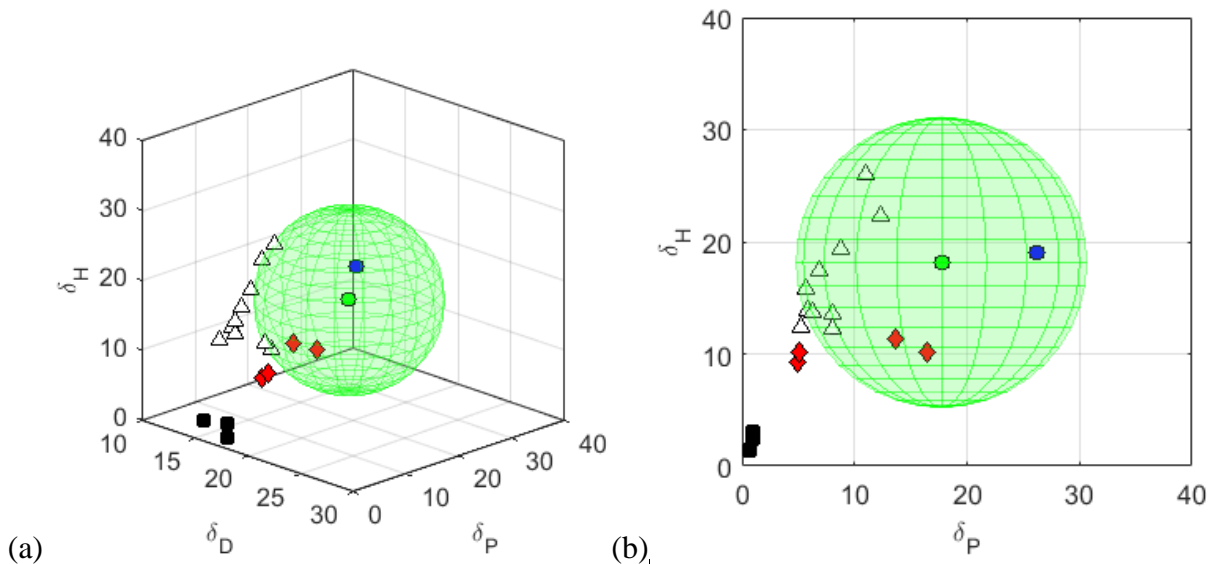
Figure 1. HSP sphere for cellobiose with 18 solvents as (a) 3D view and (b) 2D view of δ_H vs. δ_P. Polar aprotic solvents are shown as red diamonds, polar protic solvents are shown as white triangles, nonpolar solvents are shown as black squares, and formamide (can act as either hydrogen bond acceptor or donor) is shown as a blue circle. The green circle represents the center of the sphere.

The HSP for cellobiose were calculated as the following values: $\delta_D = 20.8$ MPa$^{1/2}$, $\delta_P = 17.7$ MPa$^{1/2}$, $\delta_H = 18.2$ MPa$^{1/2}$, $\delta_t = 32.8$ MPa$^{1/2}$, $R_o = 12.9$ MPa$^{1/2}$. These values are close to the HSP calculated for amorphous cellulose ($\delta_D = 24.3$ MPa$^{1/2}$, $\delta_P = 19.9$ MPa$^{1/2}$, $\delta_H = 22.5$ MPa$^{1/2}$) and sucrose ($\delta_D = 23.4$ MPa$^{1/2}$, $\delta_P = 18.4$ MPa$^{1/2}$, $\delta_H = 20.8$ MPa$^{1/2}$) by Hansen [21], so the calculated cellobiose HSP are reasonable. However, the HSP calculation method forced a 100% data fit, so morpholine had to be excluded from the sphere during the fitting procedure (score = 0) because it could not fit into the sphere without also including poor solvents such as aniline. Additionally, morpholine ($\delta_D = 18.8$ MPa$^{1/2}$, $\delta_P = 4.9$ MPa$^{1/2}$, $\delta_H = 9.2$ MPa$^{1/2}$) and aniline ($\delta_D = 19.4$ MPa$^{1/2}$, $\delta_P = 5.1$ MPa$^{1/2}$, $\delta_H = 10.2$ MPa$^{1/2}$) could be considered similar compounds based on their similar solubility parameters, however, they do not perform similarly in their abilities to dissolve cellobiose. This false negative indicates that the HSP model may not be descriptive enough to differentiate between certain compounds and, therefore, is not always reliable to predict good solvents for cellobiose.

HSP can still be used as a rough approximation for cellobiose solubility. Interestingly, *RED*, which describes how similar a solvent is to cellobiose, can be moderately correlated to cellobiose solubility. Figure 2 shows that a smaller *RED* value generally means a higher solubility while a larger *RED* value generally means a lower solubility. The best fit equation for the data is given in Eq. (22a) and the best fit equation for the data excluding morpholine is given in Eq. (22b), where S_C is the solubility of cellobiose in mole fraction. The R^2 values are 0.28 and 0.54 respectively. The solvents along the *RED* = 1 line appear to deviate the most from the linear correlation, which further suggests that the HSP model is missing important parameters necessary to accurately predict solubility. For example, benzyl alcohol (δ_D = 18.4 MPa$^{1/2}$, δ_P = 6.3 MPa$^{1/2}$, δ_H = 13.7 MPa$^{1/2}$) lies close to the edge of the HSP sphere (*RED* = 1.02) because its solubility parameters are similar to dichloroacetic acid (δ_D = 18.2 MPa$^{1/2}$, δ_P = 8.1 MPa$^{1/2}$, δ_H = 12.2 MPa$^{1/2}$) although their molecular structures and their cellobiose dissolving abilities are very different.

$$\log_{10} S_C = -4.12(RED) + 0.69 \tag{22a}$$

$$\log_{10} S_C = -5.34(RED) + 1.76 \tag{22b}$$
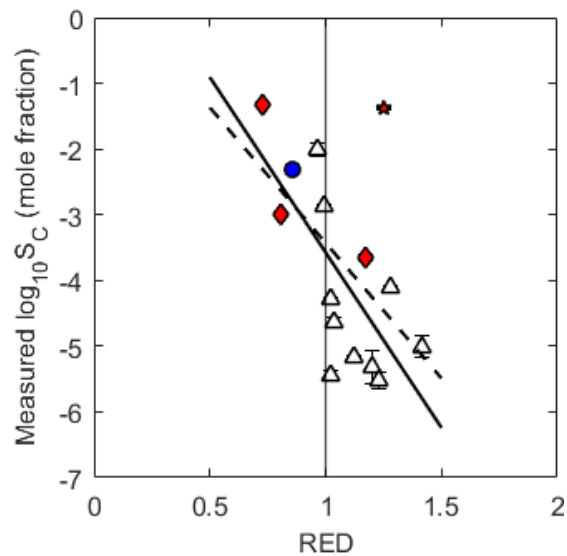
Figure 2. Correlation between RED and cellobiose solubility. The best fit line for all data points is shown as the dashed line and the best fit line excluding morpholine (shown as the red star) is shown as the solid line. The RED = 1 line is also displayed to differentiate between good and poor solvents. Polar aprotic solvents are red diamonds, polar protic solvents are white triangles, and formamide is the blue circle.

4.3. LFER Analysis

LFER is similar to HSP in the fact that both models use multiple parameters based on intermolecular forces to predict solubility. LFER descriptors for the solvents that were used in this study were found in the literature and listed in Table 3.

Table 3. LFER descriptors of the selected solvents.

Solvent	E	S	A	B	V
Dimethyl sulfoxide[a]	0.522	1.74	0.00	0.88	0.6130
Dimethylformamide[a,b]	0.367	1.31	0.00	0.74	0.6468
Formamide[b,c]	0.468	1.30	0.62	0.60	0.3650
Morpholine[c]	0.434	0.79	0.06	0.91	0.7221
Aniline[b]	0.955	0.96	0.26	0.41	0.8162
Acetic acid[a]	0.265	0.65	0.61	0.45	0.4648
Dichloroacetic acid[c,d]	0.481	1.20	0.90	0.27	0.7096
Propionic acid[b,c]	0.233	0.65	0.60	0.45	0.6057
Methanol[a,c]	0.278	0.44	0.43	0.47	0.3082
Ethanol[a,c]	0.246	0.42	0.37	0.48	0.4491
1-Propanol[a,c]	0.236	0.42	0.37	0.48	0.5900
1-Butanol[a,c]	0.224	0.42	0.37	0.48	0.7309
1-Pentanol[a,c]	0.219	0.42	0.37	0.48	0.8718
Benzyl alcohol[a,c]	0.803	0.87	0.33	0.56	0.9160
Ethylene glycol[b]	0.460	0.76	0.60	0.69	0.5078
Water[a,c]	0.000	0.45	0.82	0.35	0.1673
1-Octene[a,c]	0.094	0.08	0.00	0.07	1.1928
Ethylbenzene[a]	0.613	0.51	0.00	0.15	0.9982
p-Xylene[a]	0.613	0.52	0.00	0.16	0.9982

[a][50]
[b][30]
[c][51]
[d][52]

The LFER equation that was obtained from multiple linear regression is shown in Eq. (23), where S_C is the solubility of cellobiose in mole fraction. Table 4 shows the statistical analysis of the regressed coefficients and Table 5 shows the covariance of the descriptors. The moderate correlations between E and S, E and V, S and B, A and B, and A and V are important to note because they likely contributed to the large error of each coefficient.

$$\log_{10} S_C = -6.98 + 0.14E + 2.11S + 2.07A + 3.99B - 2.39V \tag{23}$$

The LFER equation shows that the most important descriptor is the hydrogen bond basicity (B) as it has the largest regressed coefficient (3.99). This is unsurprising because the polar aprotic solvents, which have larger B values, generally perform better than the other solvents. The polarizability (S) and molar volume (V) are also important: a larger polarizability and a smaller molar volume generally indicate a larger cellobiose solubility. The negative correlation with V agrees with the previously discussed trend of larger-sized alcohols resulting in lower cellobiose solubility. The hydrogen bond acidity (A) also appears to be important, but its high coefficient value may be due to two reasons: many more polar protic solvents (with moderate A values) were evaluated compared to polar aprotic solvents (with small A values) and dichloroacetic acid, one of the only good solvents with a large A value, may have singlehandedly enlarged the A coefficient. However, as seen in Table 5, A and B negatively co-vary, so it would actually be more desirable to have a solvent with a lower A value if it also meant a larger B value. The three nonpolar solvents are predicted to be poor solvents (dissolving 10^{-10} to 10^{-7} mole fraction of cellobiose) because they have low S, A, and B values and high V values.

