Modelling the effects of ethanol on the solubility of the proteinogenic amino acids with the NRTL, Gude and Jouyban-Acree models

Bowden, N. A., Sevillano, D. M., Sanders, J. P. M., & Bruins, M. E.

This is a "Post-Print" accepted manuscript, which has been published in "Fluid Phase Equilibria"

Please cite this publication as follows:

Bowden, N. A., Sevillano, D. M., Sanders, J. P. M., & Bruins, M. E. (2018). Modelling the effects of ethanol on the solubility of the proteinogenic amino acids with the NRTL, Gude and Jouyban-Acree models. Fluid Phase Equilibria, 459, 158-169. DOI: 10.1016/j.fluid.2017.11.036

You can download the published version at:

https://doi.org/10.1016/j.fluid.2017.11.036

Modelling the effects of ethanol on the solubility of the proteinogenic amino acids with the NRTL, Gude and Jouyban-Acree models

Nathan A. Bowden (1, 2), David Mendez Sevillano (2), Johan P.M. Sanders (1, 3), Marieke E. Bruins (3)

1) Biobased Chemistry and Technology Group, ASFG, Bornse Weilanden 9, 6708WG Wageningen, Wageningen UR, The Netherlands

2) School of Life Sciences, Lovensdijkstraat 61-63, 4818AJ Breda, Avans University of Applied Science, The Netherlands

3) Food and Biobased Research, Bornse Weilanden 9, 6107WG Wageningen, The Netherlands

Abstract: *The addition of organic solvents, such as ethanol, to molecules in solution is an effective process for crystallization and is used in industrial settings (i.e. pharmaceutical production, downstream processing, etc.). In this study, we use solubility data of all proteinogenic α-amino acids in binary ethanol/water systems to model their excess solubility. We use the empirical and regressive models of Gude and NRTL and the predictive Jouyban-Acree model. Based on the results, we hypothesize that amino acids that are spherical and lack a reactive side chain show little or no excess solubility. Being rod-like and/or having a reactive side chain leads to a positive excess solubility in a mixed solvent of ethanol and water. The empirical and regressed models, NRTL and Gude, fit the data well and the predictive Jouyban-Acree model, not originally intended to be used for small molecules, is less accurate but offers insights into the thermodynamic properties of the amino acids.*

Keywords: Thermodynamics, aqueous-solutions, equilibria, organic solvents, excess solubility

1 Introduction

In the future, products that are currently being produced using non-renewable resources (e.g. plastics, pharmaceuticals and fine chemicals) could be made from bio-based sources, such as proteins and α-amino acids[1-3]. One of the challenges in this line of research, is to find a way to separate α-amino acids from industrial residues so that the production of bio-based products can begin. This research is applicable to the industrial challenges of separating amino acids from solution.

Industrial residues can be used as a feedstock for the extraction of amino acids and other biomolecules. When amino acids are extracted, they need to be separated from aqueous solution.

Currently, the most common method of separating many amino acids from solution is by using industrial chromatography. An alternative to chromatography could be to crystallize the amino acids using an anti-solvent, such as ethanol.

The structure of every amino acid contains a carboxyl group attached to an α-carbon. This α-carbon is also attached to an amino group. The amino acids studied in this article are α-amino acids, which all have side chains also attached to the α-carbon. The exception is glycine which does not have a side chain. The side chains of α-amino acids include aliphatic groups, aromatic and non-aromatic rings, hydroxyl groups, sulphur and charged groups (e.g. a second carboxyl group, lysyl group, guanidinium group). The amino and carboxyl groups attached to the α-carbon will be charged at a pH that is not the isoelectric point. At the isoelectric point, the amino acid has a neutral charge and is called a zwitterion. All measurements in this manuscript were taken at the isoelectric point.

There has been some research on the solubility of α-amino acids in mixtures of alcohol and water[4-7]. Basic solubility measurements were reported and subsequent research focused on calculating the partition coefficients of the solubility of these α-amino acids and their phase behavior[8]. Recently, complete and reliable data has been published on the solubility of α-amino acids in ethanol/water systems[9] and mixtures of α-amino acids[10].

Many models have been proposed to model the solubility of amino acids in aqueous solution. These models include calculating partition coefficients[11], using regressed coefficients[12], examining non-ideality[13], measuring and modelling activity coefficients[14-17], activities[18] and applying a modification of the Wilson model[19]. Other models have been applied to model the solubility of amino acids in salt solutions[20-27]. Only a few models have been proposed to describe the solubility of α-amino acids in ethanol/water systems, but these manuscripts focus on a single

model and only a few α-amino acids[28-30]. This article will model all proteinogenic α-amino acids using solubility data that is available in the literature.

We use three models that represent two different modelling approaches. Of these three, two of the models use regressed parameters. The models that we use that have regressed parameters are the Gude model and the Non-Random Two Liquid (NRTL) model. While models that use regressed parameters have in general given excellent results, they do not explain what thermodynamic properties of the molecules lead to their results. The third model that we use is the Jouyban-Acree model, which is a predictive model. Predictive solubility models are based on thermodynamic properties of the molecules that they are modelling. While the thermodynamic properties of the molecules explain the results of the predictive models, predictive models have been less accurate than regressed models.

Using the different approaches allows conclusions to be made on whether the predictive model (Jouban-Acree) provides sufficient accuracy to model amino acid solubility or if a regressed model (Gude or NRTL) should be used. Other solubility models[31-36] were considered for this article, but due to their complexity were left out in favour of models with fewer variables.

The Gude[12] and NRTL[54] models were chosen in this research for their accuracy in the literature and the minimum number of parameters they use. Both the NRTL and Gude models furthermore acknowledge the lattice and therefore entropic nature of liquids, first investigated by Flory[37] and Huggins[38]. The Gude model has one parameter that is regressed to fit the data and the NRTL has two parameters that are regressed to fit the data. For this reason, it is expected that the NRTL model will have a lower error. However, it is preferential to use a regressive model with

the least number of regressed parameters. In the case where both models have similar errors, the Gude model could be used.

While the Gude and NRTL models will be accurate, in comparison, the Jouyban-Acree model is predictive and based on the bonds and forces of the molecules being modelled. The version of the Jouyban-Acree model that is used in this research has nine regressed constants. These constants are used in conjunction with Hansen solubility parameters, which are based on physical chemistry group contribution data. While the Jouyban-Acree model uses more parameters than the Gude and NRTL models, the parameters are predictive, not regressed. The Jouyban-Acree model has been shown to perform well with relatively large pharmaceutical solutes in ternary systems[39]. A version of this model with regressed parameters has been applied to only a few amino acids in ternary solution, but no α-amino acids in water and ethanol mixtures, with the exception of glycine[40]. We use the Jouyban-Acree model without regressed parameters in this research in order to evaluate the use of group contribution data to amino acid solubility models. In the future, data from this work could contribute to refining the non-regressed Jouyban-Acree parameters for amino acids.

2 Theory

2.1 Thermodynamic modelling of excess solubility

The addition of organic solvents, e.g. ethanol, to aqueous solutions of amino acids lowers the solubility of the amino acid solutes. This allows for precipitation and crystallization. The solubility of the amino acids is often lowered by organic solvents by more than 1000 times its solubility in water alone[9]. Industrial applications using organic solvents can only be designed

when this effect on the solubility is understood. This presents a challenge for chemical engineers in modelling their solubility.

Data is taken from the literature[4-7, 9] and modelled with two empirical and regressive models and with one predictive model. The two empirical and regressive models are the Gude[12] and NRTL[41-45] models and the semi-empirical and predictive model is the Jouyban-Acree model[46-50].

In order to effectively compare the performance of the models, excess solubility has been chosen as the output of the model. This decision aligns with literature[51-52] in the specific case of binary solvent mixtures. Excess solubility, represented by the mole fraction x_{aa}^{E}, can be calculated using equation (1).

$$ln\, x_{aa}^{E} \equiv ln\, x_{aa,mix} - \sum_{i=1}^{N} x_i'\, ln\, x_{aa,i} \quad (1)$$

in which case $x_{aa,mix}$ and $x_{aa,i}$ are the mole fractions of the amino acid solute *(aa)* in a mixed solvent and pure solvent, *i*, respectively. The mole fraction of the solvent *i* without the solute is denoted by x_i'.

When assuming a pure solvent phase as a standard state, such as in this research, at standard system pressure and temperature, the chemical potential of the solute is not dependent on the solvent composition. Therefore, the excess solubility can be rewritten as:

$$ln\, x_{aa}^{E} \equiv -ln\, \gamma_{aa,mix} + \sum_{i=1}^{N} x_i'\, ln\, \gamma_{aa,i} \quad (2)$$

where the dimensionless activity coefficients of the solute in saturated solutions of the mixed solvent and pure solvent are represented by $\gamma_{aa,mix}$ and $\gamma_{aa,i}$.

Cohn and Edsall[53] noted that the solubility of the solute in these systems is low. Therefore, it can be assumed that the solute is infinitely dilute and approximated as:

$$ln\, x_{aa}^{E} \equiv -ln\, \gamma_{aa,mix}^{\infty} + \sum_{i=1}^{N} x_i'\, ln\, \gamma_{aa,i}^{\infty} \quad (3)$$

2.2 Gude Model

Gude[6] developed a simplified equation to model the behaviour of amino acids in mixed solvents. This model uses 2 constants. The constant for the interaction between the solvents, $A_{j,i}$, was set to 1.55 for ethanol/water in the work of Gude and is applied in this work. The constant for the interaction between the amino acid and the solvent mixture, $C_{j,i,aa}$, is specific to each amino acid. This interaction parameter, $C_{j,i,aa}$ (mol·L^{-1}), is constant for the system and found by fitting the model to the data. Equation (4) describes the model:

$$ln\, x_{aa}^{E} \equiv ln\, \mathrm{r}' - \sum_{j=1}^{N} x_j'\, ln\, \mathrm{r}_j + r_{aa}\left(\frac{1}{r'} - \sum_j \frac{x'_j}{r_j}\right) + \sum_j \sum_i \left[A_{j,i} x'_j x'_i \left(1 + C_{j,i,aa}\right)\right] \quad (4)$$

where subscripts *j* and *i* relate to solvents and subscript *aa* relates to the solute. The values of the UNIFAC variable r were set at 0.92 for water and 2.11 for ethanol and calculated individually for the amino acids[12]. Values for r′ are the solute free value of r. The $C_{j,i,aa}$ are fitted for each amino acid from Equation (4) and are shown in Table 2.

2.3 NRTL Model

Based on the hypothesis of Wilson, that the local concentration of solvent molecules in a two-solvent system around a molecule of the solute are not the same as the concentration in the solution in general, Renon and Prausnitz[54] developed the NRTL model to calculate the interaction parameters between these molecules. In the case of this research, the mixed solvent is comprised of only two solvents, so the activity coefficient $\gamma_{aa,i}$ equation (5):

$$ln\,\gamma_{aa,i} = \frac{\sum_{i=1}^{n} x_i' \tau_{i,aa} G_{i,aa}}{\sum_{i=1}^{n} x_i' G_{i,aa}} + \sum_{i=1}^{n} \frac{x_i' G_{aa,i}}{\sum_{j=1}^{n} x_j' G_{j,i}} \left(\tau_{aa,i} - \frac{\sum_{j=1}^{n} x_j' \tau_{j,i} G_{j,i}}{\sum_{j=1}^{n} x_j' G_{j,i}} \right) \quad (5)$$

can be substituted with the NRTL equation, which yields equation (6):

$$\ln x_{aa}^{E} = \sum_{i=1}^{N} \left(\tau_{i,aa} + \tau_{i,aa} G_{i,aa} \right) x'_i - \frac{\sum_{i=1}^{n} x_i' \tau_{i,aa} G_{i,aa}}{\sum_{i=1}^{n} x_i' G_{i,aa}} - \sum_{i=1}^{n} \frac{x_i' G_{aa,i}}{\sum_{j=1}^{n} x_j' G_{j,i}} \left(\tau_{aa,i} - \frac{\sum_{j=1}^{n} x_j' \tau_{j,i} G_{j,i}}{\sum_{j=1}^{n} x_j' G_{j,i}} \right) \quad (6)$$

where $G_{mn} = exp(-\propto_{mn} \tau_{mn})$ and the dimensionless interaction parameters τ_{mn}, τ_{nm} and the non-randomness parameter $\propto_{nm}$ are represented for each system of two solvents.

The interaction parameters, τ, and the non-randomness parameters, $\propto$, for the solvents have previously been published[42]. These are $\tau_{ethanol,water} = -406.47$ and $\tau_{water,ethanol} = 1413$ at 298.15K, $\propto_{water,ethanol} = 0.1830$ and $\propto_{aa,water} = 0.05$ and $\propto_{aa,ethanol} = 0.02$. Furthermore, in this research we have assumed that the unitless interaction parameters for the

system amino acid-solvent, $\tau_{aa,i}$, and solvent-amino acid, $\tau_{i,aa}$, are the same. The $\tau_{aa,i}$ for each amino acid is calculated by regressing Equation (6) and are shown in Table 2.

2.4 Jouyban-Acree Model

Jouyban and colleagues developed a model for the excess solubility[39] based on the log-linear model developed by the group of professor Sadowski[32]. This model uses as input the Hansen solubility parameters which can be calculated from group contribution models[55].

There are several versions of the Jouyban-Acree model. The version that we use here[49], shown in equation (7), uses nine previously regressed constants that can be found in Table 1 to calculate the solubility in the mixture of solvents. Once that is calculated, equation (1) can be used to calculated the excess solubility and compare the performance with the aforementioned models.

$$\log x_{aa,mix} = f_c \log x_{aa,c} + f_w \log x_{aa,w} + \left(\frac{f_c f_w}{T}\right)\left[A_0\delta_{d,aa}\left(\delta_{d,c}-\delta_{d,w}\right)^2 + A_1\delta_{p,aa}\left(\delta_{p,c}-\delta_{p,w}\right)^2 + A_2\delta_{hb,aa}\left(\delta_{hb,c}-\delta_{hb,w}\right)^2\right] + \left(\frac{f_c f_w (f_c-f_w)}{T}\right)\left[A_3\delta_{d,aa}\left(\delta_{d,c}-\delta_{d,w}\right)^2 + A_4\delta_{p,aa}\left(\delta_{p,c}-\delta_{p,w}\right)^2 + A_5\delta_{hb,aa}\left(\delta_{hb,c}-\delta_{hb,w}\right)^2\right] + \left(\frac{f_c f_w (f_c-f_w)^2}{T}\right)\left[A_6\delta_{d,aa}\left(\delta_{d,c}-\delta_{d,w}\right)^2 + A_7\delta_{p,aa}\left(\delta_{p,c}-\delta_{p,w}\right)^2 + A_8\delta_{hb,aa}\left(\delta_{hb,c}-\delta_{hb,w}\right)^2\right] \quad (7)$$

Where subscripts w, c, p, d and hb stand for water, co-solvent, polar, dispersion and hydrogen bonding respectively. Furthermore, $\delta$and f stand for the Hansen solubility parameter, in $MPa^{0.5}$, and volume fraction respectively. The Hansen solubility parameters were calculated as discussed previously and are shown in Table 2. The solubility parameters are constant and could

be included in the A values. The A parameters show the effect of the forces in the solvent system on the amino acid. In this case, the solvent system in water and ethanol. The solubility parameters, A_0-A_8, are shown in Table 1.

Table 1: Jouyban-Acree constants

Constant	Value
A_0	0.0000
A_1	0.6060
A_2	0.0130
A_3	-8.6960
A_4	0.3760
A_5	0.0130
A_6	9.2770
A_7	-0.4610
A_8	0.0170

3 Materials and Methods

Matlab version 9.0.0341360 was used for the regression and calculations. All graphical objects in Figure 3 were obtained from Wikimedia and have been released to the public domain worldwide.