Table 4. Regressed coefficients for LFER for 4 different scenarios: (a) using all data and all descriptors; (b) using all data and two descriptors; (c) using the data excluding morpholine and all descriptors; (d) using all data and nonlinear descriptors.

no. data points	AAD	ARD	Adj. R^2	Regressed coefficient values with uncertainties					
				c	e	s	a	b	v
16[a]	66%	27%	0.50	-6.98	0.14	2.11	2.07	3.99	-2.39
				±2.4	±1.7	±0.94	±1.8	±2.7	±1.8
16[b]	71%	30%	0.53	-3.89		2.65			-3.13
				±0.90		±0.66			±1.3
15[c]	46%	16%	0.75	-4.26	-0.22	3.11	0.94	-0.88	-3.22
				±1.8	±1.1	±0.69	±1.2	±2.3	±1.2
16[d]	31%	10%	0.90	-4.03	2.20	-7.42	11.19	4.13	-2.50
				±0.84	±0.62	±1.9	±1.8	±0.96	±0.69

Table 5. Correlation coefficient table for LFER descriptors of 16 solvents.

	E	S	A	B	V
E	1				
S	0.53	1			
A	-0.31	-0.29	1		
B	0.20	0.48	-0.74	1	
V	0.57	0.16	-0.43	0.14	1

A parity plot was constructed to compare predicted LFER solubilities to experimental solubilities and is shown in Figure 3(a). LFER is able to predict solubilities relatively well for most

solvents. However, morpholine is greatly under-predicted because its descriptors do not fit the LFER trend: its polarizability descriptor is not as large as other good solvents and its molar volume is larger than most other good solvents. Two additional LFER regressions were performed to confirm the analysis of the LFER descriptors and morpholine: one regression was performed using only the S and V descriptors and another regression was performed using all five descriptors for the data excluding morpholine. The parity plots for these tests are presented in Figures 3(b) and 3(c), respectively.

The first test using only the S and V descriptors was reached by sequentially removing descriptors with the least impact on LFER predictions and then rerunning the regression analysis (removed E, then A, and then B). The AAD and ARD, shown in Table 4, are comparable to the complete LFER analysis, which further supports the conclusion that S and V are the most important descriptors necessary for predicting cellobiose solubility in most solvents. Despite excluding the B descriptor in this test, it is evident from the parity plots that B is still important to improve predictions for certain outliers such as morpholine.
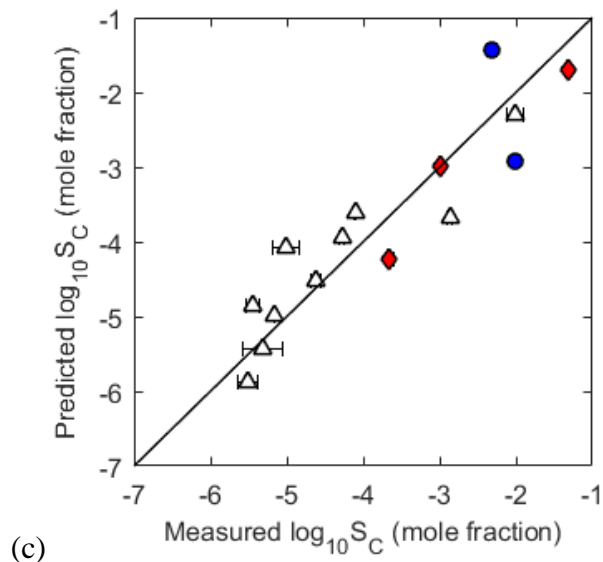

(a)   (b)

(c)
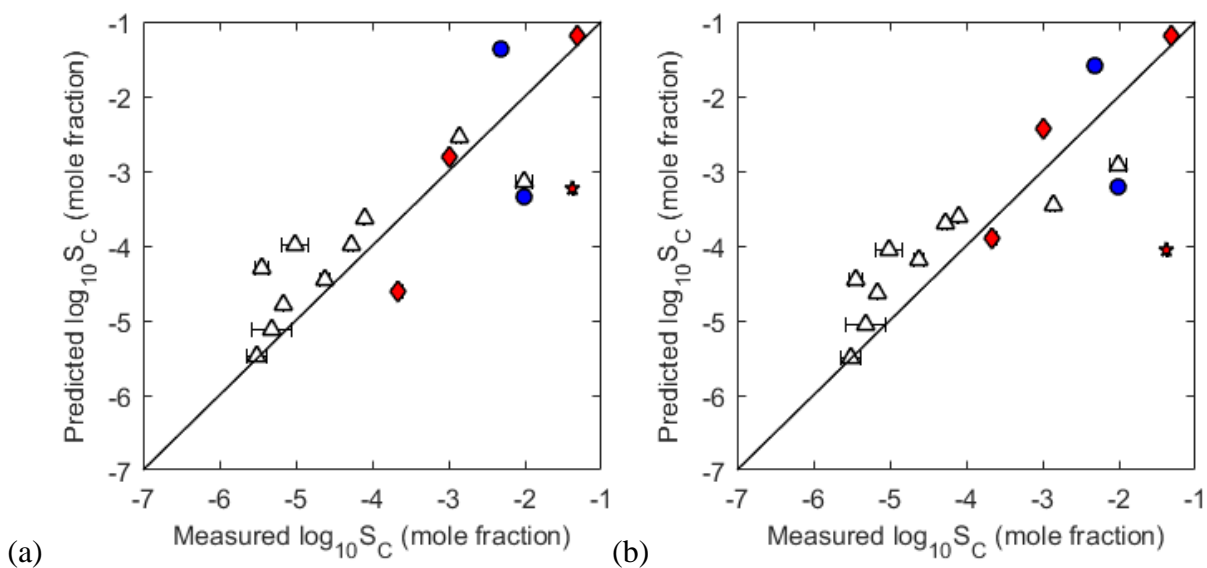
Figure 3. Parity plot comparing LFER predictions to experimental solubilities (a) for all data points in this study, (b) for all data points using only two LFER descriptors (S and V), and (c) for all data points excluding morpholine. Polar aprotic solvents are red diamonds, polar protic solvents are white triangles, amphiprotic solvents (formamide and water) are blue circles, and morpholine is shown as a red star.

The second test using all five descriptors for the data excluding morpholine was performed to observe the impact of morpholine on the LFER equation. Predictions are improved for this test; the AAD and ARD, shown in Table 4, are lower compared to the two previous LFER analyses. The new regressed coefficients for this test are also shown in Table 4. *S* and *V* have increased importance with greater certainty compared to the complete LFER analysis with all data points. *A* has less importance, suggesting that polar protic solvents such as dichloroacetic acid and ethylene glycol are good solvents because they also have high *S* values. *B* lost its importance because there are too few polar aprotic solvents compared to the number of polar protic solvents in the analysis, so this new coefficient does not carry any meaning. The large changes in each coefficient from removing one data point has two possible implications: more data is necessary for a more reliable equation and morpholine has unique properties not captured by the LFER model (and the HSP model) that are important factors that affect solubility. The molecular structure of morpholine is suspected to be a major contributor to its ability to dissolve cellobiose because tertiary amine

31

oxides such as N-methylmorpholine-N-oxide (NMMO), were specifically found to be good solvents for cellulose when mixed with water [53-55] while similar compounds such as aniline may not have the same ability to dissolve cellulose (hypothesized based on its weak ability to dissolve cellobiose).

It is slightly disappointing that the LFER model is not accurate for all solvents. Fortunately, a more complex empirical equation can be constructed based on the LFER descriptors to improve the accuracy of the model while also including morpholine. Nonlinear (squared) terms were introduced; Eq. (24) was one of the more successful models that best fit the experimental data and the regressed coefficients are listed in Table 4. There are likely many more possible equations that can fit the experimental data as well as Eq. (24), especially if more higher-order terms are included, however, the focus of this study is not to explore the infinite possible combinations of complex terms for an accurate empirical equation.

$$\log_{10} S_C = c + eE^2 + sA + aA^2 + bB^2 + vV^2 \tag{24}$$

These nonlinear terms were not derived from any theory; they were used simply to develop an accurate model that may be used to predict cellobiose solubility with less worry for outliers such as morpholine. Figure 4 shows the parity plot of the nonlinear model. The AAD is 31% and the ARD is 10%. Surprisingly, S is not included in this model, however, it is likely that it is implicitly included within the other descriptors once again due to the covariance, especially between E and S and between S and B.
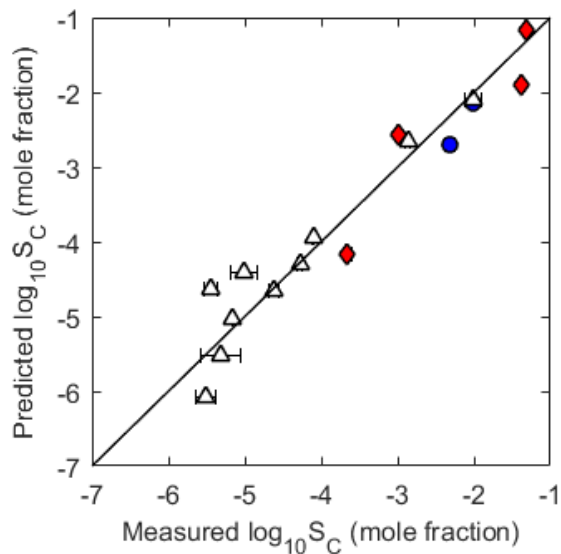
Figure 4. Parity plot comparing the nonlinear equation predictions to experimental solubilities. Polar aprotic solvents are red diamonds, polar protic solvents are white triangles, and amphiprotic solvents are blue circles.

4.4. UNIFAC and Modifications to UNIFAC

In order to calculate solubility using the SLE equation shown by Eq. (13), values for cellobiose properties (Δh_{fus}, T_m, and Δc_p) are required. Table 6 lists the selected values for these properties from the literature or group-contribution methods and also includes the calculated ideal solubility of cellobiose ($\gamma_2 = 1$) at 20°C using these property values. Sensitivity analyses were performed for each property due to the large range or unreliability of reported literature values. The justification of the selection of each property value is presented in Appendix A.

Table 6. Cellobiose properties and calculated ideal solubility.

Δh_{fus} (kJ/mol) [56]	T_m (K) [57]	Δc_p (J/mol/K) [58]	Ideal solubility at 20°C (mole fraction$\times 10^{-3}$)
73.9	510	263	0.904

UNIFAC is a useful model to study because it is semi-empirical; it can predict cellobiose solubility without experimental data. However, as mentioned previously, the accuracy of UNIFAC

depends on the accuracy of the interaction parameters regressed from equilibrium data. Many interaction parameters are currently unavailable or unreliable, which is likely one of the reasons for the low accuracy of UNIFAC for carbohydrate solubility. The parity plot comparing the predicted cellobiose solubilities from UNIFAC to experimental solubilities is presented in Figure 5(a). It shows that UNIFAC generally over-predicts most solvents to be better than what was found experimentally for cellobiose. The AAD is 130% and the ARD is 38%. Although UNIFAC could be used as a rough approximation for cellobiose solubility, it is much less accurate than the LFER model. UNIFAC predicts cellobiose solubility to be on the order of 10^{-15} to 10^{-13} mole fraction for the three nonpolar solvents, so UNIFAC has the ability to predict that cellobiose is insoluble in nonpolar solvents although it is unknown if the magnitude of the predicted value is close to the real value.

The modified UNIFAC (NIST) model was also evaluated because it is a more recently published model with more interaction parameters available [41]. This model is slightly less accurate than the original UNIFAC and it generally under-predicts most solvents to be worse than what was found experimentally. The AAD is 140% and the ARD is 49%. The large differences between the predictions of the UNIFAC models show the importance of the availability of reliable interaction parameters.

Given that the original UNIFAC generally over-predicts solubility while modified UNIFAC (NIST) generally under-predicts solubility, it would be interesting to try to counteract each model's inaccuracies by combining the two UNIFAC models. The combined model was constructed by simply taking an average of the results of the two models as shown in Eq. (25), where S_{UNIFAC} is the predicted solubility from the original UNIFAC, S_{NIST} is the predicted solubility from modified UNIFAC (NIST), and $S_{Combined}$ is the predicted solubility of the combined model.

$$\ln S_{Combined} = \frac{\ln S_{UNIFAC} + \ln S_{NIST}}{2} \qquad (25)$$

The AAD of the combined model is 73% and the ARD is 27%. Although this combined model is comparable to the LFER model, it does not solve the problems of UNIFAC's inability to distinguish between carbohydrate isomers and its usage of unreliable interaction parameters.
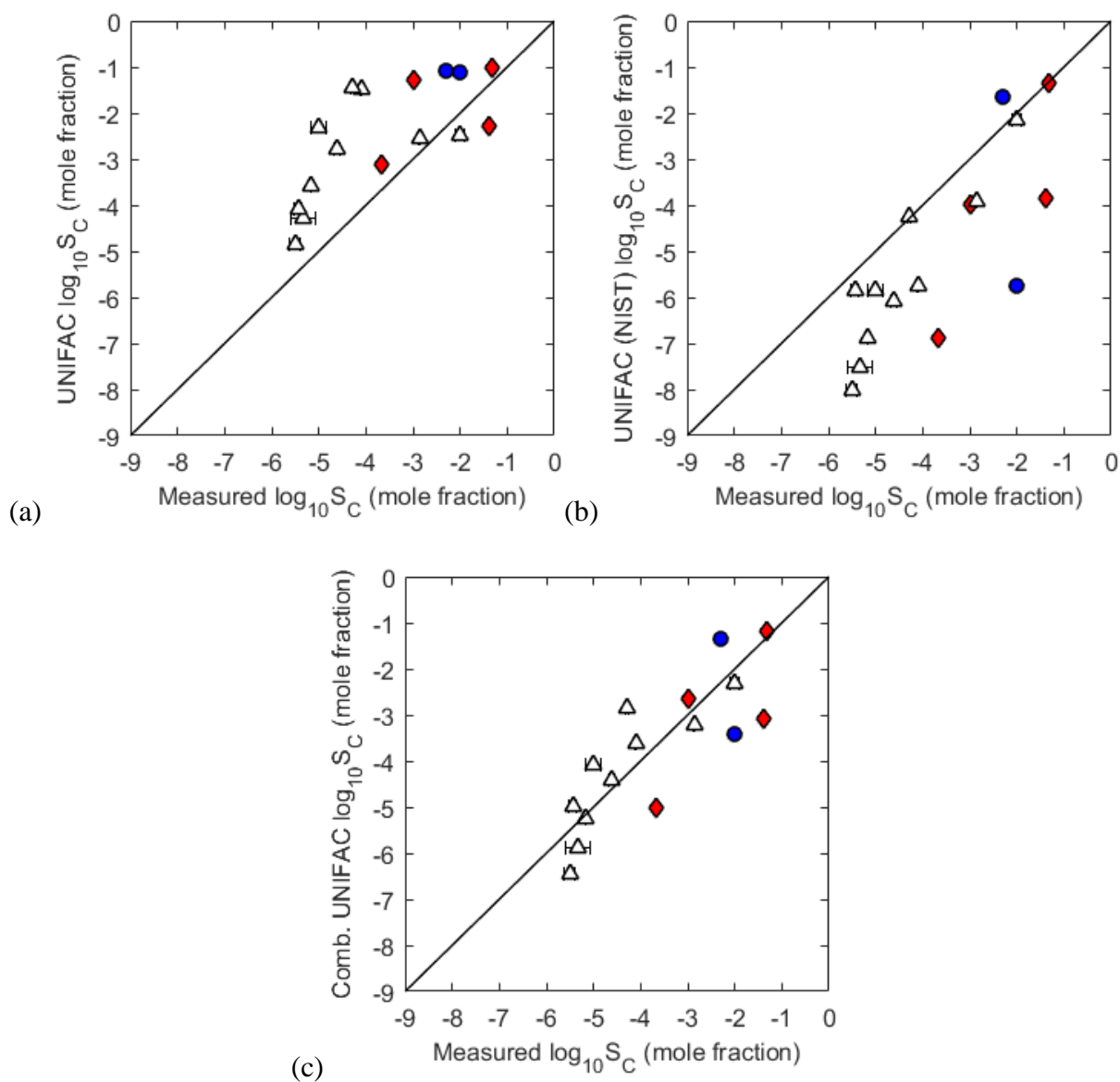

(a)

(b)

(c)

Figure 5. Parity plot comparing (a) UNIFAC predictions (b) modified UNIFAC (NIST) predictions and (c) combined (averaged UNIFAC and modified UNIFAC (NIST)) predictions to experimental solubilities. Polar aprotic solvents are red diamonds, polar protic solvents are white triangles, and amphiprotic solvents are blue circles.

A-UNIFAC, which is a modification of UNIFAC specifically for carbohydrates, provides more promising results, but only applies to water and alcohols. A comparison of UNIFAC, modified UNIFAC (NIST), and A-UNIFAC solubility predictions for applicable solvents are shown in Table 7. Benzyl alcohol was not able to be defined by the groups available for A-UNIFAC and was excluded in the comparison. Ethylene glycol, however, was included in the comparison because it can be defined by the alcohol groups (CH_2, OH). v^{OH} for ethylene glycol was set to 1.

Table 7. Comparison of UNIFAC, modified UNIFAC (NIST), and A-UNIFAC predictions for cellobiose solubility in alcohols and water. Solubilities are in mole fraction×10^{-3}.

Solvent	Experimental Solubility	UNIFAC	Mod. UNIFAC (NIST)	A-UNIFAC
Methanol	0.0524	34.9	0.0557	0.197
Ethanol	0.0237	1.70	0.000816	0.0181
1-Propanol	0.00688	0.264	0.000128	0.00676
1-Butanol	0.00476	0.0539	0.0000301	0.00314
1-Pentanol	0.00309	0.0138	0.00000934	0.00160
Ethylene Glycol	1.36	2.73	0.126	0.00124
Water	9.78	75.7	0.00185	19.9

As mentioned previously, the original UNIFAC generally over-predicts cellobiose solubilities. The AAD and ARD for UNIFAC for only alcohols and water (excluding ethylene glycol) are 150% and 35%. A-UNIFAC is able to predict solubilities in alcohols and water more accurately, which is expected due to the fitting procedure of A-UNIFAC and the extra association term to account for the self- and cross-associations in mixtures [17]. Ethylene glycol, however, is better predicted by the original UNIFAC. A-UNIFAC cannot be extended to glycols and is

therefore limited to simple alcohols and water. The AAD and ARD for A-UNIFAC excluding ethylene glycol are 25% and 6.7%, respectively. Although the A-UNIFAC parameters were not fit to cellobiose solubility data, it was able to reasonably predict cellobiose solubility.

A-UNIFAC was evaluated for ten additional carbohydrates to verify its ability to predict carbohydrate solubility. Experimental data for these carbohydrates were extracted from the literature at 18-25°C in water, methanol, and ethanol [13,20,43,44,59-62]. Property values (Δh_{fus}, T_m, and Δc_p) for each carbohydrate are listed in Appendix A. Figure 6 shows the parity plot for A-UNIFAC, which indicates that it works well for many carbohydrates and is therefore the most promising solubility model. The AAD is 22% and the ARD is 10%.
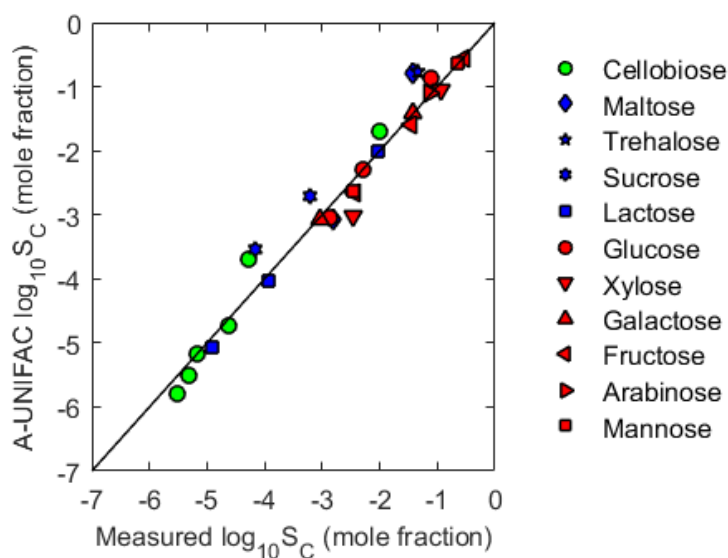
Figure 6. Parity plot comparing A-UNIFAC predictions for carbohydrate solubility in water and alcohols (primarily methanol and ethanol) to experimental solubilities found in the literature. Monomers are in red, dimers are in blue, and cellobiose data (from this study) are in green.

Additional work is required to extend the UNIFAC and A-UNIFAC models to accurately predict cellobiose solubility in other solvents. A much larger solubility database for cellobiose in additional solvents and at various temperatures will allow new interaction parameters between the new sugar groups in A-UNIFAC and more functional groups to be defined.