The data from the literature that is used in all of the models is shown in the supplementary data. In this table, the solubility of each of the 20 proteinogenic amino acids in mole fraction is given, along with the ethanol mole fraction in the solvent without the solute, the standard deviation (labelled “+/-”) and the source of the data. The standard deviation was calculated by the root of the sum of the square of the difference between each of the measurements and the average of the measurements, divided by the number of measurements minus one. All data were measured at the isoelectric point. This means that the amino acids are present as neutral zwitterions and therefore carry no net charge..

The interaction parameters of the NRTL and Gude models are regressed by minimizing the normalized root-mean-square error (NRMSE). The NRMSE was calculated for all three models by equation (8), where x_i' is the mole fraction of ethanol in the solute free solvent, $\hat{y}_{x_i'}$ is the predicted excess solubility, $y_{x_i'}$ is the measured excess solubility and y_{max} and y_{min} are the maximum and minimum excess solubility. Normalizing the root-mean-square-error by dividing by the range facilitates the comparison between amino acids that are on different scales.

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{x_i'} - y_{x_i'})^2}{n}}}{y_{max} - y_{min}} \quad (8)$$

4 Results and discussion

The regression coefficients, $\tau_{aa,i}$ and $\tau_{i,aa}$, of the NRTL model for the interaction between the amino acid and ethanol and the amino acid and water are shown in Table 2. The regression coefficients of the Gude model for each amino acid, $C_{j,i,aa}$, are also shown in Table 2. These coefficients were calculated by minimizing the NRMSE of the excess solubility values that were modelled to the excess solubility measured. The Jouyban-Acree parameters that were calculated are shown in Table 2.

The modelled fits of the Gude and NRTL models and the application of the Jouyban-Acree model are shown along with the data points in Figure 1-20 for all 20 proteinogenic amino acids. If the standard deviation of the data was available, this was included in the figures. If multiple data were available for ethanol mole fractions of 0.000 and 1.000, then preference was given to the data that has been shown to be more accurate[5]. A fit where the excess solubility was equal to 0 was added to each of the amino acids in Figures 1-20 to guide the eye.

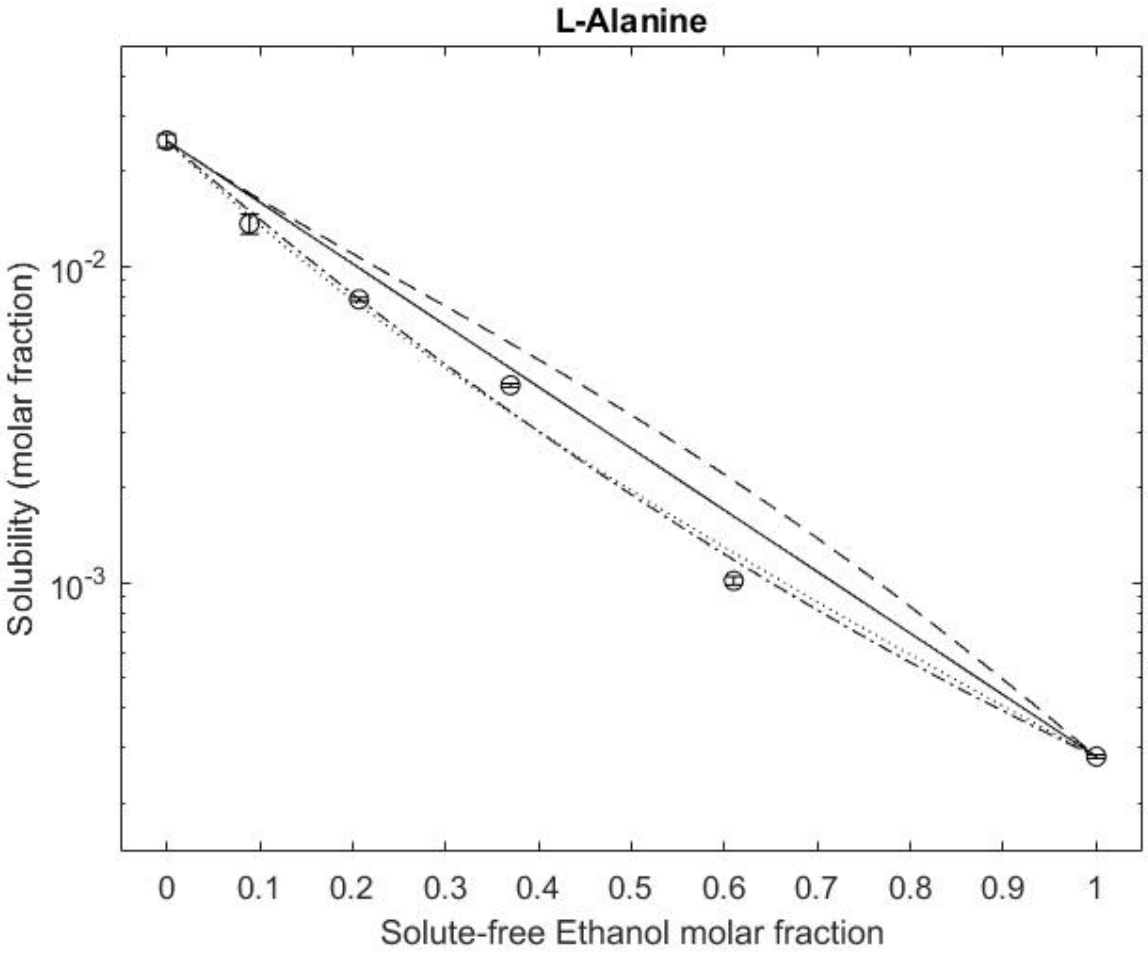


Figure 1: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).

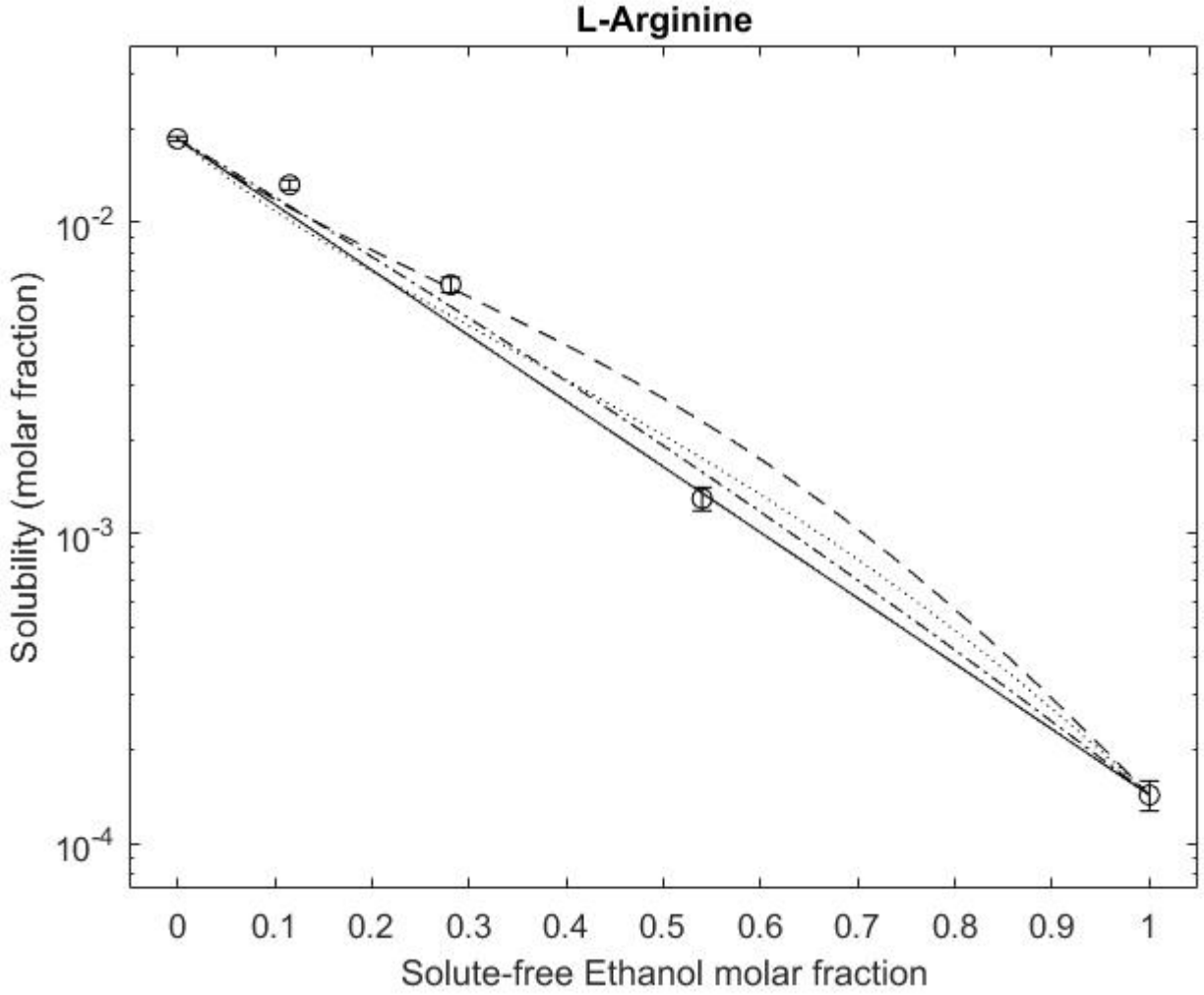


Figure 2: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).

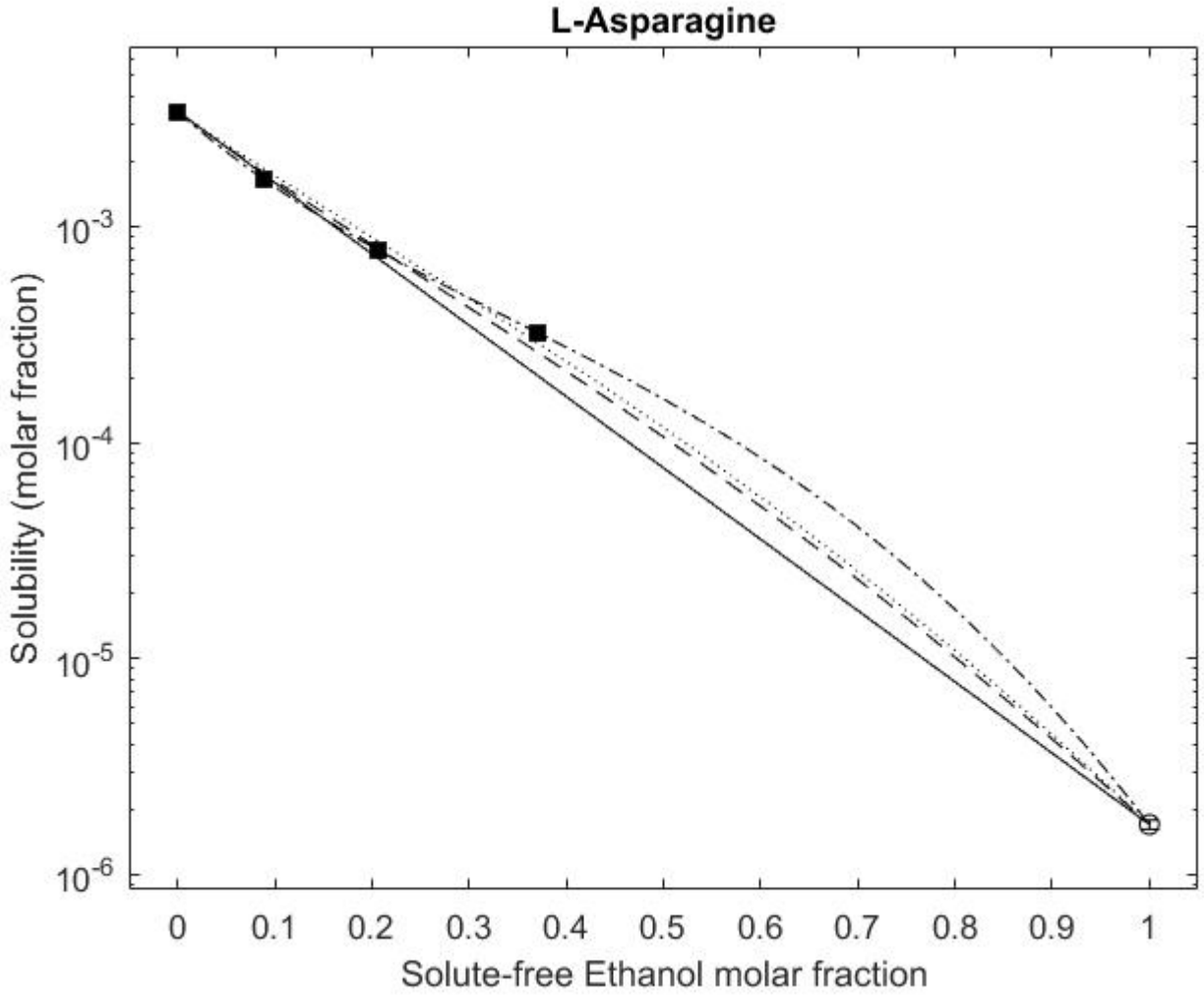
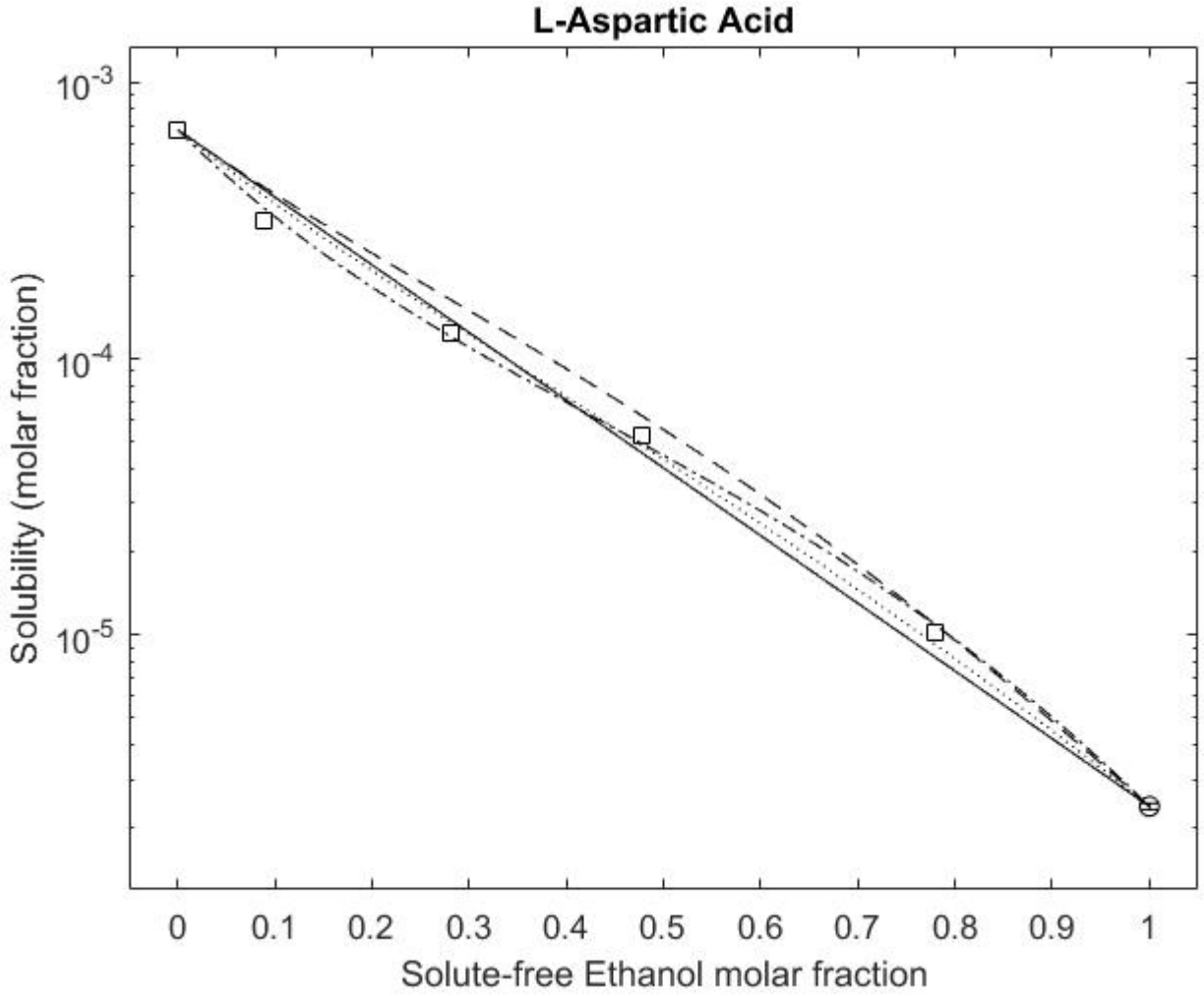


Figure 3: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).