4.5. Model Predictions

This study contains a limited dataset for cellobiose solubility, so it may be useful to list a few additional organic compounds that could be good solvents for cellobiose based on the literature and predictions from HSP, LFER, and UNIFAC. The solubility of most carbohydrates is hypothesized to be similar or related, so solvent systems for cellulose are also expected to work well for cellobiose and other similar carbohydrates. One notable solvent system for cellulose is a solution of lithium chloride (LiCl) and dimethylacetamide (DMA or DMAc). N-methyl-2-pyrrolidone (NMP), 1,3-dimethyl-2-imidazolidinone (DMI), and hexamethylphosphoric triamide (HMPT) are a few compounds that can substitute DMAc in the solvent system [11,55]. Tertiary amine oxides were also found to be good solvents for cellulose when mixed with water. Examples include N-methylmorpholine-N-oxide (NMMO), N,N-dimethylethanolamine-N-oxide, and N-methylpiperidine-N-oxide [53-55]. Table 8 lists available HSP and LFER parameters for 11 additional organic solvents that are predicted to dissolve cellobiose well based on their use in the literature and Table 9 lists cellobiose solubility predictions for these organic solvents using the HSP, LFER, and UNIFAC models. It is important to note that pyridine, piperidine, and piperazine have similar structures (and HSP and LFER parameters) to morpholine, so these models may not be able to accurately predict solubilities for these compounds.

Table 8. HSP and LFER parameters for additional organic solvents.

Solvent	δ_D[a] (MPa$^{1/2}$)	δ_P[a] (MPa$^{1/2}$)	δ_H[a] (MPa$^{1/2}$)	E	S	A	B	V
N-methyl-2-pyrrolidone	18.0	12.3	7.2	-	-	-	-	-
1,3-dimethyl-2-imidazolidinone	-	-	-	-	-	-	-	-
Hexamethylphosphoric triamide	18.5	8.6	11.3	-	-	-	-	-
Dimethylacetamide	16.8	11.5	10.2	0.363[b]	1.33[b]	0[b]	0.78[b]	0.788[b]
Ethanolamine	17.0	15.5	21.2	-	-	-	-	-
Butylamine	16.2	4.5	8.0	0.224[c]	0.35[c]	0.16[c]	0.61[c]	0.772[c]
1,3-Butanediol	16.6	10.0	21.5	-	-	-	-	-
Diethylene glycol	16.6	12.0	20.7	-	-	-	-	-
Pyridine	19.0	8.8	5.9	0.631[c]	0.84[c]	0[c]	0.52[c]	0.675[c]
Piperidine	17.6	4.5	8.9	0.422[c]	0.46[c]	0.10[c]	0.69[c]	0.804[c]
Piperazine	18.1	5.6	8.0	0.57[b]	0.83[b]	0.11[b]	1.14[b]	0.763[b]

[a][21]
[b][30]
[c][52]

Table 9. Predicted solubility of cellobiose in additional organic solvents in mole fraction×10⁻³. For HSP predictions, Eq. (22a) was used and RED values are in parentheses. For LFER predictions, Eq. (23) was used.

Solvent	HSP	LFER	UNIFAC	Mod. UNIFAC (NIST)
N-methyl-2-pyrrolidone	0.256 (1.04)	-	112	13.3
1,3-dimethyl-2-imidazolidinone	-	-	68.1	41.6
Hexamethylphosphoric triamide	0.591 (0.95)	-	-	-
Dimethylacetamide	0.391 (0.99)	1.28	58.7	0.0501
Ethanolamine	10.1 (0.65)	-	29.3	16.1
Butylamine	0.00420 (1.47)	0.0051	3.18	0.103
1,3-Butanediol	5.49 (0.72)	-	0.676	0.0136
Diethylene glycol	2.33 (0.81)	-	1.15	0.0264
Pyridine	0.0527 (1.21)	0.0220	5.70	0.0383
Piperidine	0.0143 (1.34)	0.0123	0.00485	0.0047
Piperazine	0.0231 (1.29)	6.36	4.44	51.4

4.6. Correlation with Cellulose Swelling

Good solvents for cellobiose do not directly translate to good solvents for cellulose because the crystallinity of cellulose is a major contributor to its resistance to dissolution [7,10]. Therefore, it is also important to study the intra-crystalline and inter-crystalline swelling of cellulose, a phenomenon in which the solvent penetrates and expands cellulose fibers without dissolution, which reduces cellulose crystallinity [63]. Both cellulose dissolution and swelling involve the disruption of the supramolecular structure and are therefore controlled by the same cellulose-solvent interactions [64]. Therefore, a correlation between cellobiose solubility and cellulose swelling would be quite useful for future work involving cellulose.

Swelling data for α-cellulose (one of two distinct crystalline forms of cellulose) was found in the literature [63-66] for 13 of the solvents used in this study. There is some disagreement on

the degree of swelling in water, which also affects the swelling data point for morpholine because Onyianta et al. [66] only reported cellulose swelling in morpholine to be approximately 130% of that in water. The larger literature value for water was selected for a better correlation. Figure 7 shows a strong correlation between cellobiose solubility and cellulose swelling. The best fit line is given in Eq. (26), where *%Sw* is the percentage of swelling of cellulose. The R^2 value is 0.85.

$$\%Sw = 37.6 \log_{10} S_C + 243 \tag{26}$$
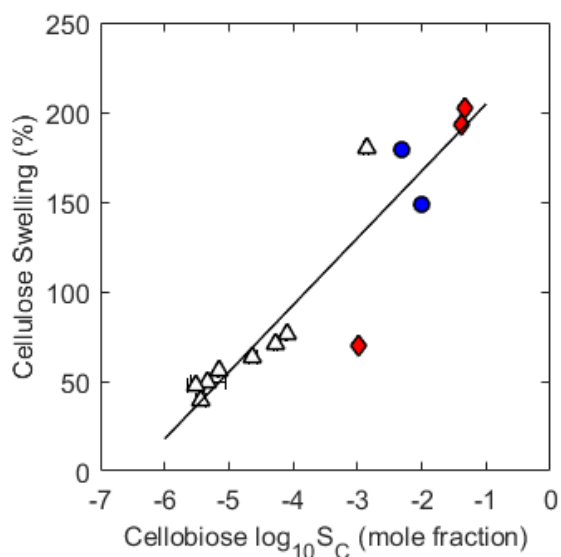
Figure 7. Correlation between cellobiose solubility and cellulose swelling.

This correlation supports the idea that solubility and swelling are related. Good solvents for cellobiose tend to also be able to swell cellulose. It is helpful to predict and find solvents that swell cellulose because disrupting the crystalline structure improves cellulose reactivity [4].

5. Conclusions and Future Work

Solubility of cellobiose in various organic solvents and water were determined. Polar aprotic solvents performed better than polar protic and nonpolar solvents, with DMSO as the best solvent out of the ones tested in this study. The HSP of cellobiose were calculated to be the following values: $\delta_D = 20.8$ MPa$^{1/2}$, $\delta_P = 17.7$ MPa$^{1/2}$, $\delta_H = 18.2$ MPa$^{1/2}$, $\delta_t = 32.8$ MPa$^{1/2}$, $R_o = 12.9$

MPa$^{1/2}$. A linear relation between solubility and *RED* was proposed as a rough numerical prediction for the HSP model. The LFER model is the most accurate model of the three that were evaluated in this work. The regressed equation shows that a more polar and a smaller-sized solvent will tend to dissolve cellobiose well. The AAD is 66% and the ARD is 27%. UNIFAC and modified UNIFAC (NIST) are much less accurate in its cellobiose solubility predictions with AAD values of 130% and 140%, respectively, and ARD values of 38% and 49%, respectively. A-UNIFAC is more accurate, but only applies to water and alcohols. The AAD is 25% and the ARD is 6.7%. These modifications to UNIFAC show that it is the model with the most potential. Future modifications to the UNIFAC equations and interaction parameters will continuously improve its accuracy in SLE predictions. Cellobiose solubility was found to have a strong correlation with cellulose swelling. Solvents that are predicted to be good for cellobiose may therefore be used in future work to improve processes involving cellulose depolymerization.

In the future, it is also recommended to perform a time study to verify or improve the procedure for measuring solubility. 24 hours was assumed to be sufficient to reach equilibrium for each solvent, however, the time it takes to reach equilibrium is different for each solvent. Morpholine was observed to be slower at dissolving cellobiose than other solvents such as DMSO. In order to confirm that equilibrium is reached, cellobiose-solvent mixtures should be analyzed after every 12-hour interval starting at 24 hours. If the concentration of the mixture remains constant for each time frame after 24 hours, then it would be acceptable to assume equilibrium has been reached after 24 hours. Additionally, it is important to measure the solution density in order to convert solubility from g L^{-1} to mole fraction.

6. References

[1] Barclay, W. B., Wheeler, M. C., Moh, A., & van Walsum, G. P. (2017). Hydrolysis of Cellulose and Glucose Using Recyclable α-Hydroxysulfonic Acids. *Industrial & Engineering Chemistry Research, 56*(44), 12529-12537. doi:10.1021/acs.iecr.7b03355

[2] Isikgor, F. H., & Becer, C. R. (2015). Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polymer Chemistry, 6*(25), 4497-4559. doi:10.1039/C5PY00263J

[3] McKendry, P. (2002). Energy production from biomass (part 1): overview of biomass. *Bioresource Technology, 83*(1), 37-46. doi:https://doi.org/10.1016/S0960-8524(01)00118-3

[4] Tyufekchiev, M., Kolodziejczak, A., Duan, P., Foston, M., Schmidt-Rohr, K., & Timko, M. T. (2019). Reaction engineering implications of cellulose crystallinity and water-promoted recrystallization. *Green Chemistry*. doi:10.1039/C9GC02466B

[5] Olsson, C., & Westman, G. (2013). Direct dissolution of cellulose: background, means and applications. *Cellulose-Fundamental Aspects, 10*, 52144.

[6] Sun, Y., Sun, K., Zhang, L., Zhang, S., Liu, Q., Wang, Y., . . . Hu, X. (2020). Impacts of Solvents on the Stability of the Biomass-Derived Sugars and Furans. *Energy & Fuels, 34*(3), 3250-3261. doi:10.1021/acs.energyfuels.9b03921

[7] Pinkert, A., Marsh, K. N., & Pang, S. (2010). Reflections on the Solubility of Cellulose. *Industrial & Engineering Chemistry Research, 49*(22), 11121-11130. doi:10.1021/ie1006596

[8] Klemm, D., Heublein, B., Fink, H.-P., & Bohn, A. (2005). Cellulose: Fascinating Biopolymer and Sustainable Raw Material. *Angewandte Chemie International Edition, 44*(22), 3358-3393. doi:10.1002/anie.200460587

[9] Medronho, B., & Lindman, B. (2014). Competing forces during cellulose dissolution: From solvents to mechanisms. *Current Opinion in Colloid & Interface Science, 19*(1), 32-40. doi:https://doi.org/10.1016/j.cocis.2013.12.001

[10] Ghasemi, M., Alexandridis, P., & Tsianou, M. (2017). Cellulose dissolution: insights on the contributions of solvent-induced decrystallization and chain disentanglement. *Cellulose, 24*(2), 571-590. doi:10.1007/s10570-016-1145-1

[11] Liebert, T. (2010). Cellulose Solvents – Remarkable History, Bright Future. In *Cellulose Solvents: For Analysis, Shaping and Chemical Modification* (Vol. 1033, pp. 3-54): American Chemical Society.