230

231 **Figure 4: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
232 **Data are from the authors (open circles) or from the literature (closed squares).**

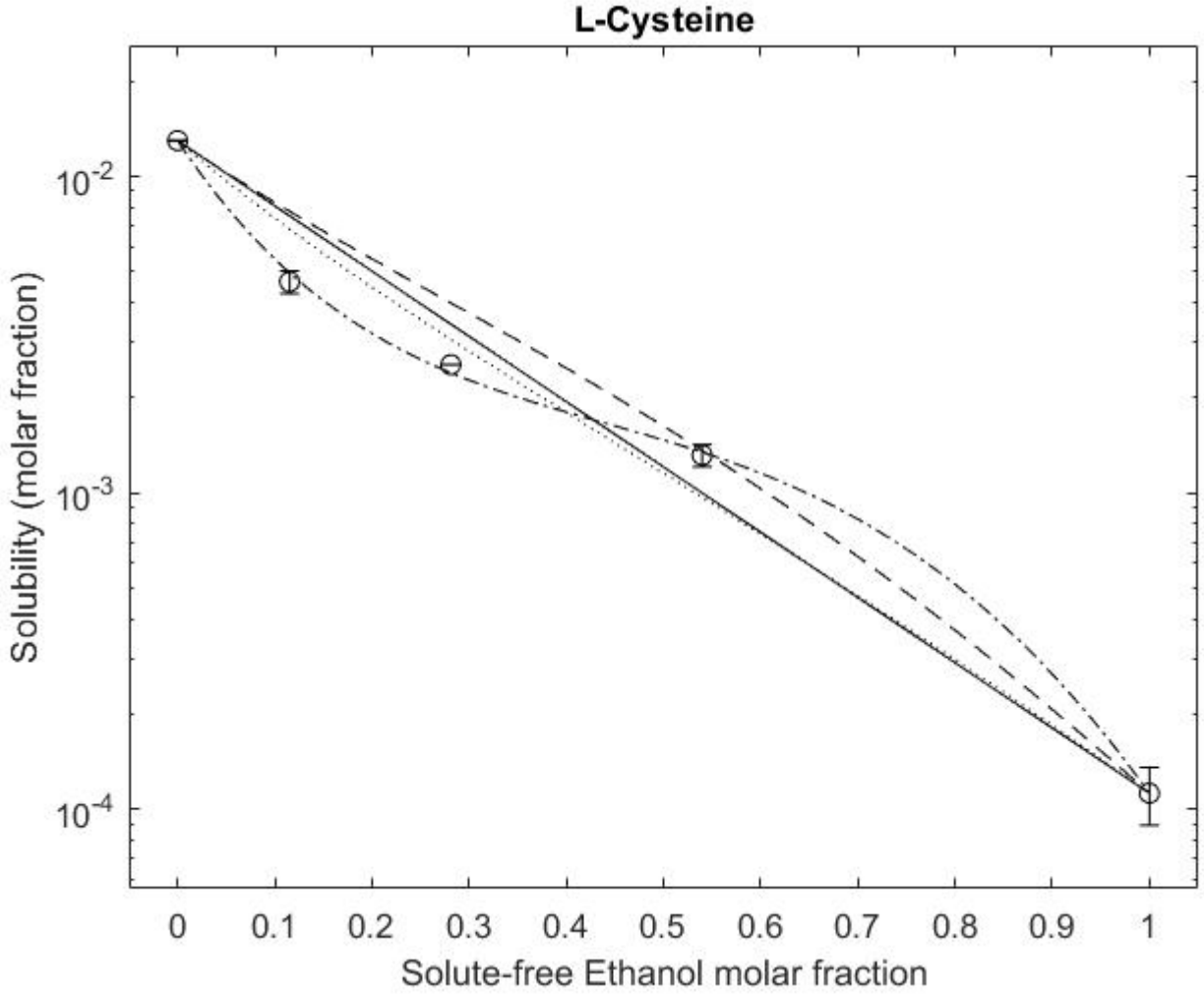


Figure 5: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).

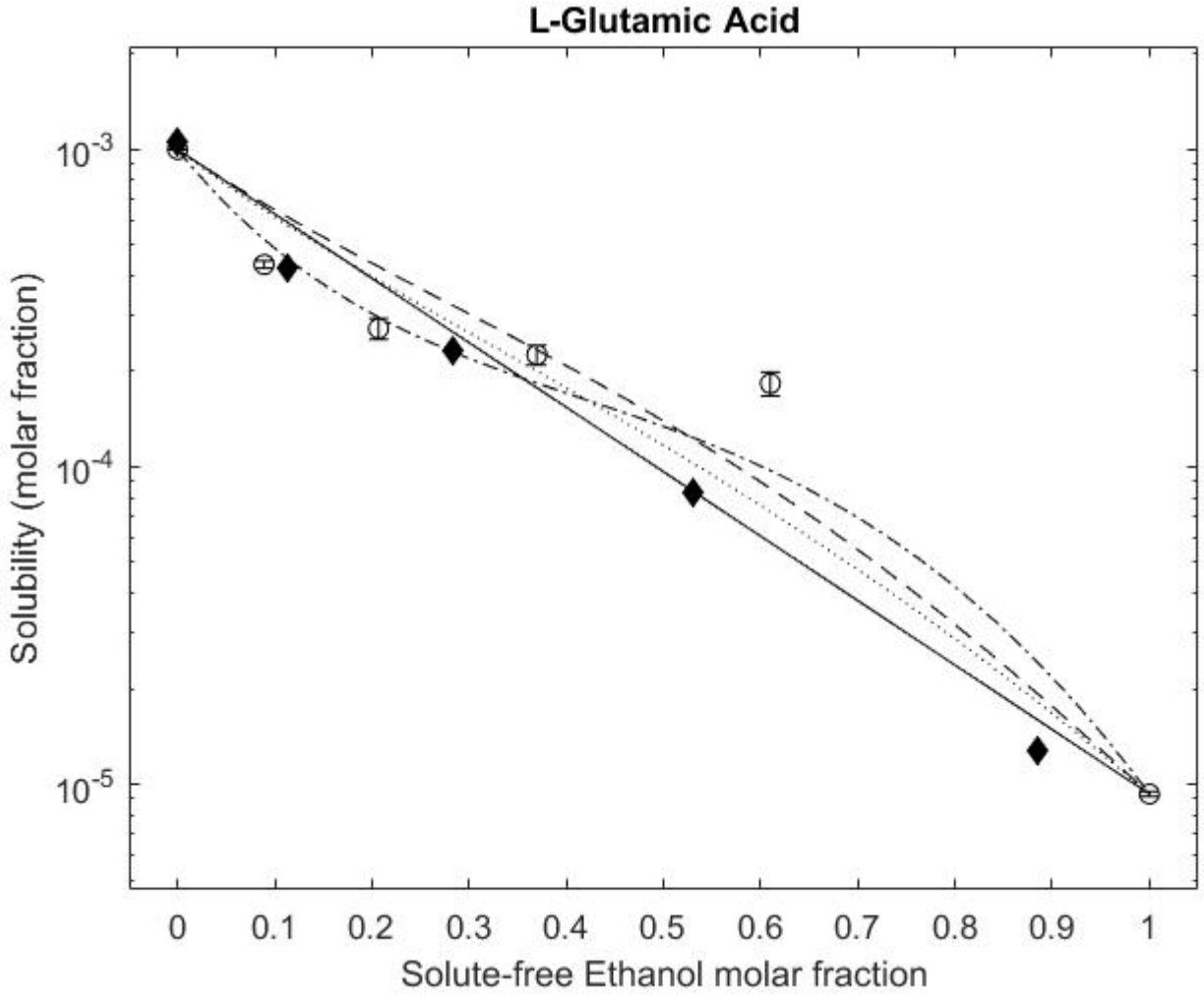
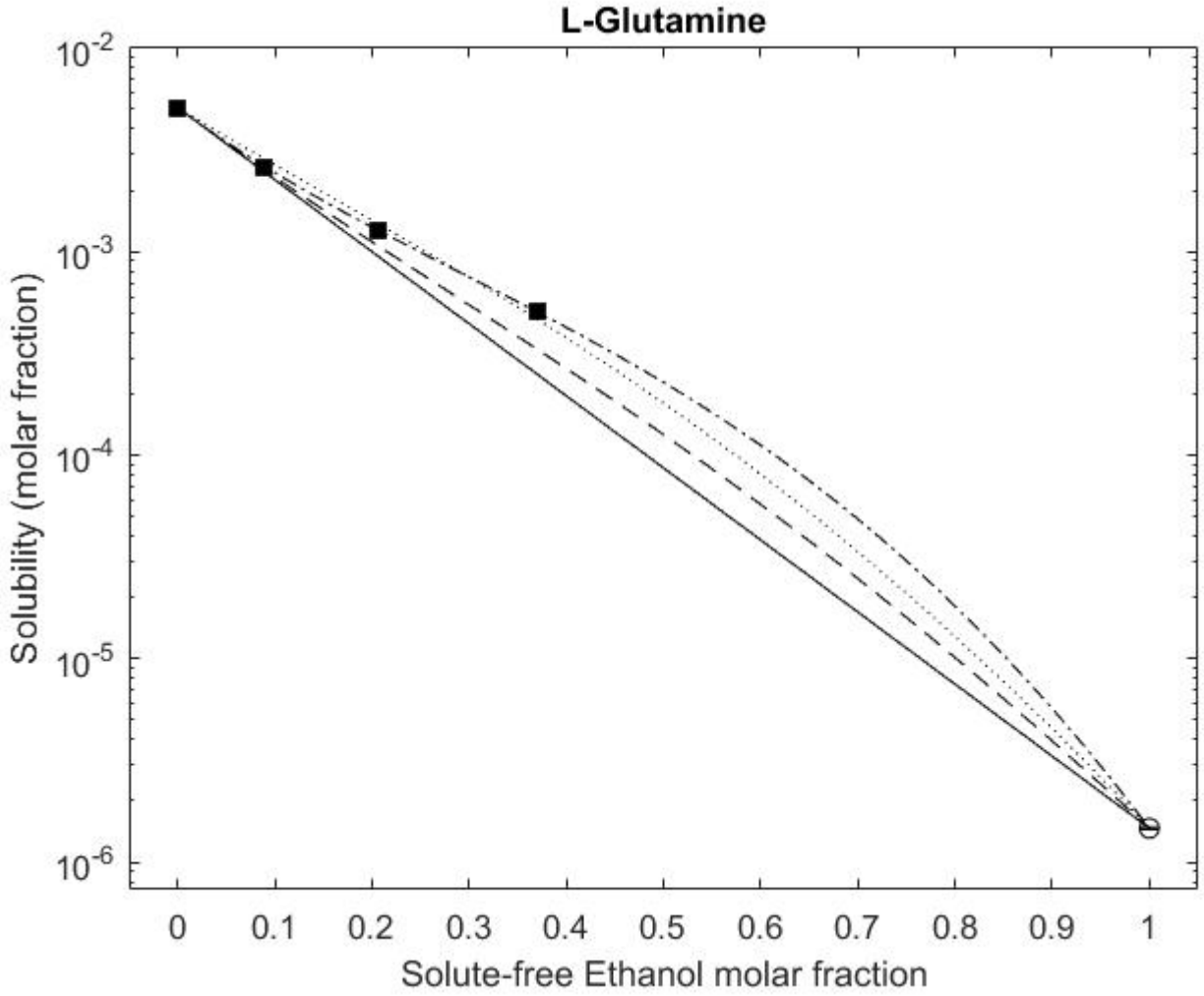
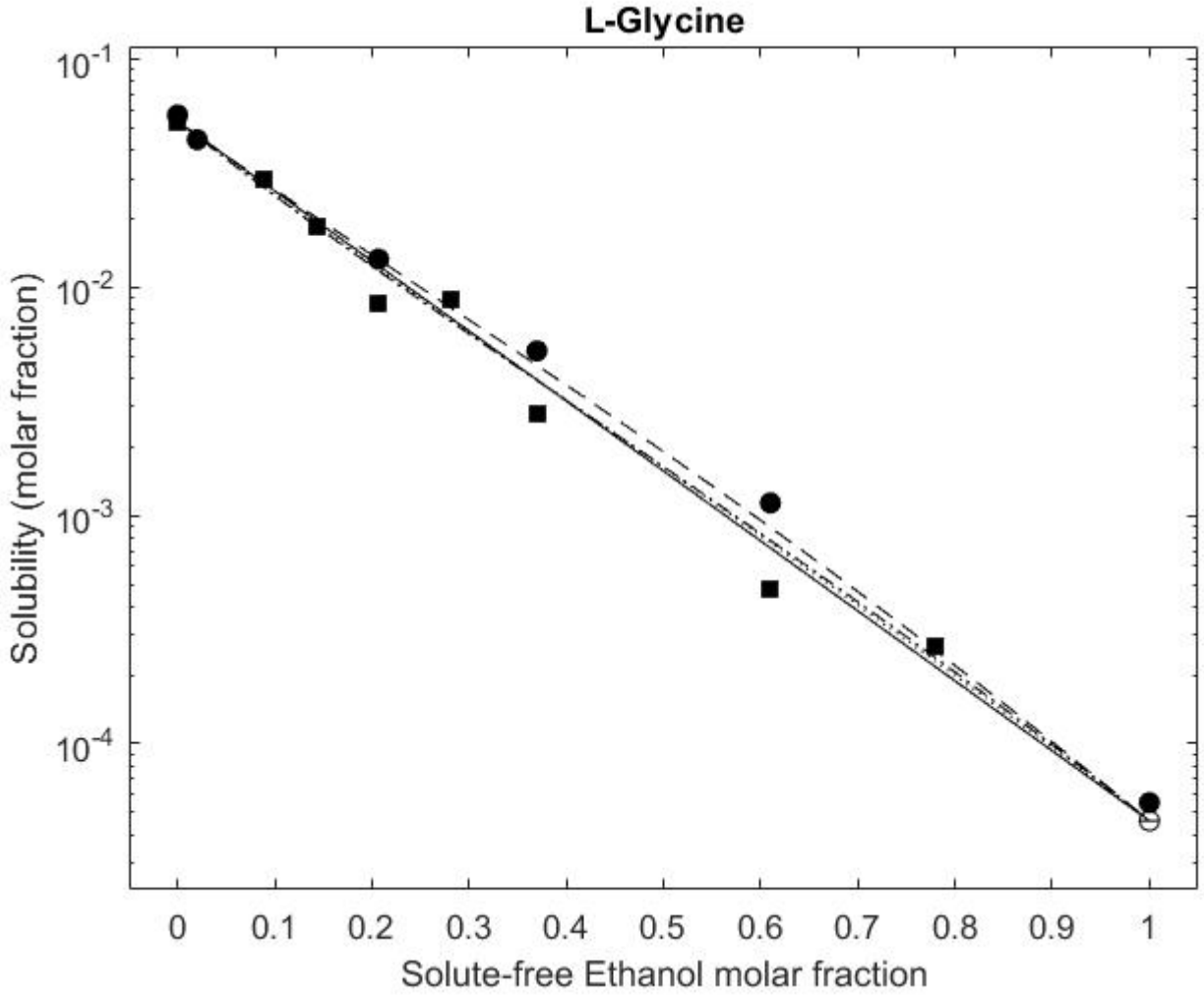


Figure 6: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).