[12] Kobayashi, H., Yabushita, M., Hasegawa, J.-y., & Fukuoka, A. (2015). Synergy of Vicinal Oxygenated Groups of Catalysts for Hydrolysis of Cellulosic Molecules. *The Journal of Physical Chemistry C, 119*(36), 20993-20999. doi:10.1021/acs.jpcc.5b06476

[13] Jónsdóttir, S. Ó., Cooke, S. A., & Macedo, E. A. (2002). Modeling and measurements of solid–liquid and vapor–liquid equilibria of polyols and carbohydrates in aqueous solution. *Carbohydrate Research, 337*(17), 1563-1571. doi:https://doi.org/10.1016/S0008-6215(02)00213-6

[14] Kononenko, O. K., & Herstein, K. M. (1956). Nonaqueous Solvents for Sucrose. *Industrial & Engineering Chemistry Chemical & Engineering Data Series, 1*(1), 87-92. doi:10.1021/i460001a017

[15]   Catté, M., Dussap, C.-G., Achard, C., & Gros, J.-B. (1994). Excess properties and solid-liquid equilibria for aqueous solutions of sugars using a UNIQUAC model. *Fluid Phase Equilibria, 96*, 33-50. doi:https://doi.org/10.1016/0378-3812(94)80086-3

[16]   Catté, M., Dussap, C.-G., & Gros, J.-B. (1995). A physical chemical UNIFAC model for aqueous solutions of sugars. *Fluid Phase Equilibria, 105*(1), 1-25. doi:https://doi.org/10.1016/0378-3812(94)02604-Y

[17]   Ferreira, O., Brignole, E. A., & Macedo, E. A. (2003). Phase Equilibria in Sugar Solutions Using the A-UNIFAC Model. *Industrial & Engineering Chemistry Research, 42*(24), 6212-6222. doi:10.1021/ie030246n

[18]   Kuramochi, H., Noritomi, H., Hoshino, D., & Nagahama, K. (1997). Representation of activity coefficients of fundamental biochemicals in water by the UNIFAC model. *Fluid Phase Equilibria, 130*(1), 117-132. doi:https://doi.org/10.1016/S0378-3812(96)03209-8

[19]   Peres, A. M., & Macedo, E. A. (1997). A modified UNIFAC model for the calculation of thermodynamic properties of aqueous and non-aqueous solutions containing sugars. *Fluid Phase Equilibria, 139*(1), 47-74. doi:https://doi.org/10.1016/S0378-3812(97)00196-9

[20]   Spiliotis, N., & Tassios, D. (2000). A UNIFAC model for phase equilibrium calculations in aqueous and nonaqueous sugar solutions. *Fluid Phase Equilibria, 173*(1), 39-55. doi:https://doi.org/10.1016/S0378-3812(00)00387-3

[21]   Hansen, C. M. (2007). *Hansen solubility parameters: a user's handbook*: CRC press.

[22]   Blanks, R. F., & Prausnitz, J. M. (1964). Thermodynamics of Polymer Solubility in Polar and Nonpolar Systems. *Industrial & Engineering Chemistry Fundamentals, 3*(1), 1-8. doi:10.1021/i160009a001

[23]   Stefanis, E., & Panayiotou, C. (2008). Prediction of Hansen Solubility Parameters with a New Group-Contribution Method. *International Journal of Thermophysics, 29*(2), 568-585. doi:10.1007/s10765-008-0415-z

[24]   Sato, T., Araki, S., Morimoto, M., Tanaka, R., & Yamamoto, H. (2014). Comparison of Hansen Solubility Parameter of Asphaltenes Extracted from Bitumen Produced in Different Geographical Regions. *Energy & Fuels, 28*(2), 891-897. doi:10.1021/ef402065j

[25]   Vitha, M., & Carr, P. W. (2006). The chemical interpretation and practice of linear solvation energy relationships in chromatography. *Journal of Chromatography A, 1126*(1), 143-194. doi:https://doi.org/10.1016/j.chroma.2006.06.074

[26]   Abraham, M. H., Ibrahim, A., & Zissimos, A. M. (2004). Determination of sets of solute descriptors from chromatographic measurements. *Journal of Chromatography A, 1037*(1), 29-47. doi:https://doi.org/10.1016/j.chroma.2003.12.004

[27]   Abraham, M. H., & McGowan, J. C. (1987). The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. *Chromatographia, 23*(4), 243-246. doi:10.1007/BF02311772

[28]   Abraham, M. H. (1993). Application of solvation equations to chemical and biochemical processes. In *Pure and Applied Chemistry* (Vol. 65, pp. 2503).

[29]   Platts, J. A., Butina, D., Abraham, M. H., & Hersey, A. (1999). Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach. *Journal of Chemical Information and Computer Sciences, 39*(5), 835-845. doi:10.1021/ci980339t

[30]   Zissimos, A. M., Abraham, M. H., Klamt, A., Eckert, F., & Wood, J. (2002). A Comparison between the Two General Sets of Linear Free Energy Descriptors of Abraham and Klamt. *Journal of Chemical Information and Computer Sciences, 42*(6), 1320-1331. doi:10.1021/ci025530o

[31]   Prausnitz, J. M., Lichtenthaler, R. N., & De Azevedo, E. G. (1998). *Molecular thermodynamics of fluid-phase equilibria*: Pearson Education.

[32]   Fredenslund, A., Jones, R. L., & Prausnitz, J. M. (1975). Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal, 21*(6), 1086-1099. doi:10.1002/aic.690210607

[33]   Skjøld-Jørgensen, S., Kolbe, B., Gmehling, J., & Rasmussen, P. (1979). Vapor—liquid equilibria by UNIFAC group contribution. Revision and extension. *Ind. Eng. Chem. Process Des. Dev., 18*(4), 714-722.

[34]   Gmehling, J., Rasmussen, P., & Fredenslund, A. (1982). Vapor-liquid equilibriums by UNIFAC group contribution. Revision and extension. 2. *Industrial & Engineering Chemistry Process Design and Development, 21*(1), 118-127. doi:10.1021/i200016a021

[35]   Macedo, E. A., Weidlich, U., Gmehling, J., & Rasmussen, P. (1983). Vapor-liquid equilibriums by UNIFAC group contribution. Revision and extension. 3. *Industrial & Engineering Chemistry Process Design and Development, 22*(4), 676-678. doi:10.1021/i200023a023

[36]   Tiegs, D., Rasmussen, P., Gmehling, J., & Fredenslund, A. (1987). Vapor-liquid equilibria by UNIFAC group contribution. 4. Revision and extension. *Industrial & Engineering Chemistry Research, 26*(1), 159-161. doi:10.1021/ie00061a030

[37]   Rasmussen, P., Fredenslund, A., Schiller, M., Gmehling, J., & Hansen, H. (1991). Vapor-Liquid Equilibria by UNIFAC Group Contribution. 5. *Revision and Extension. Ind. Eng. Chem. Res, 30*(10), 2352.

[38]   Wittig, R., Lohmann, J., & Gmehling, J. (2003). Vapor−Liquid Equilibria by UNIFAC Group Contribution. 6. Revision and Extension. *Industrial & Engineering Chemistry Research, 42*(1), 183-188. doi:10.1021/ie020506l

[39]   Weidlich, U., & Gmehling, J. (1987). A modified UNIFAC model. 1. Prediction of VLE, hE, and .gamma..infin. *Industrial & Engineering Chemistry Research, 26*(7), 1372-1381. doi:10.1021/ie00067a018

[40]   Larsen, B. L., Rasmussen, P., & Fredenslund, A. (1987). A modified UNIFAC group-contribution model for prediction of phase equilibria and heats of mixing. *Industrial & Engineering Chemistry Research, 26*(11), 2274-2286. doi:10.1021/ie00071a018

[41]   Kang, J. W., Diky, V., & Frenkel, M. (2015). New modified UNIFAC parameters using critically evaluated phase equilibrium data. *Fluid Phase Equilibria, 388*, 128-141. doi:https://doi.org/10.1016/j.fluid.2014.12.042

[42]   Mengarelli, A. C., Brignole, E. A., & Bottini, S. B. (1999). Activity coefficients of associating mixtures by group contribution. *Fluid Phase Equilibria, 163*(2), 195-207. doi:https://doi.org/10.1016/S0378-3812(99)00227-7

[43]   Montañés, F., Olano, A., Ibáñez, E., & Fornari, T. (2007). Modeling solubilities of sugars in alcohols based on original experimental data. *AIChE Journal, 53*(9), 2411-2418. doi:10.1002/aic.11258

[44]   Gong, X., Wang, C., Zhang, L., & Qu, H. (2012). Solubility of Xylose, Mannose, Maltose Monohydrate, and Trehalose Dihydrate in Ethanol–Water Solutions. *Journal of Chemical & Engineering Data, 57*(11), 3264-3269. doi:10.1021/je300885g

[45]   Taylor, J. B. (1957). The water solubilities and heats of solution of short chain cellulosic oligosaccharides. *Transactions of the Faraday Society, 53*(0), 1198-1203. doi:10.1039/TF9575301198

[46]   Gharagheizi, F. (2007). New procedure to calculate the Hansen solubility parameters of polymers. *Journal of Applied Polymer Science, 103*(1), 31-36. doi:10.1002/app.23874

[47]   Saeed (2019). UNIFAC group contribution method activity calculator function (https://www.mathworks.com/matlabcentral/fileexchange/64885-unifac-group-contribution-method-activity-calculator-function), MATLAB Central File Exchange. Retrieved August 29, 2019.

[48]   Moye, C. J. (1972). Non-Aqueous Solvents for Carbohydrates. In R. S. Tipson & D. Horton (Eds.), *Advances in Carbohydrate Chemistry and Biochemistry* (Vol. 27, pp. 85-125): Academic Press.

[49]   Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization, 9*(1), 112-147. doi:10.1137/S1052623496303470

[50]   Abraham, M. H., Andonian-Haftvan, J., Whiting, G. S., Leo, A., & Taft, R. S. (1994). Hydrogen bonding. Part 34. The factors that influence the solubility of gases and vapours in water at 298 K, and a new method for its determination. *Journal of the Chemical Society, Perkin Transactions 2*(8), 1777-1791. doi:10.1039/P29940001777

[51]   Abraham, M. H. (1993). Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chemical Society Reviews, 22*(2), 73-83. doi:10.1039/CS9932200073

[52]   Abraham, M. H., Chadha, H. S., Whiting, G. S., & Mitchell, R. C. (1994). Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the Δlog p parameter of seiler. *Journal of Pharmaceutical Sciences, 83*(8), 1085-1100. doi:10.1002/jps.2600830806

[53]   Franks, N. E., & Varga, J. K. (1979). Process for making precipitated cellulose. In: Google Patents.