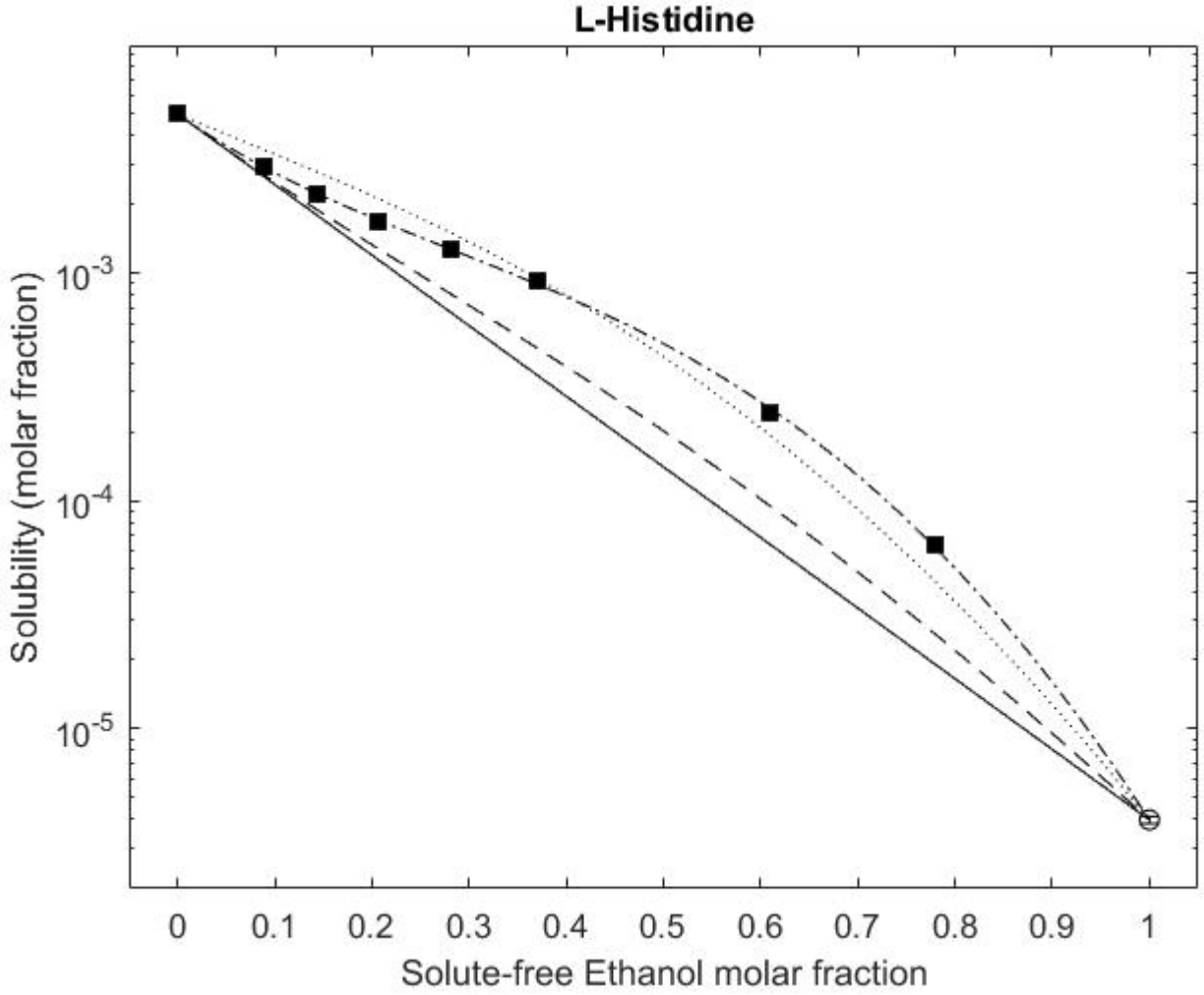
239

240 **Figure 7: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
241 **Data are from the authors (open circles) or from the literature (closed squares).**



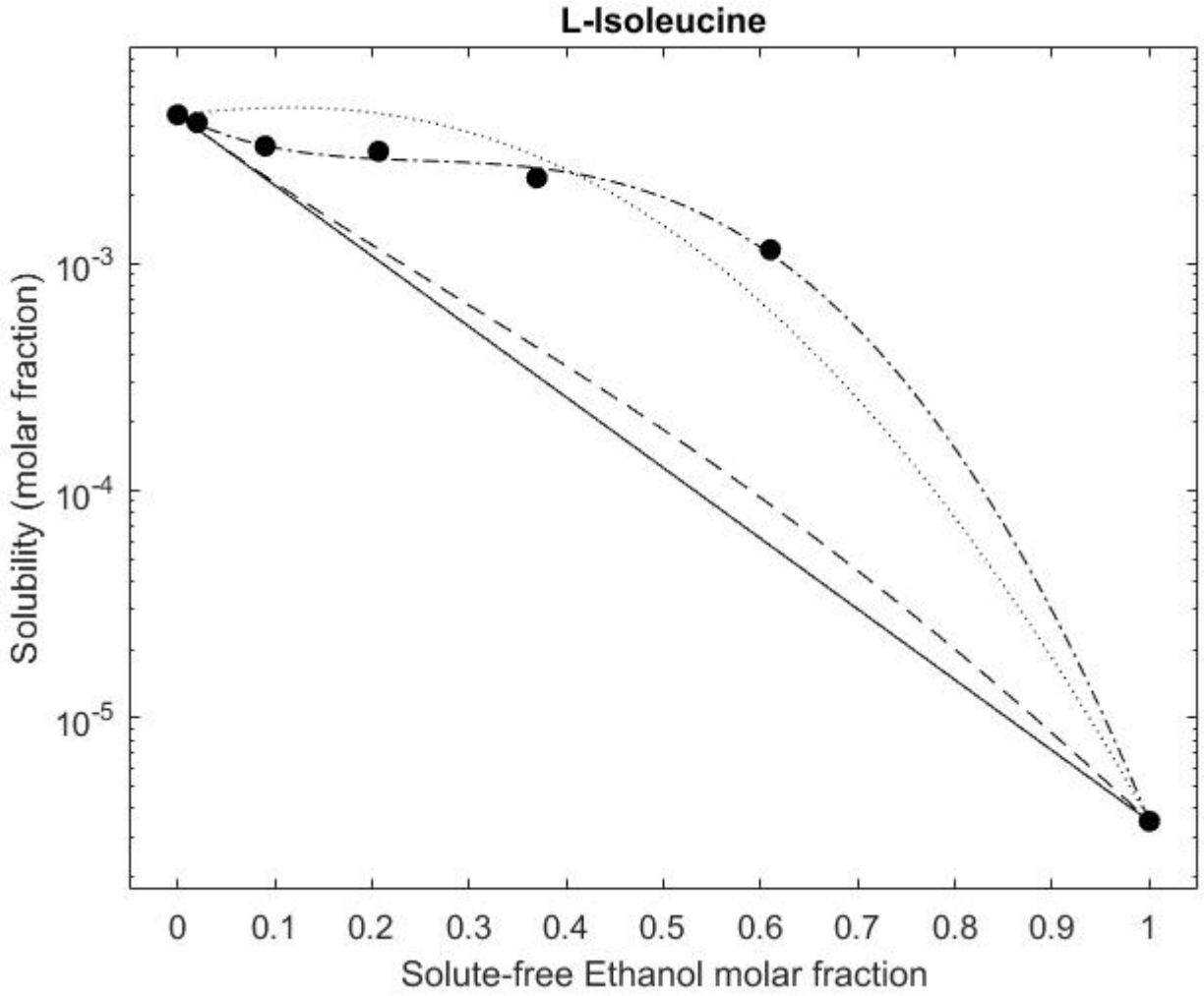
242

243 **Figure 8: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
244 **Data are from the authors (open circles) or from the literature (closed squares).**



245

246 **Figure 9: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
247 **Data are from the authors (open circles) or from the literature (closed squares).**



248

249 **Figure 10: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
250 **Data are from the authors (open circles) or from the literature (closed squares).**

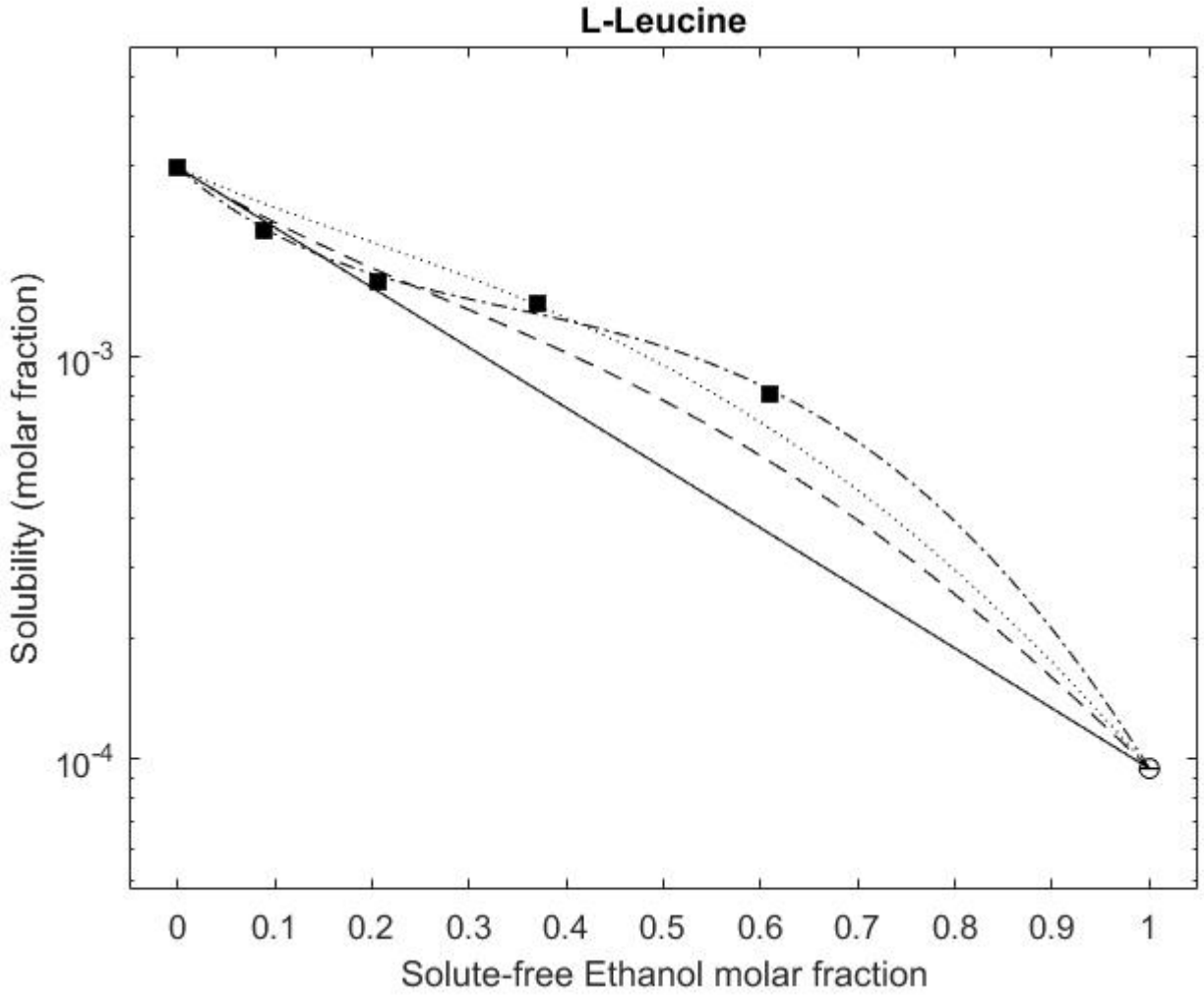


Figure 11: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).

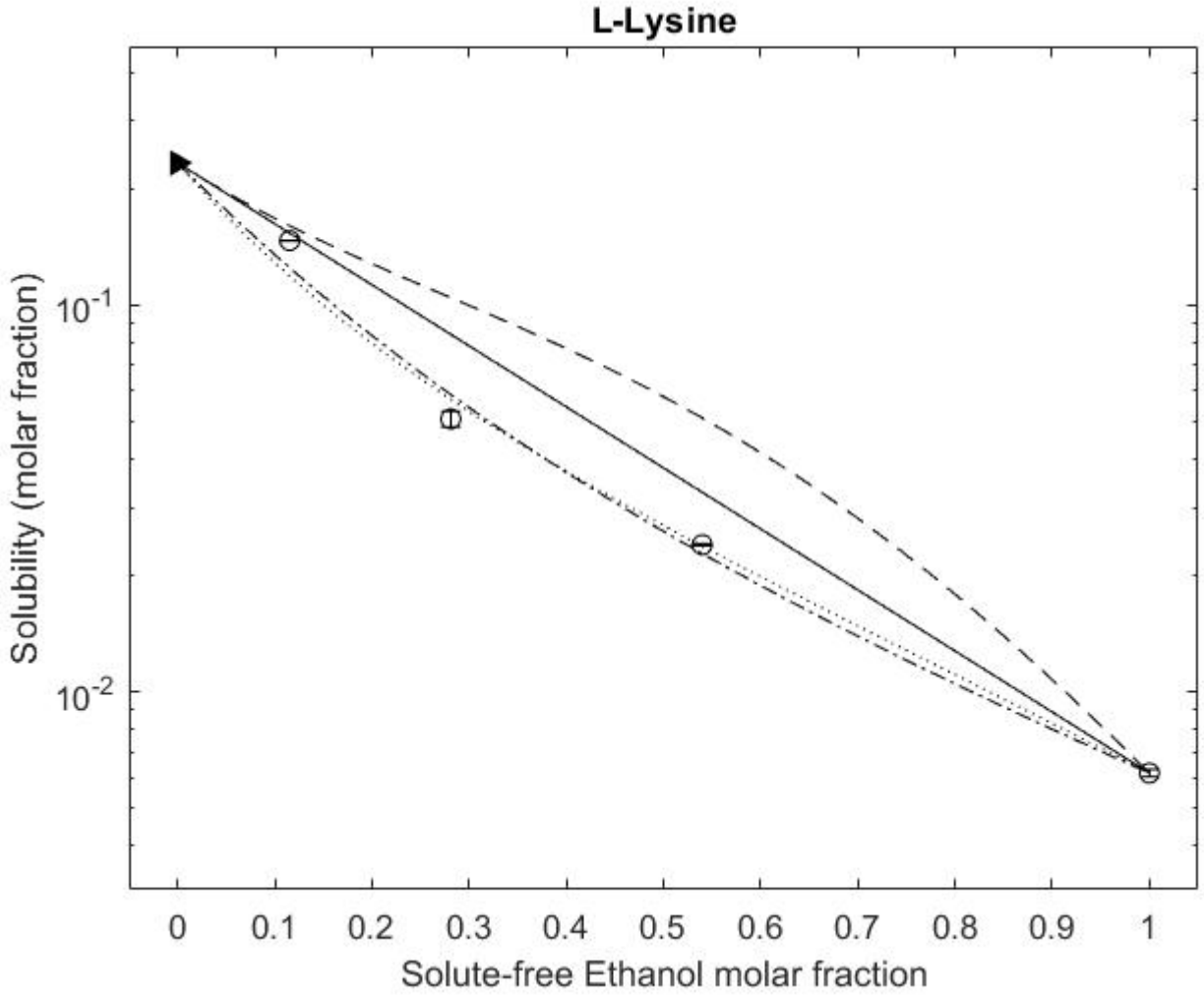


Figure 12: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).

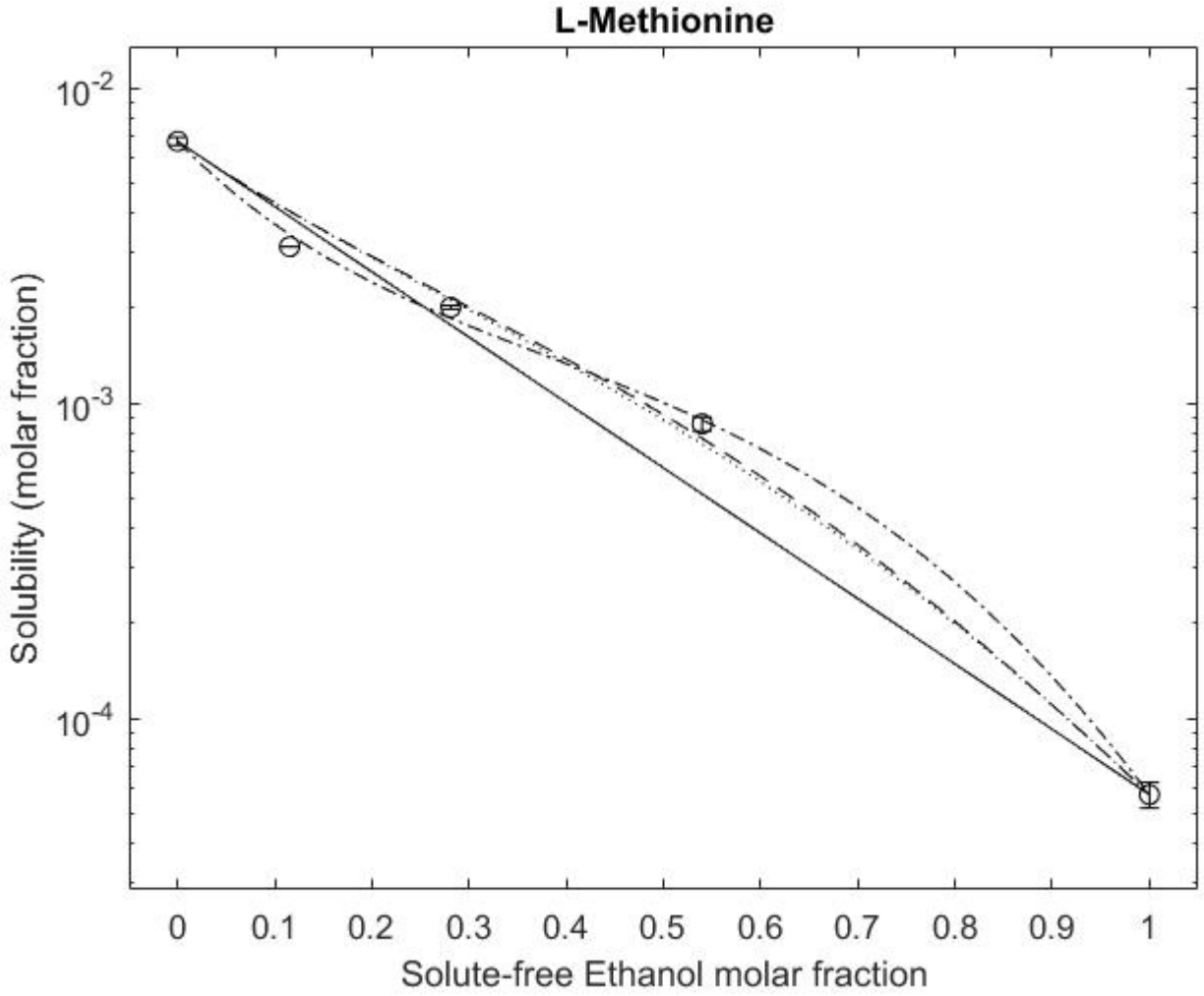


Figure 13: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).

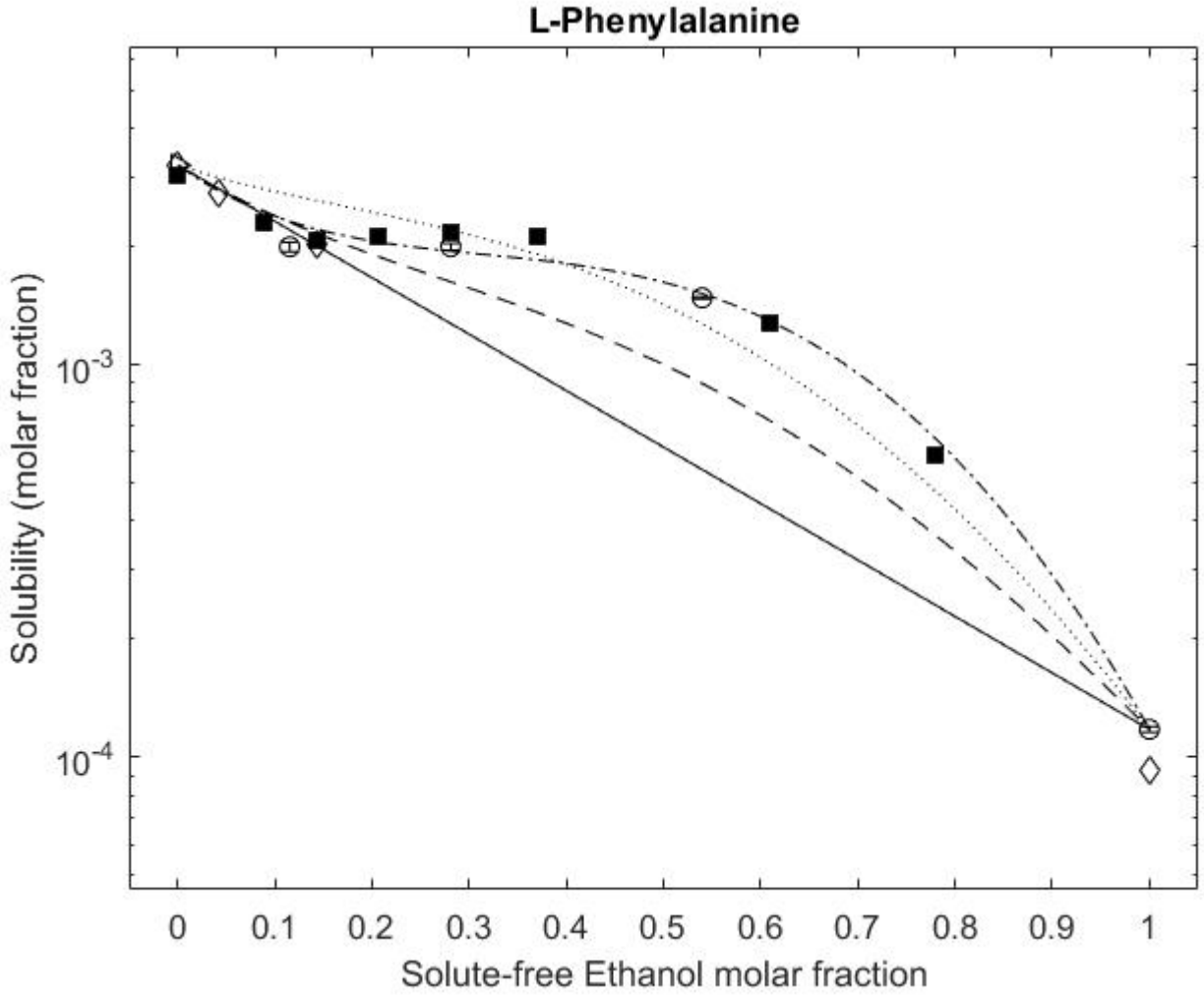
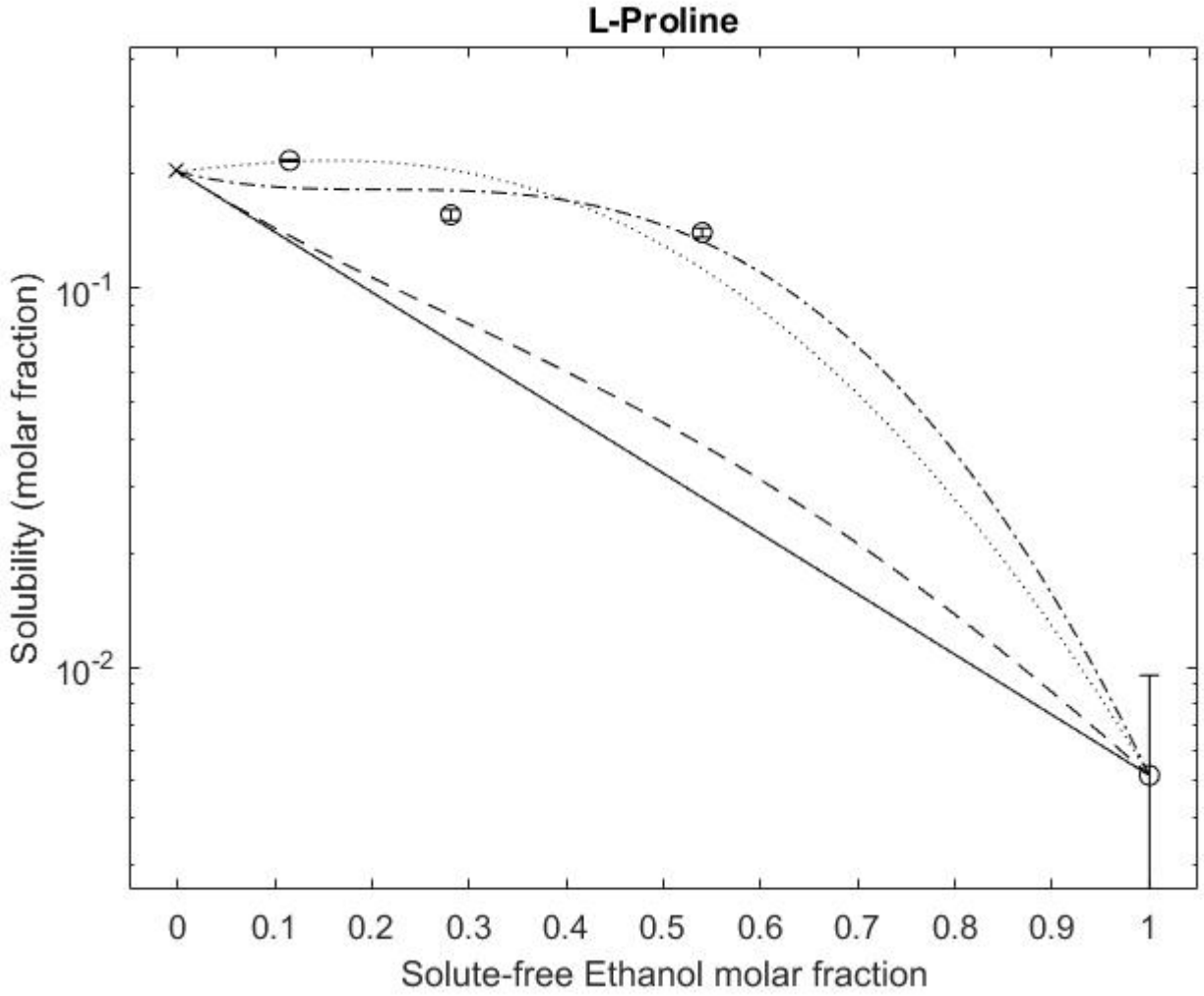


Figure 14: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).



263

264 **Figure 15: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
265 **Data are from the authors (open circles) or from the literature (closed squares).**

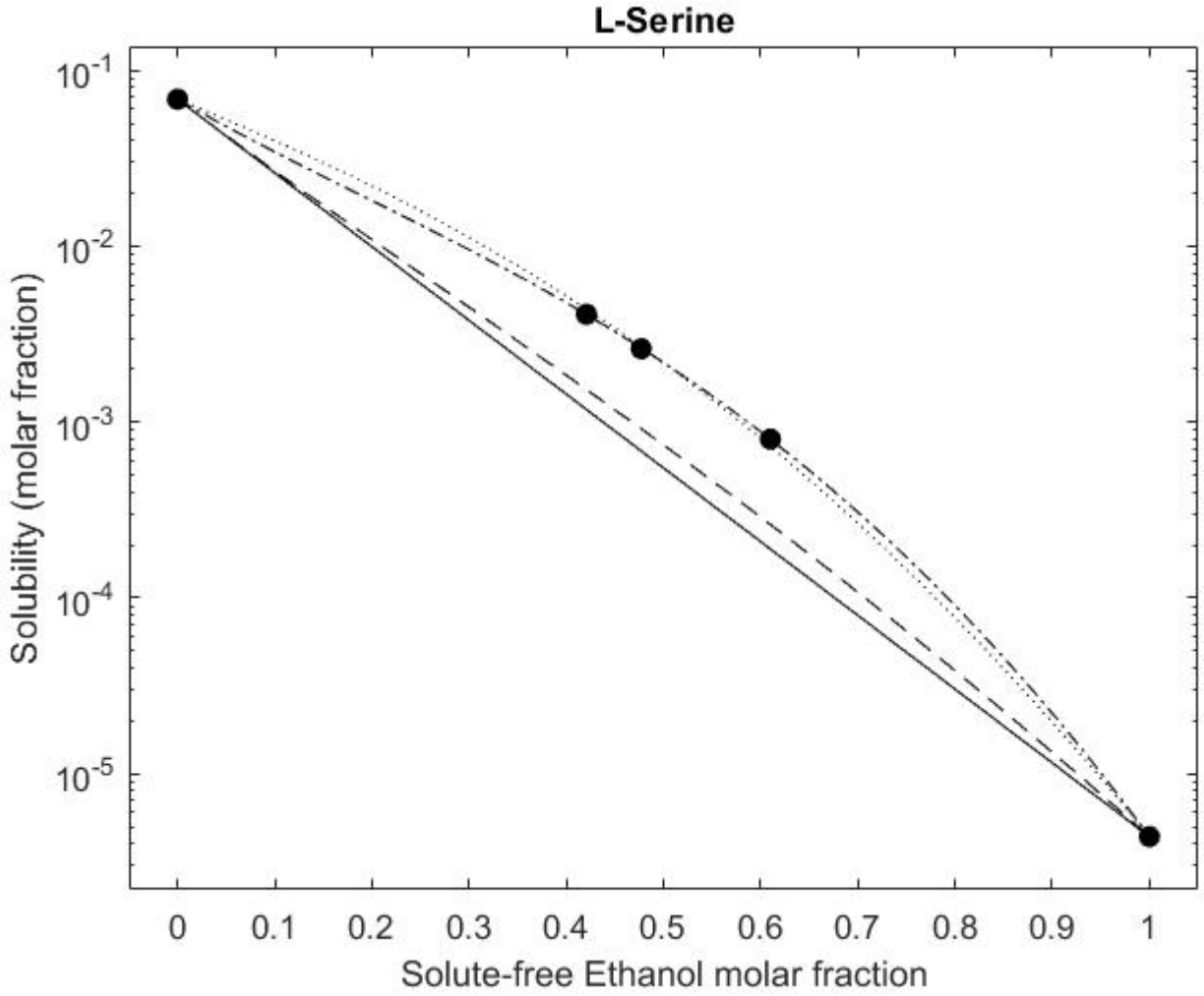


Figure 16: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).