[54]   Franks, N. E., & Varga, J. K. (1980). Process for making a shapeable cellulose and shaped cellulose products. In: Google Patents.

[55]   Johnson, D. (1985). Solvents for cellulose. *Cellulose chemistry and its applications*, 181-201.

[56]   Marrero, J., & Gani, R. (2001). Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria, 183-184*, 183-208. doi:https://doi.org/10.1016/S0378-3812(01)00431-9

[57]   Pinto, S. S., Diogo, H. P., Nunes, T. G., & Moura Ramos, J. J. (2010). Molecular mobility studies on the amorphous state of disaccharides. I—thermally stimulated currents and differential scanning calorimetry. *Carbohydrate Research, 345*(12), 1802-1807. doi:https://doi.org/10.1016/j.carres.2010.05.023

[58]   Miller, D. P., & de Pablo, J. J. (2000). Calorimetric Solution Properties of Simple Saccharides and Their Significance for the Stabilization of Biological Structure and Function. *The Journal of Physical Chemistry B, 104*(37), 8876-8883. doi:10.1021/jp000807d

[59]   Gong, X., Wang, S., & Qu, H. (2011). Solid-Liquid Equilibria of D-Glucose, D-Fructose and Sucrose in the Mixture of Ethanol and Water from 273.2 K to 293.2 K. *Chinese Journal of Chemical Engineering, 19*(2), 217-222. doi:https://doi.org/10.1016/S1004-9541(11)60157-2

[60]   Gray, M. C., Converse, A. O., & Wyman, C. E. (2003). Sugar Monomer and Oligomer Solubility. In B. H. Davison, J. W. Lee, M. Finkelstein, & J. D. McMillan (Eds.),

Biotechnology for Fuels and Chemicals: The Twenty-Fourth Symposium (pp. 179-193). Totowa, NJ: Humana Press.

[61]   Crestani, C. E., Bernardo, A., Costa, C. B. B., & Giulietti, M. (2013). Fructose Solubility in Mixed (Ethanol + Water) Solvent: Experimental Data and Comparison among Different Thermodynamic Models. *Journal of Chemical & Engineering Data, 58*(11), 3039-3045. doi:10.1021/je400471m

[62]   Hunziker, O. F., & Nissen, B. H. (1926). Lactose Solubility and Lactose Crystal Formation: I. Lactose Solubility. *Journal of Dairy Science, 9*(6), 517-537. doi:https://doi.org/10.3168/jds.S0022-0302(26)93924-6

[63]   Mantanis, G. I., Young, R. A., & Rowell, R. M. (1995). Swelling of compressed cellulose fiber webs in organic liquids. *Cellulose, 2*(1), 1-22. doi:10.1007/BF00812768

[64]   Fidale, L. C., Ruiz, N., Heinze, T., & Seoud, O. A. E. (2008). Cellulose Swelling by Aprotic and Protic Solvents: What are the Similarities and Differences? *Macromolecular Chemistry and Physics, 209*(12), 1240-1254. doi:10.1002/macp.200800021

[65]   El Seoud, O. A., Fidale, L. C., Ruiz, N., D'Almeida, M. L. O., & Frollini, E. (2008). Cellulose swelling by protic solvents: which properties of the biopolymer and the solvent matter? *Cellulose, 15*(3), 371-392. doi:10.1007/s10570-007-9189-x

[66]   Onyianta, A. J., Dorris, M., & Williams, R. L. (2018). Aqueous morpholine pre-treatment in cellulose nanofibril (CNF) production: comparison with carboxymethylation and TEMPO oxidisation pre-treatment methods. *Cellulose, 25*(2), 1047-1064. doi:10.1007/s10570-017-1631-0

[67]   Hatakeyama, T., Yoshida, H., Nagasaki, C., & Hatakeyama, H. (1976). Differential scanning calorimetric studies on phase transition of glucose and cellulose oligosaccharides. *Polymer, 17*(7), 559-562. doi:https://doi.org/10.1016/0032-3861(76)90268-8

[68]   Slade, L., & Levine, H. (1988). Non-equilibrium behavior of small carbohydrate-water systems. In *Pure and Applied Chemistry* (Vol. 60, pp. 1841).

[69]   Raemy, A., & Schweizer, T. (1983). Thermal behaviour of carbohydrates studied by heat flow calorimetry. *Journal of Thermal Analysis and Calorimetry, 28*(1), 95-108. doi:10.1007/BF02105282

[70]   Joback, K. G., & Reid, R. C. (1987). ESTIMATION OF PURE-COMPONENT PROPERTIES FROM GROUP-CONTRIBUTIONS. *Chemical Engineering Communications, 57*(1-6), 233-243. doi:10.1080/00986448708960487

[71]   Wu, M., & Yalkowsky, S. (2009). Estimation of the Molar Heat Capacity Change on Melting of Organic Compounds. *Industrial & Engineering Chemistry Research, 48*(2), 1063-1066. doi:10.1021/ie801587c

[72]   Mohan, M., Banerjee, T., & Goud, V. V. (2016). Solid Liquid Equilibrium of Cellobiose, Sucrose, and Maltose Monohydrate in Ionic Liquids: Experimental and Quantum Chemical Insights. *Journal of Chemical & Engineering Data, 61*(9), 2923-2932. doi:10.1021/acs.jced.5b00914

[73]   Chickos, J. S., Braton, C. M., Hesse, D. G., & Liebman, J. F. (1991). Estimating entropies and enthalpies of fusion of organic compounds. *The Journal of Organic Chemistry, 56*(3), 927-938. doi:10.1021/jo00003a007

[74]   Jain, A., Yang, G., & Yalkowsky, S. H. (2004). Estimation of Melting Points of Organic Compounds. *Industrial & Engineering Chemistry Research, 43*(23), 7618-7621. doi:10.1021/ie049378m

[75]   Pappa, G. D., Voutsas, E. C., Magoulas, K., & Tassios, D. P. (2005). Estimation of the Differential Molar Heat Capacities of Organic Compounds at Their Melting Point. *Industrial & Engineering Chemistry Research, 44*(10), 3799-3806. doi:10.1021/ie048916s

[76]   Roos, Y. (1993). Melting and glass transitions of low molecular weight carbohydrates. *Carbohydrate Research, 238*, 39-48. doi:https://doi.org/10.1016/0008-6215(93)87004-C

[77]   Parks, G. S., Snyder, L. J., & Cattoir, F. R. (1934). Studies on Glass. XI. Some Thermodynamic Relations of Glassy and Alpha-Crystalline Glucose. *The Journal of Chemical Physics, 2*(9), 595-598. doi:10.1063/1.1749540

[78]   Orford, P. D., Parker, R., & Ring, S. G. (1990). Aspects of the glass transition behaviour of mixtures of carbohydrates of low molecular weight. *Carbohydrate Research, 196*, 11-18. doi:https://doi.org/10.1016/0008-6215(90)84102-Z

[79]   Bondi, A. A. (1968). Physical properties of molecular crystals liquids, and glasses.

[80]   Ferreira, O., Macedo, E. A., & Bottini, S. B. (2005). Extension of the A-UNIFAC model to mixtures of cross- and self-associating compounds. *Fluid Phase Equilibria, 227*(2), 165-176. doi:https://doi.org/10.1016/j.fluid.2004.11.009

7. Appendices

7.1. Appendix A: Sensitivity Analysis for Cellobiose Properties

Most literature values for T_m of cellobiose range between 498 K and 522 K [57,60,67,68]. Raemy and Schweizer [69] report 468 K as the onset temperature and 493 K as the peak temperature, but this source will be excluded as an outlier. Group-contribution methods estimate T_m to be 513 K [56] and 768 K [70]. The sensitivity range was chosen to be 495 K to 525 K. For $\Delta h_{fus} = 73.9$ kJ/mol and $\Delta c_p = 263$ J/mol/K, the ideal solubility (mole fraction) of cellobiose varies from 7.81×10^{-4} at $T_m = 495$ K to 1.11×10^{-3} at $T_m = 525$ K. T_m has only a small effect on ideal solubility, so it was sufficient to select $T_m = 510$ K [57] because it is at the midpoint of the range of literature values and closely matches the group-contribution estimation by Marrero and Gani [56] of 513 K, which uses a third-order approximation.

The literature reports Δc_p to be 263 J/mol/K [58] for cellobiose. Additionally, Δc_p was reported to be 132.2 J/mol/K [57] for the glass transition and was estimated to be 130 J/mol/K using a group-contribution method by Wu et al. [71]. The sensitivity range was chosen to be 0 to 300 J/mol/K to compare these values to the ideal (and often assumed) value of $\Delta c_p = 0$. For $T_m = 510$ K and $\Delta h_{fus} = 73.9$ kJ/mol, the ideal solubility (mole fraction) of cellobiose varies from 2.52×10^{-6} at $\Delta c_p = 0$ J/mol/K to 2.07×10^{-3} at $\Delta c_p = 300$ J/mol/K. Ideal solubility is quite sensitive to Δc_p, therefore, $\Delta c_p = 0$ cannot be assumed for cellobiose. $\Delta c_p = 263$ J/mol/K was reasonable to use because it is the only available literature value; $\Delta c_p = 132.2$ J/mol/K was reported for the glass transition and not the solid-liquid phase change.

Δh_{fus} for cellobiose and other saccharides with more than one monosaccharide unit is difficult to measure experimentally because they often decompose easily before melting [13,57]. The literature reports two different values for the enthalpy of fusion for cellobiose: 31.1 kJ/mol [60,72] and 54.8 kJ/mol [13,69]. Several group-contribution methods [56,70,73,74] were

performed and ranged from 55.7 kJ/mol [74] to 96.5 kJ/mol [73]. The sensitivity range was chosen to be 30 to 100 kJ/mol. For $T_m = 510$ K and $\Delta c_p = 263$ J/mol/K, the ideal solubility (mole fraction) varies from 1.92 at $\Delta h_{fus} = 30$ kJ/mol to 9.52×10^{-6} at $\Delta h_{fus} = 100$ kJ/mol. Ideal solubility is most sensitive to Δh_{fus}, so it is important to select a reasonable value for Δh_{fus}. The ideal solubility exceeded 1 for low Δh_{fus} values, therefore, the literature value of 31.1 kJ/mol was unreasonable to use.