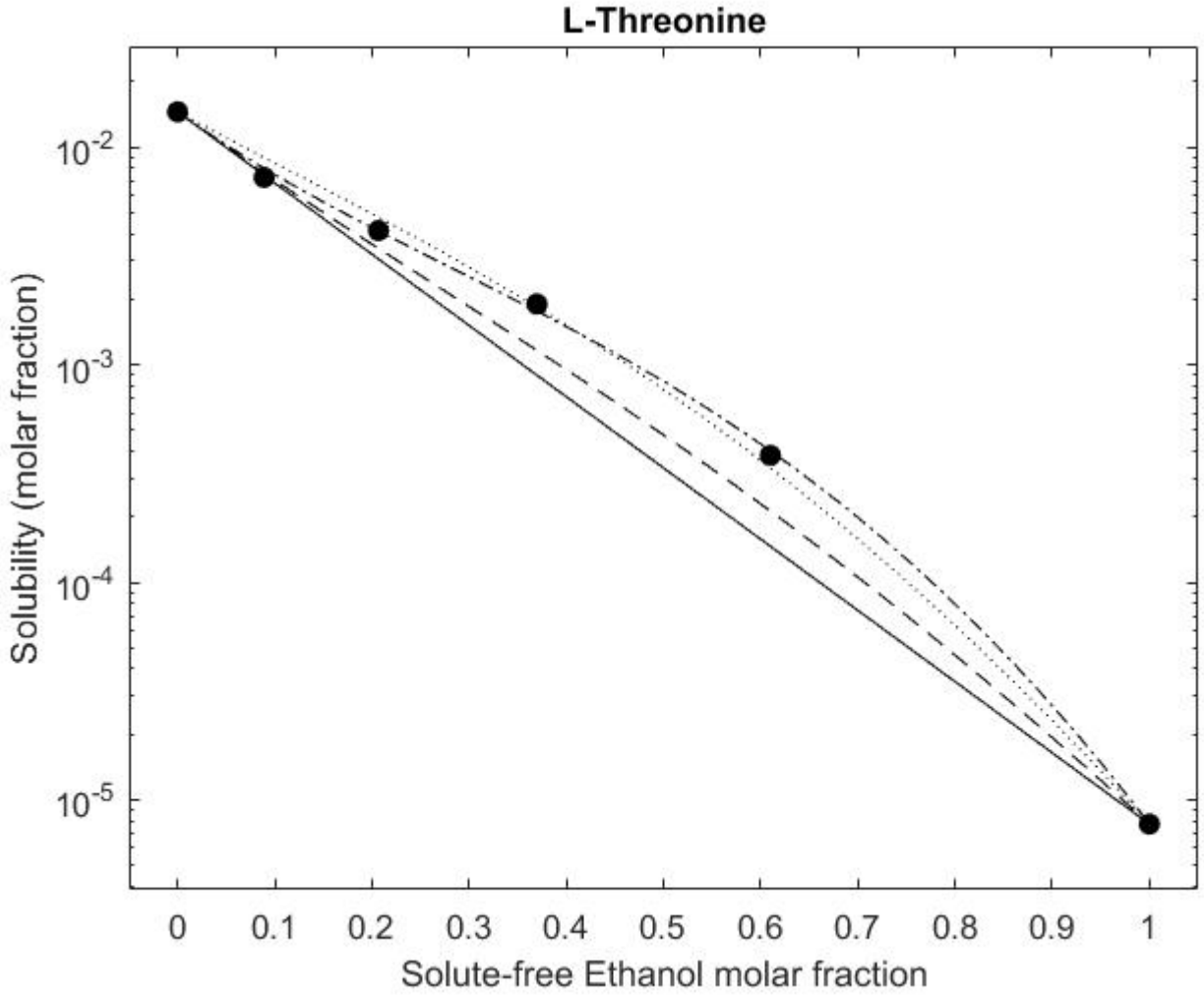
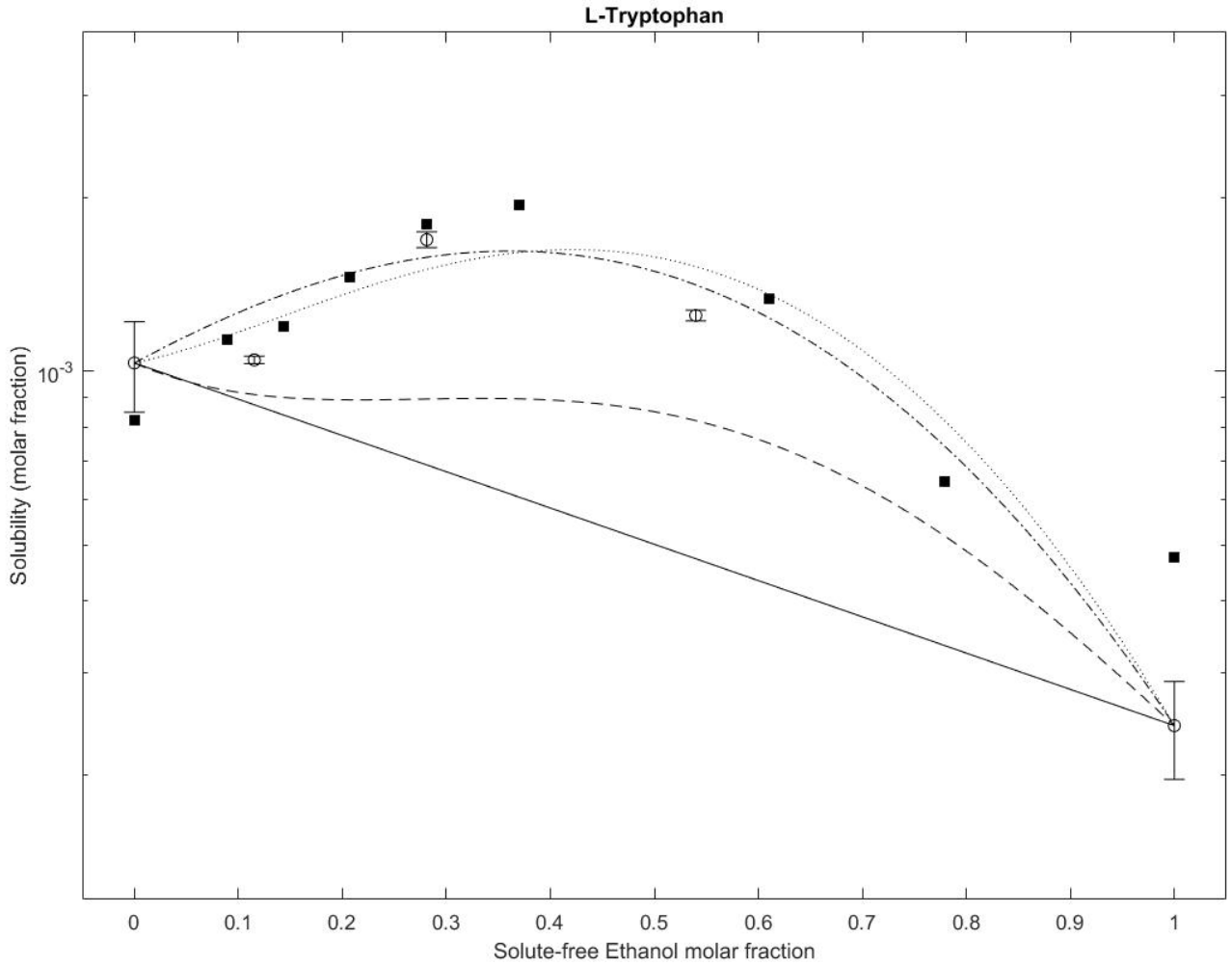


Figure 17: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).



272

273 **Figure 18: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
274 **Data are from the authors (open circles) or from the literature (closed squares).**

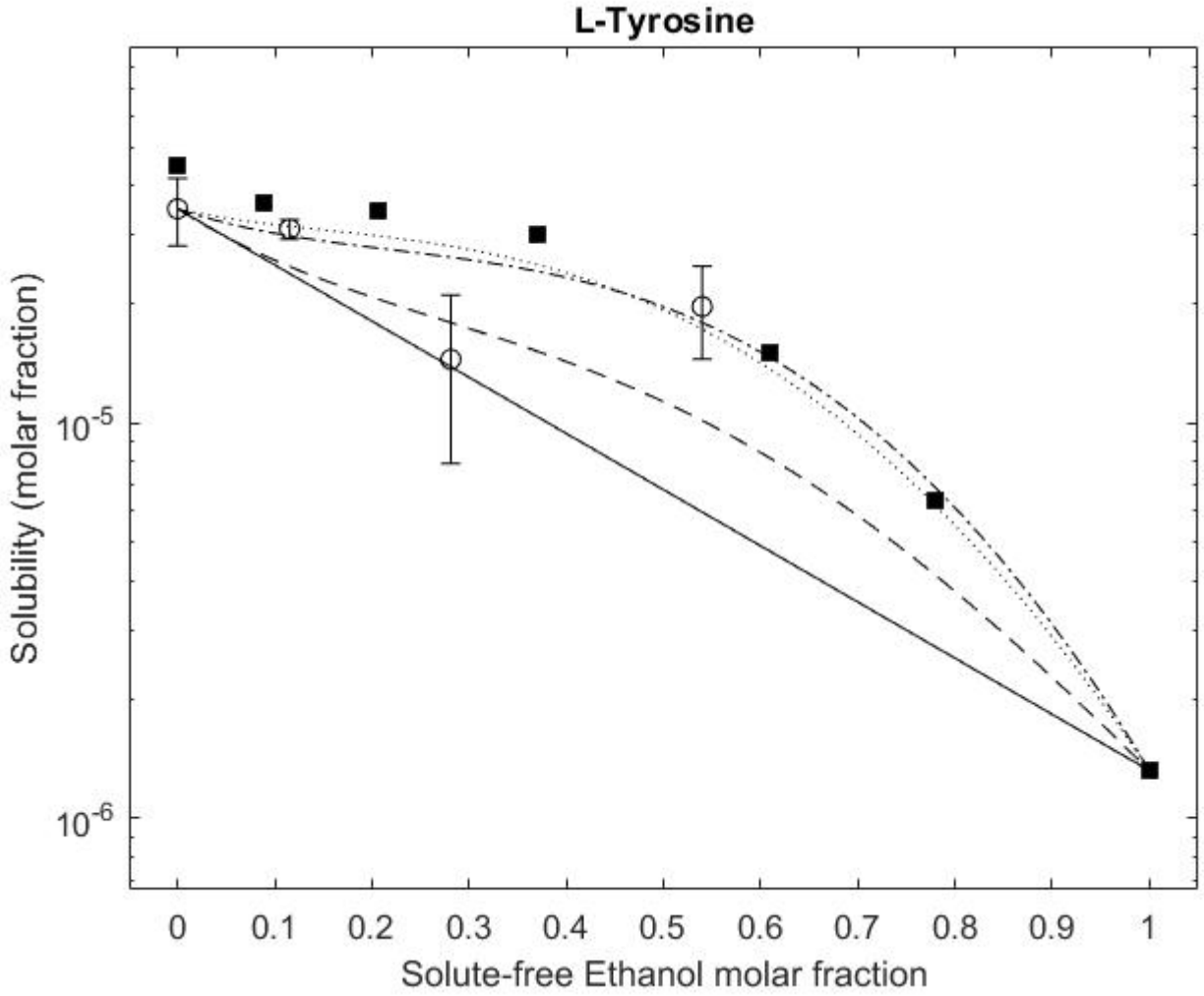
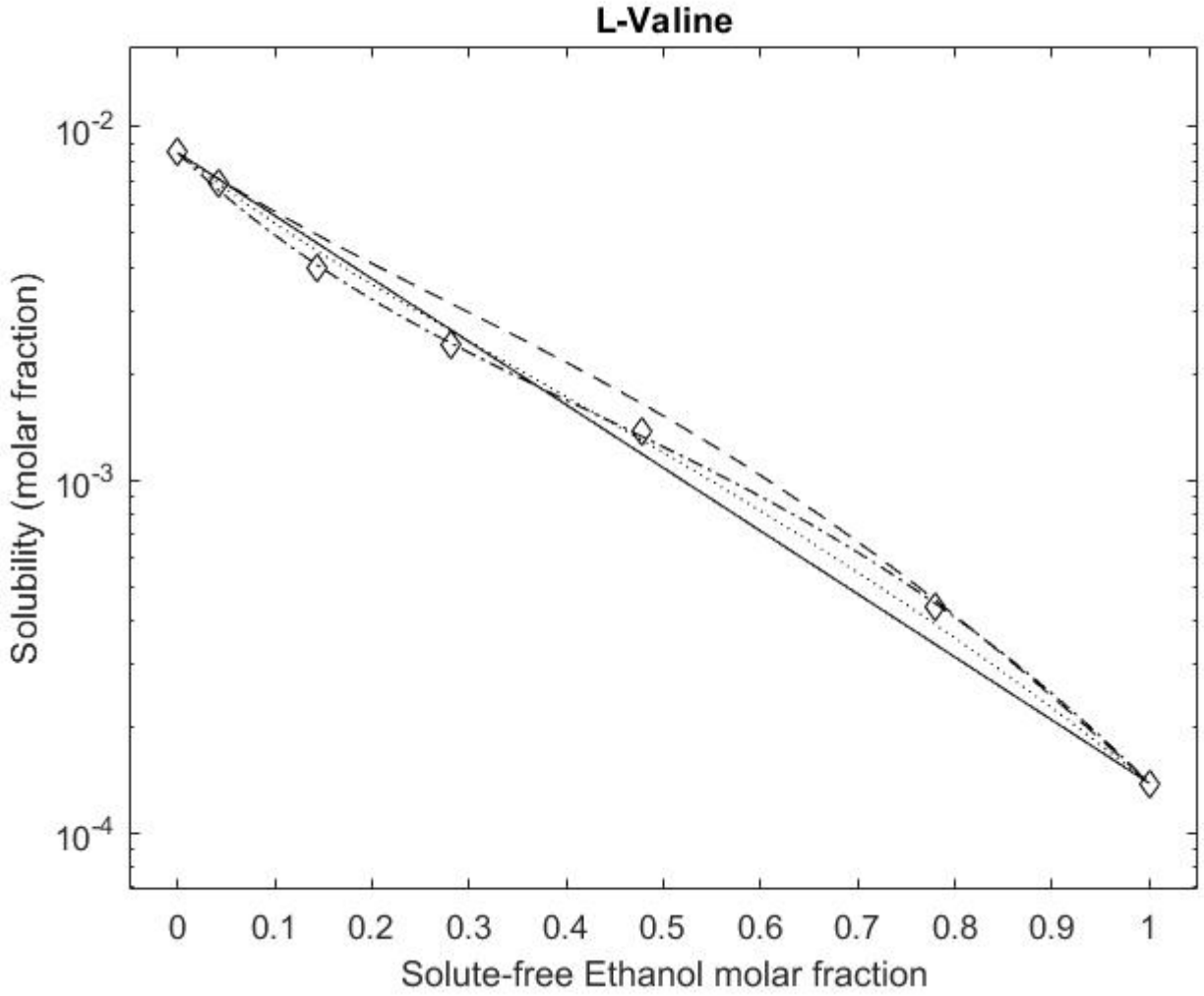


Figure 19: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids. Data are from the authors (open circles) or from the literature (closed squares).



278

279 **Figure 20: Plots of excess solubility = 0 (solid line), Gude model (dotted line), NRTL model (dot-dash line) and Jouyban-Acree model (dashed line) of the proteinogenic amino acids.**
280 **Data are from the authors (open circles) or from the literature (closed squares).**

Table 2: Calculated parameters for the Jouyban-Acree model and regressed parameters for the Gude and NRTL models for each amino acid

Model	Jouyban-Acree			Gude	NRTL	
Parameter	δ_d MPa$^{0.5}$	δ_p MPa$^{0.5}$	δ_{hb} MPa$^{0.5}$	$C_{j,i,aa}$ mol·L^{-1}	τ (water, aa) * 10^6	τ (ethanol, aa) * 10^6
L-ARGININE	18.2312	8.0426	18.7229	1.5926	1.7003	4.2508
L-CYSTEINE	18.2152	6.2829	16.6663	-0.0542	0.9855	2.4638
Glycine	16.3684	10.0170	14.8238	-0.3007	1.2510	3.1276
L-ALANINE	16.0719	5.1966	12.4649	-0.9696	1.6393	4.0982
L-ASPARAGINE	16.8666	13.1746	17.4297	1.3097	1.0379	2.5947
L-ASPARTIC ACID	16.7254	7.2224	17.7194	0.3348	1.0962	2.7404
L-GLUTAMIC ACID	16.6985	6.9179	17.3075	0.8557	1.0147	2.5369
L-GLUTAMINE	16.8397	12.8701	17.0178	2.3001	1.0566	2.6416
L-HISTIDINE	19.2245	4.8443	14.8368	3.2647	1.0297	2.5743
L-ISOLEUCINE	15.7186	3.8964	11.0699	6.7822	0.9472	2.3681
L-LEUCINE	15.7646	3.8983	11.3848	1.9626	1.0476	2.6190
L-SERINE	16.7016	8.5020	19.1997	3.6126	1.0840	2.7100
L-THREONINE	16.4021	7.8108	18.6285	2.4094	1.0718	2.6796
L-VALINE	15.7915	4.2028	11.7967	0.4935	1.1135	2.7837
L-LYSINE	16.3246	7.5725	18.0542	-0.2720	1.2858	3.2146
L-METHIONINE	17.0776	5.3406	11.4124	1.3551	1.0421	2.6053
L-PHENYLALANINE	17.7072	4.5880	10.6483	3.0520	1.0343	2.5857
L-PROLINE	19.1658	6.1022	13.9127	3.6895	1.0573	2.6430
L-TRYPTOPHAN	20.3128	5.1780	8.4406	4.1462	1.2889	3.2223
L-TYROSINE	17.2033	3.2604	18.1645	3.8473	1.0968	2.7420
Water	15.6	16	42.3	N/A	N/A	N/A
Ethanol	15.8	8.8	19.4	N/A	N/A	N/A

Comparing regressed to predictive models of excess solubility

The NRMSE values and the number of measurements, n, for all of the models for each amino acid are shown in Table 3. The model with the lowest NRMSE value is the most accurate. For some amino acids, the number of data points were low, with only 5 or 6 data points. Some of these amino acids with only 5 or 6 data points show the highest NRMSE values and therefore the most error. However, other amino acids with 5 data points (e.g. L-serine, L-methionine) had low error values. It is possible to compare the accuracy of the models for each amino acid since all models used the same data points. However, since the number of data points for some amino acids is limited, we cannot draw conclusions on the amino acids by comparing the NRMSE values.

For all amino acids, the NRTL model had the lowest error and is therefore the most accurate. The second most accurate for all amino acids, except for L-methionine, was the Gude model. The predictive Jouyban-Acree model was more accurate than the Gude model for L -methionine. Both the NRTL and Gude models had lower error values for all (in the case of NRTL) or most (in the case of Gude) amino acids. The predictive Jouyban-Acree model had a higher error value for all amino acids when compared to the NRTL model. The Jouyban-Acree model had a higher error value for all amino acids except L-methionine when compared to the Gude model.

The NRTL model described the empirical data well for all of the amino acids. All error values for the NRTL model were below 0.500, except for L-arginine, which had only 5 data points.

Table 3: NRMSE values for each amino acid for the Gude, NRTL and Jouyban-Acree models

Amino Acid	**n**	Gude	NRTL	Jouyban-Acree
l-Arginine	5	0.816	0.531	1.060
l-Cysteine	5	0.401	0.070	0.522
Glycine	15	0.286	0.285	0.310
l-Alanine	6	0.423	0.379	1.270
l-Asparagine	5	0.210	0.009	0.255
l-Aspartic Acid	6	0.284	0.161	0.476
l-Glutamic Acid	11	0.257	0.217	0.264
l-Glutamine	5	0.125	0.003	0.413
l-Histidine	9	0.182	0.016	0.483
l-Isoleucine	7	0.131	0.020	0.499
l-Leucine	6	0.191	0.042	0.260
l-Serine	5	0.360	0.021	5.470
l-Threonine	6	0.147	0.067	0.402
l-Valine	7	0.217	0.069	0.436
l-Lysine	5	0.304	0.280	1.320
l-Methionine	5	0.237	0.098	0.227
l-Phenylalanine	17	0.134	0.073	0.214
l-Proline	5	0.181	0.118	0.773
l-Tryptophan	14	0.174	0.170	0.354
l-Tyrosine	11	0.222	0.215	0.407

While the Gude model fits had higher NRMSE values than the NRTL model, the values of the error of the Gude model were under 0.500 for 19 of the 20 proteinogenic amino

acids. The exception is L-arginine (NRMSE = 0.816). Since the errors are low, the Gude model could be used for drawing conclusions as we do in the next section. However, when more accurate calculations are needed, e.g. when designing an industrial process, we advise using the NRTL model.

Of the 20 amino acids, 14 of the amino acids modelled by the Jouyban-Acree model were under 0.500 except for L-arginine, L-cysteine, L-alanine, L-serine, L-lysine and L-proline. These 6 amino acids had only 5 or 6 data points each and were some of the most soluble amino acids. Furthermore, 5 of these 6 amino acids with NRMSE values above 0.500 in the Jouyban-Acree model had low NRMSE values using one or both of the other models. Even without using regressed paramaters, the Jouyban-Acree model predicts the amino acid solubility for most of the amino acids well, but not as well as the Gude and NRTL models. The Jouyban-Acree model could be used when there are no or few solubility data available.

Effect of molecular shape on excess solubility of amino acids

As discussed earlier, the work of Flory-Huggins shows that liquids, similar to solids, have an entropic and lattice structure. Due to this entropy, Prausnitz *et al*[56] showed that the shape of a solute has an effect on the solubility of the solute. In their work, they used the relative van der Waals variables Q, surface area, and r, radius of the molecule, to describe the shape of the molecule and therefore how it influences this entopic and lattice structure. The shape of spherical solutes (Q/r = 1.00) showed no effects on the excess solubility of a solute. Straight-chain solutes (Q/r = 0.788) showed strong effects on the excess solubility of the solute, while rod-like solutes (Q/r = 0.394) showed an even greater effect on the excess solubility of the solute.

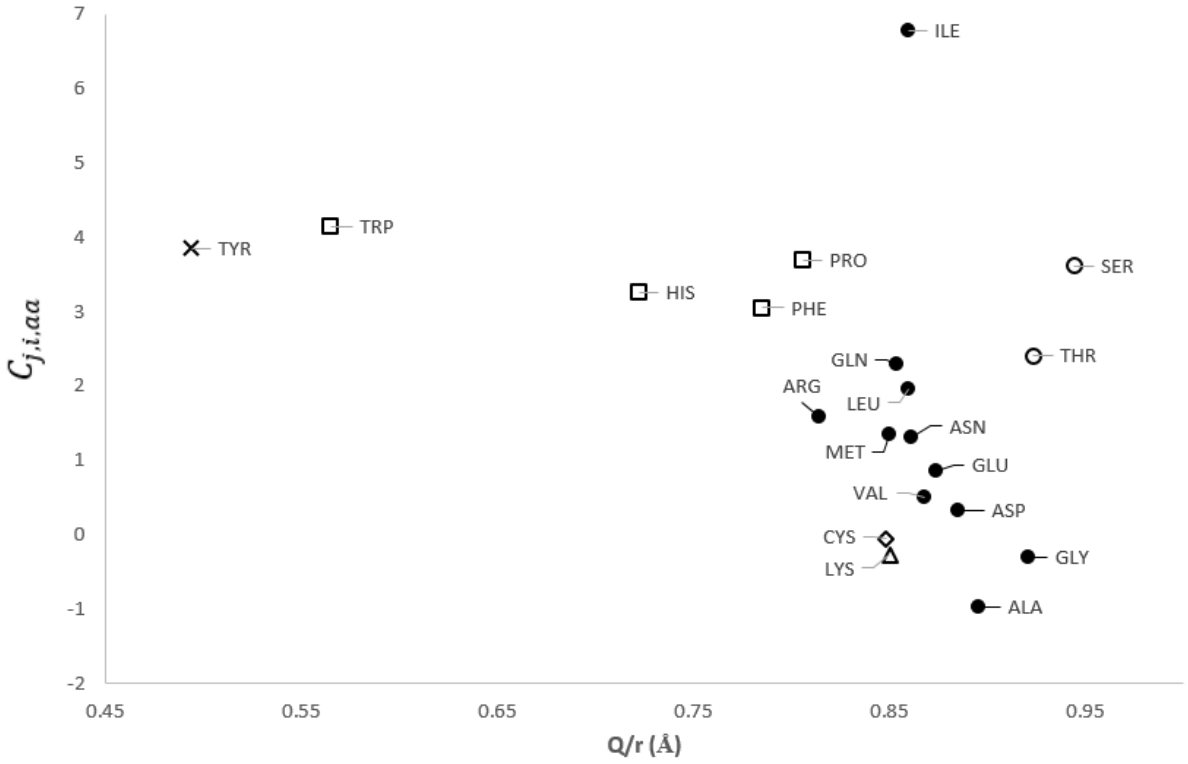


Figure 21: Regressed Gude model solubility parameter, $C_{j,i,aa}$, in relation to UNIFAC surface and radius parameters, Q/r showing non-reactive polar and aliphatic side chains (solid circles), hydroxyl side chains (open circles), lysyl side chain (open triangle), ringed side chains (open square), sulphur (open diamond) and hydroxyl ringed side chains (cross)

In Figure 2, the UNIFAC variables Q/r for each α-amino acid are plotted against the regressed constant in the Gude model, $C_{j,i,aa}$. A Q/r ratio close to unity means that the molecule is spherical and a lower ratio means that the molecule is rod-like. The $C_{j,i,aa}$ denotes the degree of excess solubility. A $C_{j,i,aa}$ close to 0 means that there is no excess solubility. A positive $C_{j,i,aa}$ means there is positive excess solubility and negative means there is negative excess solubility.

Spherical α-amino acids, like glycine, L-alanine and L-aspartic acid, with Q/R ratios from 0.89 to 0.92, react with less molecules of solvent. The spherical amino acids are surrounded by less water molecules than the rod-like amino acids, as their local concentration of ethanol is close to the concentration of the whole solution. As an organic anti-solvent is

added, the lattice structure of these amino acids in solution is disrupted. This leads to little or no excess solubility.

Some rod-like α-amino acids show slightly positive excess solubility. The α-amino acids L-arginine, L-glycine, L-leucine, L-methionine and L-asparagine have Q/r ratios ranging from 0.81 to 0.85 and positive excess solubilities. The evidence supports the conclusion that they have a lower concentration of ethanol molecules around them locally than in the solution in general because of their shape. This would lead to their higher solubility than expected.

Even more pronounced rod-like amino acids, L-tyrosine, L-tryptophan, L-histidine, L-phenylalanine and L-proline, with Q/r ratios between 0.49 and 0.81, could react with even more molecules of solvent, due to their shape.

However, the shape of the amino acid molecules and therefore their effect on the entropic and lattice structure is only a part of the effect that the side chain of the amino acid has on its excess solubility. In Figure 2 there are exceptions to the general trend of the Q/r ratio of the amino acid and its excess solubility. These exceptions are the amino acids with reactive side chains. Therefore, in the next two sections we will examine the effect of the reactivity of the side chain to the excess solubility.

Amino Acids with non-reactive side chains

Eleven amino acids were identified as having non-reactive side chains. Non-reactive side chains are defined here as side chains that are either aliphatic or as measured at their isoelectric point, such as the data in this article, do not have a charge. These are shown in Figure 3 as black circles.

Glycine shows no excess solubility. Glycine has no side chain and has only an amino group and a carboxyl group. This supports the conclusion that lacking a reactive side chain, glycine follows the solubility predicted by the mole fraction of the solubility of both solvents. All other amino acids can be classified as glycine and a side chain. Glycine is therefore the

null amino acid from which the change in excess solubility, not explained by its shape, due to the side chain can be discussed.

L-Glutamine, L-asparagine and L-arginine show little excess solubility. The first two amino acids have an amide in the side chain, while the last one has a guanidinium group in its side chain. At maximum solubility, the solution is at the isoelectric point, meaning that the side chains would not have a charge. Building on the evidence of glycine, the addition of an amide group or an amine group also has little effect on the excess solubility. Their slight increase in excess solubility could be explained by their shape alone as shown by the Q/r ratio.

L-Aspartic acid and L-glutamic acid are negatively charged amino acids. However, as discussed previously with L-arginine, since by definition, maximum solubility is measured at the isoelectric point, L-aspartic acid and L-glutamic acid would not be charged. This could mean that having no charge and being mostly spherical with a non-reactive side chain has no effect on the excess solubility in a two-solvent system. Similar to the previous amino acids, any small increase in excess solubility could possibly be explained by their slightly rod-like shape.

L-Alanine, L-valine, L-methionine, L-leucine and L-isoleucine are aliphatic amino acids. L-alanine has only one methylene group, L-valine and L-methionine have three and L-leucine and L-isoleucine have four. L-Methionine is slightly longer than L-valine because of a sulphur atom in between the second and third methylene. These amino acids show increasing excess solubility in order of their decreasing Q/r ratios. This means that as they become more rod-like, their excess solubility has been shown to increase. However, this does not explain why L-isoleucine has an even higher increased solubility than L-leucine. Further research should be focused on the effect of the position of the branching on the side-chain to understand its effects on excess solubility.

Amino acids with reactive side chains

Nine amino acids have reactive side chains. These amino acids therefore would not follow the trend of higher Q/r ratios leading to lower excess solubility.

The only amino acid to show a large negative excess solubility is L-lysine. L-Lysine has a lysyl group in its side chain. This negative excess solubility is most pronounced around equal mole fractions of ethanol and water. The lysyl group is less attractive to the solvents as the water and ethanol are to each other, leading to lower solubility than expected.

All five amino acids with rings on their side chain have high positive excess solubilities. These amino acids include all three phenylic amino acids: L-phenylalanine, L-tryptophan and L-tyrosine. L-Histidine, which has imidazole on its side chain, shows positive excess solubility as well as L-proline, which has pyrrolidine as a side chain. It is possible that the two solvents act as affinity molecules, bringing these amino acids further into solution. However, it is also possible that their rod-like shape is causing this effect.

The three amino acids with a hydroxylic side chain show positive excess solubility. These include L-tyrosine, which is also has a phenyl group, L-serine and L-threonine. A side chain with a hydroxyl group leads to a preferential reaction to the solvents ethanol and water than ethanol to water. This cannot be explained by the shape of the amino acids, since both L-serine and L-threonine are spherical. Therefore, it may be concluded that an addition of a hydroxyl group leads to a marked increase in excess solubility.

5 Conclusion

The results support a hypothesis that both the shape of an amino acid and the activity of the side chain of an amino acid influence the solubility of the amino acid in mixed solvent solutions. Results support the conclusion that if the amino acid is spherical and does not have a reactive side chain, then there will be no change in the excess solubility as expected from

the solvent mole fraction of ethanol and water. Spherical amino acids with reactive side chains, like L-serine and L-threonine, will have positive excess solubilities. Rod-like amino acids with either a long side chain or a reactive side chain, such as the presence of a phenyl group and/or hydroxyl group, react preferentially to water and ethanol than water and ethanol do to each other and will have the greatest positive excess solubilities.