Another way to calculate Δh_{fus} is to use the entropy of fusion (Δs_{fus}) in the following relation [60,73]:

$$\Delta h_{fus} = T_m \Delta s_{fus} \qquad (A.1)$$

Δc_p has also been proposed to be related to Δs_{fus} at the melting temperature [71,75]. Pappa et al. [75] proposed an empirical correlation between Δc_p and Δs_{fus} for each group of compounds (alkanes, alcohols, etc.). Assuming that a similar correlation can be found for most carbohydrates, literature values for Δc_p and Δs_{fus} for a few common carbohydrates were found in order to determine an empirical correlation. Property values of common carbohydrates are listed in Table A.1. Δc_p is approximately twice the value of Δs_{fus} for each of the four listed carbohydrates. If this correlation also applies to cellobiose, then the literature value of 54.8 kJ/mol is most likely too low because the Δc_p-to-Δs_{fus} ratio would be 2.45.

The group-contribution method using a third-order approximation by Marrero and Gani [56] estimates Δh_{fus} to be 73.9 kJ/mol for cellobiose. This estimated value is the best fit for the Δc_p and Δs_{fus} correlation with a ratio of 1.82 and is therefore an acceptable value to use for cellobiose solubility calculations. However, it is important to note that the group-contribution method cannot distinguish between carbohydrate isomers and does not account for the effect of the glycosidic bond and therefore may not be accurate. Nevertheless, this value will be used

because the group-contribution method was able to closely estimate T_m and therefore may also be able to estimate close to the actual Δh_{fus} value. Additionally, Jónsdóttir et al. [13] suspected that 54.8 kJ/mol was an inaccurate value because it did not fit well in the UNIQUAC model they used and found that 70 kJ/mol was a better fit to the model. Similarly, the UNIFAC model used in this study found 73.9 kJ/mol as a better fit than 54.8 kJ/mol.

Table A.1 shows two additional trends. First, it appears that a larger T_m value tends to correlate with a larger Δh_{fus} value. Second, Δc_p for monosaccharides (glucose and fructose) appear to have similar values near 130 J/mol/K and Δc_p for disaccharides (sucrose and maltose) appear to have similar values near 250 J/mol/K. Due to the large T_m value of cellobiose compared to other carbohydrates, it is reasonable to expect a Δh_{fus} value larger than the Δh_{fus} value for sucrose, which further supports the use of the $\Delta h_{fus} = 73.9$ kJ/mol. It is also reasonable to use Δc_p = 263 J/mol/K because it is similar to Δc_p values for other disaccharides.

Table A.1. Properties of common carbohydrates.

	Δh_{fus} (J/mol)	T_m (K)	Δc_p (J/mol/K)	$\Delta s_{fus} = \frac{\Delta h_{fus}}{T_m}$ (J/mol/K)	$\frac{\Delta c_p}{\Delta s_{fus}}$
Glucose	32 248[a]	423.15[b]	138[c]	76.2	1.81
Fructose	26 030[d,e]	378.15[d,e]	135[a]	68.9	1.96
Sucrose	57 000[f]	459.15[f]	254[d,g]	124.1	2.05
Maltose	45 400[d]	379.15[d]	231[d,g]	119.7	1.93
Trehalose	48 048[d]	368.15[d]	241[d,g]	130.5	1.85
Xylose	31 700[h]	416.15[h]	97[h]	76.2	1.27
Galactose	43 778[d]	436.15[d]	139[h]	100.4	1.38
Arabinose	35 700[i]	433.15[i]	120[j]	82.4	1.46
Mannose	24 687[d]	407.15[d]	130[a]	60.6	2.15
Lactose	75 306[d]	474.15[d]	239[d,g]	158.8	1.51
Cellobiose	73 900[k]	510[l]	263[g]	144.9	1.82

[a][76]
[b][19]
[c][77]
[d][17]
[e][61]
[f][20]
[g][58]
[h][13]
[i][60]
[j][78]
[k][56]
[l][57]

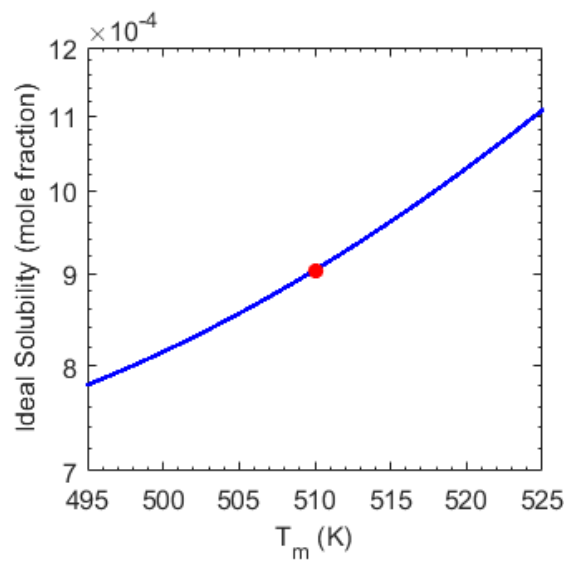*Figure A.1. Sensitivity of the melting temperature T_m on the solubility of cellobiose.*
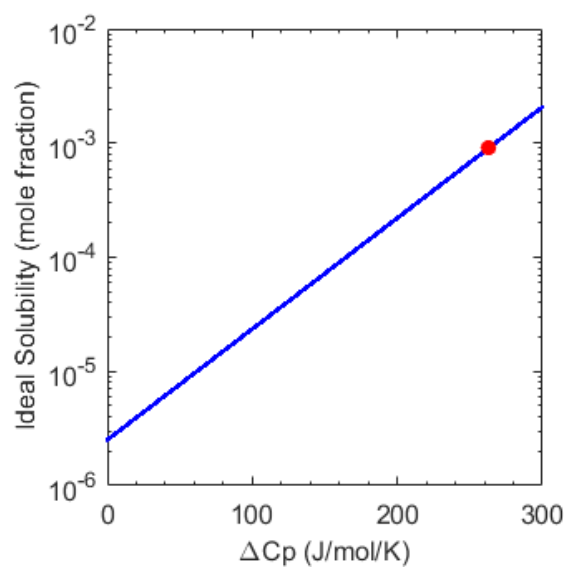

Figure A.2. Sensitivity of the heat capacity Δc_p on the solubility of cellobiose.
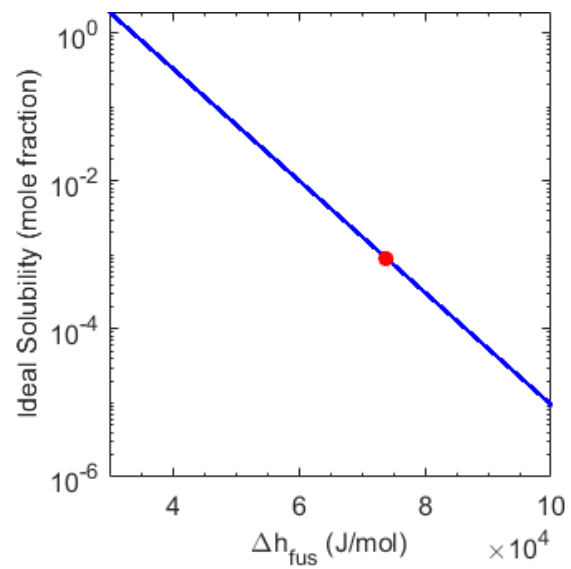
Figure A.3. Sensitivity of the heat of fusion Δh$_{fus}$ on the solubility of cellobiose.

7.2. Appendix B: UNIFAC Equations and Parameters

7.2.1. Original UNIFAC Equations

The activity coefficient consists of a combinatorial part and a residual part in the UNIFAC model. Eqs. (B.1-3) are used to calculate the combinatorial part, where x_i is the mole fraction of component i in the mixture, θ_i is the area fraction, and Φ_i is the segment fraction (similar to the volume fraction).

$$\ln \gamma_i^C = \ln \frac{\Phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\theta_i}{\Phi_i} + l_i - \frac{\Phi_i}{x_i} \Sigma_j x_j l_j \tag{B.1}$$

$$l_i = \frac{z}{2}(r_i - q_i) - (r_i - 1) \tag{B.2}$$

$$z = 10 \tag{B.3}$$

Eqs. (B.4-5) show the calculations for θ_i and Φ_i, where r_i is the molecular (van der Waals) volume parameter and q_i is the molecular surface area parameter.

$$\theta_i = \frac{q_i x_i}{\Sigma_j q_j x_j} \tag{B.4}$$

$$\Phi_i = \frac{r_i x_i}{\Sigma_j r_j x_j} \tag{B.5}$$

r_i and q_i are calculated by group-contribution volume and surface area parameters R_k and Q_k, respectively, where v_k^i is the number of groups of type k in molecule i.

$$r_i = \Sigma_k v_k^i R_k \tag{B.6}$$

$$q_i = \Sigma_k v_k^i Q_k \tag{B.7}$$

R_k and Q_k are calculated from the van der Waals group volume and surface areas V_{wk} and A_{wk}, respectively, which are given by Bondi [79]. R_k and Q_k are available in the literature and listed in the Dortmund Databank.

$$R_k = \frac{V_{wk}}{15.17} \tag{B.8}$$

$$Q_k = \frac{A_{wk}}{2.5 \times 10^9} \qquad \text{(B.9)}$$

The residual part of the activity coefficient is calculated by Eq. (B.10), where Γ_k is the functional group residual activity coefficient and Γ_k^i is the activity coefficient of group k in a reference solution containing only molecules of type i. Γ_k^i is necessary to have the activity coefficient γ_i reach unity as x_i approaches 1.

$$\ln \gamma_i^R = \sum_k v_k^i \left[\ln \Gamma_k - \ln \Gamma_k^i \right] \qquad \text{(B.10)}$$

Eq. (B.11) is used to calculate Γ_k and Γ_k^i, where Θ_m is the area fraction of group m and is calculated similarly to θ_i with X_m as the mole fraction of group m in the mixture.

$$\ln \Gamma_k = Q_k \left[1 - \ln \sum_m \Theta_m \Psi_{mk} - \sum_m \left(\frac{\Theta_m \Psi_{km}}{\sum_n \Theta_n \Psi_{nm}} \right) \right] \qquad \text{(B.11)}$$

$$\Theta_m = \frac{Q_m X_m}{\sum_n Q_n X_n} \qquad \text{(B.12)}$$

The group interaction parameter Ψ_{mn} is given in Eq. (B.13), where U_{mn} is a measure of energy of interaction between groups m and n. a_{mn} and a_{nm}, where $a_{mn} \neq a_{nm}$, represent the two group interaction parameters for a binary mixture, which is evaluated from experimental phase equilibrium data and is often tabulated.

$$\Psi_{mn} = \exp \left[-\left(\frac{U_{mn} - U_{nn}}{RT} \right) \right] = \exp \left(-\frac{a_{mn}}{T} \right) \qquad \text{(B.13)}$$

7.2.2. A-UNIFAC Equations and Parameters

The expression to calculate the activity coefficient for component i using the A-UNIFAC model is given in Eq. (B.14).

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R + \ln \gamma_i^A \qquad \text{(B.14)}$$

The combinatorial and residual parts are calculated the same way as the original UNIFAC model by Fredenslund et al. [32]. The association part is calculated using Eq. (B.15), where $v^{OH,i}$ is the number of OH associating groups in component i.

57

$$\ln \gamma_i^A = v^{OH,i} \left[2 \ln \left(\frac{X_{OH}}{X_{OH,i}} \right) + \left(X_{OH,i} - X_{OH} \right) \right] - (1 - X_{OH})(v^{OH,i} - r_i \rho_{OH}) \qquad \text{(B.15)}$$

The fraction of nonbonded sites in the mixture (X_{OH}) and the fraction of nonbonded sites in pure component i ($X_{OH,i}$) are calculated by Eqs. (B.16-17).

$$X_{OH} = \frac{-1 + \sqrt{1 + 4\rho_{OH}\Delta_{OH}}}{2\rho_{OH}\Delta_{OH}} \qquad \text{(B.16)}$$

$$X_{OH,i} = \frac{1 + \sqrt{1 + 4(\rho_{OH})_i \Delta_{OH}}}{2(\rho_{OH})_i \Delta_{OH}} \qquad \text{(B.17)}$$

ρ_{OH} is the concentration of the associating group in the mixture and $(\rho_{OH})_i$ is the concentration of the associating group in pure component i. Eqs. (B.18-19) are used to calculate these parameters, where x_i is the molar fraction of component i in the mixture and r_i is the UNIQUAC molecular volume of component i, which is calculated by summing the UNIQUAC volumes of each functional group in component i.

$$\rho_{OH} = \frac{\sum_{i=1}^{NC} v^{OH,i} x_i}{\sum_{i=1}^{NC} r_i x_i} \qquad \text{(B.18)}$$

$$(\rho_{OH})_i = \frac{v^{OH,i}}{r_i} \qquad \text{(B.19)}$$

Δ_{OH}, the dimensionless association strength, is a function of the energy of association (ϵ_{OH}/k) and the volume of association (κ_{OH}). For OH and H_2O self-associations and cross-association, $\kappa_{OH} = 0.0062$ and $\epsilon_{OH}/k = 3125$ K [17,80].

$$\Delta_{OH} = \kappa_{OH} \left[\exp \left(\frac{\epsilon_{OH}}{kT} \right) - 1 \right] \qquad \text{(B.20)}$$

v^{OH} is another parameter that was estimated simultaneously with the group interaction parameters. This parameter for cellobiose was set to the same value as sucrose of 4.3 due to the lack of available data, as suggested by Ferreira et al. [17]. v^{OH} was reasonably set to 1 for water, methanol, and ethanol and v^{OH} was set to 0.5 for propanol, butanol, and pentanol because Montañés et al. [43]

had to use a value of 0.5 for v^{OH} for 1-propanol and isopropanol. v^{OH} was set to 1 (rather than 2) for ethylene glycol simply because it was a slightly better fit to the experimental data.

Table B.1. Volume and area parameters for the new sugar groups and existing alcohol groups.

	R_k	Q_k
PYR1	2.4784	1.380
PYR2	2.7059	1.692
FUR1	1.8041	0.924
FUR2	2.0315	1.152
-O-	0.2439	0.240
OH$_{ring}$	1.0000	1.200
CH$_3$	0.9011	0.8480
CH$_2$	0.6744	0.5400
OH	1.0000	1.2000
CH$_3$OH	1.4311	1.4320
H$_2$O	0.9200	1.4000

Table B.2. Functional group composition of each compound.

	PYR1	-O-	CH$_3$	CH$_2$	OH	CH$_3$OH	OH$_{ring}$	H$_2$O
Cellobiose	2	1	0	2	0	0	8	0
Water	0	0	0	0	0	0	0	1
Methanol	0	0	0	0	0	1	0	0
Ethanol	0	0	1	1	1	0	0	0
1-Propanol	0	0	1	2	1	0	0	0
1-Butanol	0	0	1	3	1	0	0	0
1-Pentanol	0	0	1	4	1	0	0	0
Ethylene glycol	0	0	0	2	2	0	0	0

Table B.3. Group interaction parameters (K) [43].

a_{nm}	PYR/FUR	-O-	CH_2/CH_3	OH	CH_3OH	H_2O	OH_{ring}
PYR/FUR	0.0	0.0	0.0	176.5	-33.8	-154.3	0.0
-O-	0.0	0.0	0.0	-721.0	-323.5	-508.0	0.0
CH_2/CH_3	0.0	0.0	0.0	50.4	122.7	308.5	-60.2
OH	387.4	-876.6	387.4	0.0	110.9	-127.3	72.2
CH_3OH	-197.2	-278.7	-19.78	60.2	0.0	-167.6	31.9
H_2O	108.4	155.3	136.8	70.7	251.2	0.0	87.8
OH_{ring}	0.0	0.0	703.4	715.1	681.8	-174.4	0.0

7.3. Appendix C: HPLC Chromatograms for Cellobiose Solubility
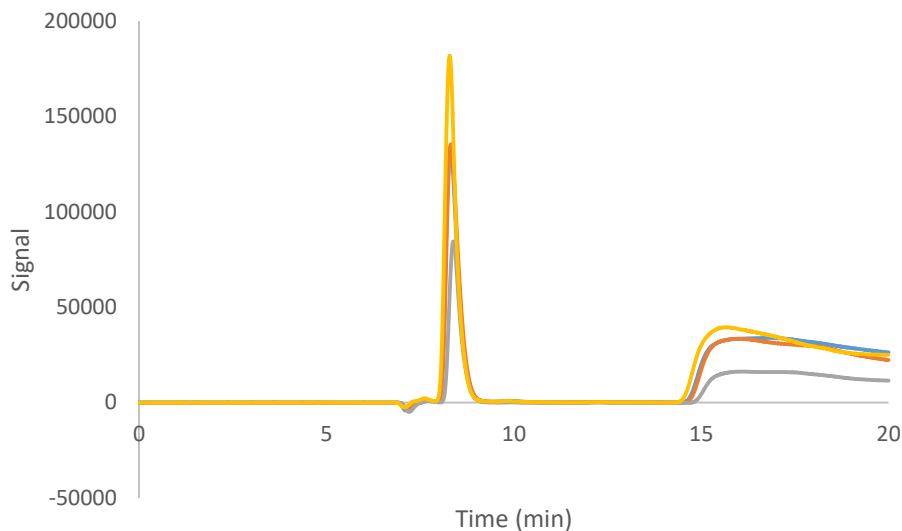

Figure C.1. HPLC chromatogram for formamide samples. Chromatograms for other solvents such as DMSO, DMF, etc. were similar and not included. Cellobiose peak is at t = 8.3 min.
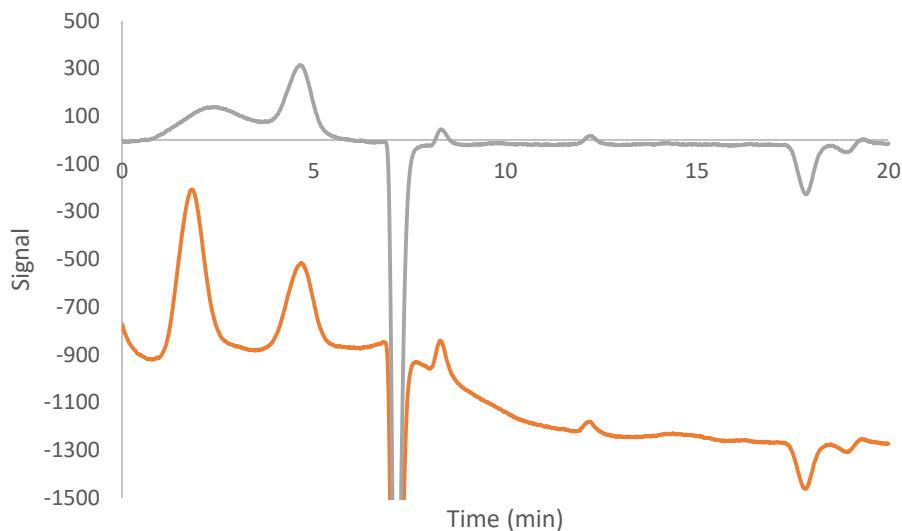

Figure C.2. HPLC chromatogram for 1-pentanol samples. Baselines were not perfect, but still provided reasonable results. Chromatograms for 1-propanol, 1-butanol, and benzyl alcohol also had slightly imperfect baselines and were not included. Cellobiose peak is at t = 8.3 min.

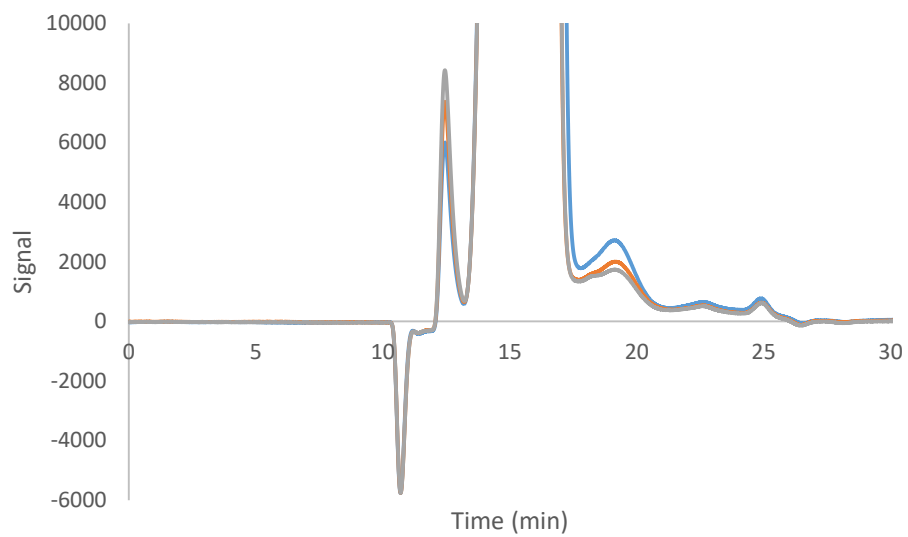*Figure C.3. HPLC chromatogram for dichloroacetic acid samples. Cellobiose peak is at t = 12.4 min; HPLC flow rate was lowered to separate the cellobiose and solvent peaks.*