This hypothesis is artistically rendered in Figure 3 for four amino acids. In all four amino acids, the mole fraction of ethanol is 0.2. In the top left, L-alanine, a spherical amino acid (Q/r = 0.90; $C_{j,i,aa}$ = -0.97) with a non-reactive side chain, is shown. Here the ethanol disrupts the water molecule lattice and there is a slight decrease in excess solubility. In the top right, L-serine, a spherical amino acid (Q/r = 0.94; $C_{j,i,aa}$ = 3.61) with a reactive hydroxyl group on its side chain, is shown. The ethanol does not disrupt the lattice, rather it joins the lattice, being attracted to the hydroxyl group. Given small to medium molar concentrations of ethanol, there is marked positive excess solubility. In the bottom left, L-arginine, a rod-like amino acid (Q/r = 0.81; $C_{j,i,aa}$ = 1.59) with a non-reactive side chain, is shown. Here, the lattice of water molecules is not disrupted, because it has contact with many water molecules. Given small molar concentrations of ethanol, there is a small amount of excess solubility. In the bottom right, L-tyrosine, a rod-like amino acid (Q/r = 0.49; $C_{j,i,aa}$ = 3.85) with a reactive ring and hydroxyl groups on its side chain, is shown. Here, the ethanol and the water form a tight lattice around the molecule. In this case, even at medium concentrations of ethanol, there will be great excess solubility. At low concentrations of ethanol, the relative solubility has even been shown to increase.

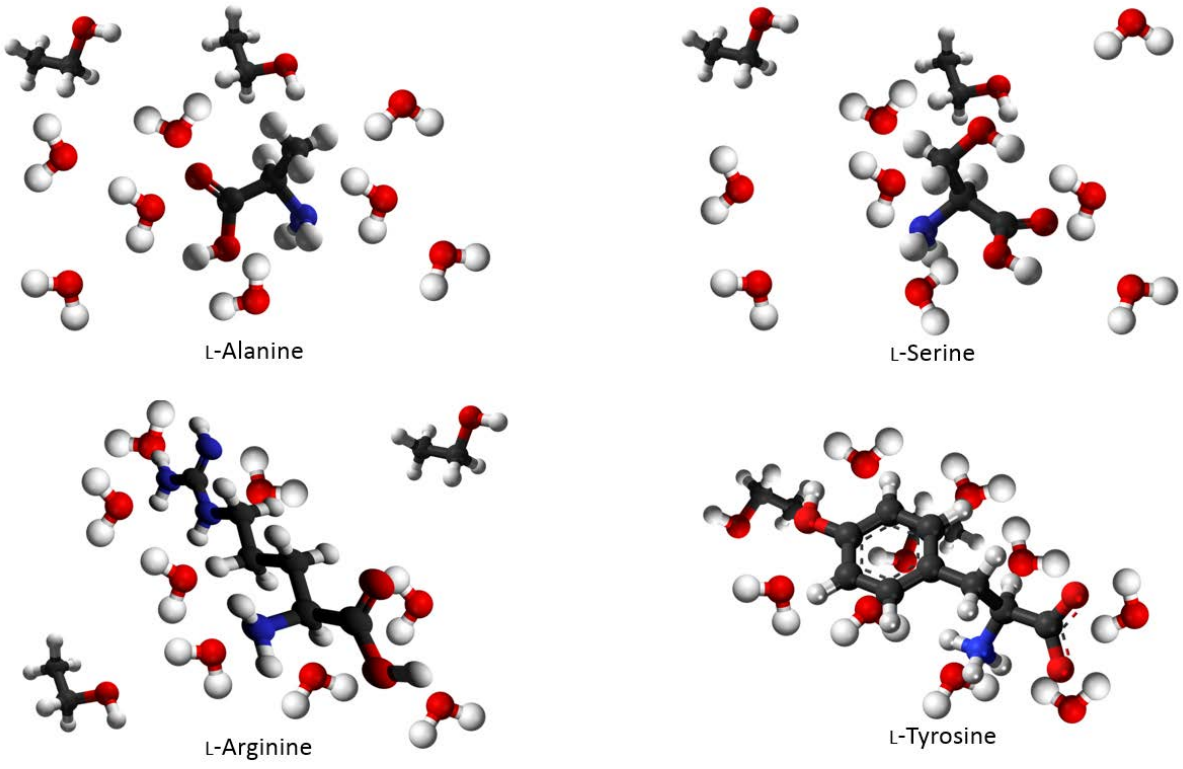


Figure 22: A depiction of the effects of amino acid shape and side chain composition in solution. Top left, L-alanine, spherical and non-reactive. Top right, L-serine, spherical and reactive. Bottom left, L-arginine, rod-like and non-reactive. Bottom right, L-tyrosine, rod-like and reactive.

Regressed models describe the solubility of the amino acids well. The NRTL model is better than the Gude model in this regard. However, since the Gude model has only one regressed parameter, it may be preferential to use it. The predictive Jouyban-Acree model performs well for some amino acids but not as well as both the Gude and NRTL models. Future research on group contribution in amino acid side chains is encouraged, in order to improve the accuracy of predictive models. The model that the end-user should use depends on the accuracy that is required. If the highest accuracy is required and solubility data is abundant, then a regressed model could be used. If the highest accuracy is not required, and there is no or few data, then a predictive model could be used.

The effect of the charge of an amino acid on the solubility of the amino acid has not been studied in this research. All the solubility data were taken at the isoelectric point,

meaning that the amino acid was not charged. Further work on the effect of ethanol on a charged amino acid is encouraged.

6 References

1. Scott, E.; Peter, F.; Sanders, J., Biomass in the Manufacture of Industrial Products – The use of Proteins and Amino Acids. *Applied Microbiology and Biotechnology* **2007,** *75*, 751-762;

2. Lammens, T. M.; Franssen, M. C. R.; Scott, E. L.; Sanders, J. P. M., Availability of protein-derived amino acids as feedstock for the production of bio-based chemicals. *Biomass and Bioenergy* **2012,** *44* (0), 168-181;

3. Lammens, T. M., Bio-based Industrial Chemicals from Glutamic Acid. **2011**. PhD Thesis, Wageningen University Press, Wageningen, the Netherlands

4. Dunn, M.S.; Ross, F.J.; Quantitative Investigations of Amino Acids and Peptides: IV. The Solubilities of the Amino Acids in Water-Ethyl Alcohol Mixtures *Journal of Biological Chemistry* **1938**, 125, 309-322

5 McMeekin, T.L.; Cohn, E.J.; Weare, J.H.; Studies in the Physical Chemistry of Amino Acids, Peptides and Related Substances. III. The Solubility of Derivatives of the Amino Acids in Alcohol—Water Mixtures *Journal of the American Chemical Society* **1935** *57* (4), 626-633

6 Ferreira, L. A.; Pinho, S. P.; Macedo, E. A., Solubility of L-serine, L-threonine and L-isoleucine in aqueous aliphatic alcohol solutions. *Fluid Phase Equilibria* **2008,** *270* (1-2), 1-9;

7 Nozaki, Y; Tanford, C; The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions *Journal of Biological Chemistry* **1971** (246) 7, 2211-2217

8. Gude, M. T.; Meuwissen, H. H. J.; van der Wielen, L. A. M.; Luyben, K. C. A. M., Partition Coefficients and Solubilities of α-Amino Acids in Aqueous 1-Butanol Solutions. *Industrial & Engineering Chemistry Research* **1996,** *35* (12), 4700-4712;

9. Bowden, N.; Sanders, J. P. M.; Bruins, Marieke, Solubility of the proteinogenic α-amino acids in water, ethanol-water mixtures and ethanol. *Journal of Chemical Engineering Data* **2017 (Submitted)**.

10. Bowden, N.; Sanders, J. P. M.; Bruins, Marieke, Solid-liquid phase equilibria and thermodynamic properties of mixtures of 20 α-amino acids in a model system of industrial residues. *Journal of Chemical Thermodynamics* **2017 (Submitted)**.

11. van Berlo, M.; Gude, M. T.; van der Wielen, L. A. M.; Luyben, K., Partition coefficients and solubilities of glycine in the ternary solvent system 1-butanol plus ethanol plus water. *Industrial & Engineering Chemistry Research* **1997,** *36* (6), 2474-2482;

12. Gude, M. T.; van der Wielen, L. A. M.; Luyben, K. C. A. M., Phase behavior of α-amino acids in multicomponent aqueous alkanol solutions. *Fluid Phase Equilibria* **1996,** *116* (1–2), 110-117.

13. Figueiredo, B. R.; Da Silva, F. A.; Silva, C. M., Non-ideality and Solubility Modeling of Amino Acids and Peptides in Aqueous Solutions: New Physical and Chemical Approach. *Industrial & Engineering Chemistry Research* **2013,** *52* (45), 16044-16056;

14. Held, C.; Cameretti, L. F.; Sadowski, G., Measuring and Modeling Activity Coefficients in Aqueous Amino-Acid Solutions. *Industrial & Engineering Chemistry Research* **2011,** *50* (1), 131-141;

15. Hempel, S.; Sadowski, G., Water activity coefficients in aqueous amino acid solutions by molecular dynamics simulation: 1. Force field development. *Mol. Simul.* **2012,** *38* (2), 132-138;

16. Khoshkbarchi, M. K.; Vera, J. H., A Simplified Perturbed Hard-Sphere Model for the Activity Coefficients of Amino Acids and Peptides in Aqueous Solutions. *Industrial & Engineering Chemistry Research* **1996,** *35* (11), 4319-4327;

17. Lee, B.-S.; Kim, K.-C., Study on the activity coefficients and solubilities of amino acids in aqueous solutions with perturbed-chain statistical associating fluid theory. *Korean Journal of Chemical Engineering* **2010,** *27* (1), 267-277;

18. Rudakov, A. M.; Sergievskii, V. V.; Khomchenko, E. O., Modeling component activities in binary aqueous solutions of some amino acids. *Russian Journal of Physical Chemistry* **2004,** *78*, S77-S81;

19. Sadeghi, R., Thermodynamic representation of phase equilibrium behavior of aqueous solutions of amino acids by the modified Wilson model. *Fluid Phase Equilibria* **2007,** *260* (2), 266-274;

20. Soto, A.; Arce, A.; Khoshkbarchi, M. K., Experimental data and modelling of apparent molar volumes, isentropic compressibilities and refractive indices in aqueous solutions of glycine+NaCl. *Biophysical Chemistry* **1998,** *74* (3), 165-173.

21. Gao, W. C.; Li, Z. B., Determination and Chemical Modeling of Phase Equilibria for the Glycine-KCl-NaCl-H2O System and Its Application To Produce Crystals with Anticaking Characteristics. *Industrial & Engineering Chemistry Research* **2012,** *51* (24), 8315-8325;

22. Held, C.; Prinz, A.; Wallmeyer, V.; Sadowski, G., Measuring and modeling alcohol/salt systems. *Chemical Engineering Science* **2012,** *68* (1), 328-339;

23. Khoshkbarchi, M. K.; Vera, J. H., Measurement and modeling of activities of amino acids in aqueous salt systems. *Aiche Journal* **1996,** *42* (8), 2354-2364;

24. Seyfi, S.; Pazuki, G.; Aghamiri, S. F.; Beheshti, M., On the prediction of equilibrium phase behavior of amino acids in aqueous and aqueous-electrolyte solutions using SAFT equation of state. *Fluid Phase Equilibria* **2009,** *287* (1), 15-22;

517 25. Soto, A.; Arce, A.; Khoshkbarchi, M. K., Experimental data and modelling of apparent molar volumes,
518 isentropic compressibilities and refractive indices in aqueous solutions of glycine plus NaCl. *Biophysical*
519 *Chemistry* **1998,** *74* (3), 165-173;

520 26. Tome, L. I. N.; Jorge, M.; Gomes, J. R. B.; Coutinho, J. A. P., Toward an Understanding of the Aqueous
521 Solubility of Amino Acids in the Presence of Salts: A Molecular Dynamics Simulation Study. *Journal of*
522 *Physical Chemistry B* **2010,** *114* (49), 16450-16459;

523 27. Tome, L. I. N.; Pinho, S. P.; Jorge, M.; Gomes, J. R. B.; Coutinho, J. A. P., Salting-in with a Salting-out
524 Agent: Explaining the Cation Specific Effects on the Aqueous Solubility of Amino Acids. *Journal of Physical*
525 *Chemistry B* **2013,** *117* (20), 6116-6128.

526 28. Ferreira, L. A.; Breil, M. P.; Pinho, S. P.; Macedo, E. A.; Mollerup, J. M., Thermodynamic Modeling of
527 Several Aqueous Alkanol Solutions Containing Amino Acids with the Perturbed-Chain Statistical Associated
528 Fluid Theory Equation of State. *Industrial & Engineering Chemistry Research* **2009,** *48* (11), 5498-5505;

529 29. Fuchs, D.; Fischer, J.; Tumakaka, F.; Sadowski, G., Solubility of Amino Acids: Influence of the pH value
530 and the Addition of Alcoholic Cosolvents on Aqueous Solubility. *Industrial & Engineering Chemistry Research*
531 **2006,** *45* (19), 6578-6584;

532 30. Ji, P.; Zou, J.; Feng, W., Effect of alcohol on the solubility of amino acid in water. *Journal of Molecular*
533 *Catalysis B: Enzymatic* **2009,** *56* (2–3), 185-188.

534 31. Cameretti, L. F.; Sadowski, G., Modeling of aqueous amino acid and polypeptide solutions with PC-SAPT.
535 *Chem. Eng. Process.* **2008,** *47* (6), 1018-1025;

536 32. Gross, J.; Sadowski, G., Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for
537 Chain Molecules. *Industrial & Engineering Chemistry Research* **2001,** *40* (4), 1244-1260;

538 33. GrosseDaldrup, J.-B.; Held, C.; Ruether, F.; Schembecker, G.; Sadowski, G., Measurement and Modeling
539 Solubility of Aqueous Multisolute Amino-Acid Solutions. *Industrial & Engineering Chemistry Research* **2010,**
540 *49* (3), 1395-1401;

541 34. Held, C.; Reschke, T.; Muller, R.; Kunz, W.; Sadowski, G., Measuring and modeling aqueous
542 electrolyte/amino-acid solutions with ePC-SAFT. *Journal of Chemical Thermodynamics* **2014,** *68*, 1-12;

543 35. Ji, P. J.; Feng, W., Solubility of amino acids in water and aqueous solutions by the statistical associating fluid
544 theory. *Industrial & Engineering Chemistry Research* **2008,** *47* (16), 6275-6279;

545 36. Klamt, A.; Krooshof, G. J. P.; Taylor, R., COSMOSPACE: Alternative to conventional activity-coefficient
546 models. *AIChE Journal* **2002,** *48* (10), 2332-2349.

547 37. Flory, P. J., Thermodynamics of High Polymer Solutions. *the Journal of Physical Chemistry* **1942,** *10* (51).

548 38. Huggins, M. L., Solutions of Long Chain Compounds. *Journal of Physical Chemistry* **1941,** *9* (440).

549 39. Jouyban-Gharamaleki, A.; Valaee, L.; Barzegar-Jalali, M.; Clark, B. J.; Acree Jr, W. E., Comparison of
550 various cosolvency models for calculating solute solubility in water–cosolvent mixtures. *International Journal of*
551 *Pharmaceutics* **1999,** *177* (1), 93-101.

552 40. Jouyban, A.; Khoubnasabjafari, M.; Chan, H.K.; Acree, W.E. Jr., Mathematical representation of amino
553 acids in binary aqueous-organic solvent mixtures at various temperatures using the Jouyban-Acree model.
554 *Pharmazie* **2005** 61: 789-792

555 41. Soto, A.; Arce, A.; K. Khoshkbarchi, M.; Vera, J. H., Measurements and modelling of the solubility of a
556 mixture of two amino acids in aqueous solutions. *Fluid Phase Equilibria* **1999,** *158–160* (0), 893-901;

557 42. Ferreira, L. A.; Macedo, E. A.; Pinho, S. P., Solubility of amino acids and diglycine in aqueous–alkanol
558 solutions. *Chemical Engineering Science* **2004,** *59* (15), 3117-3124;

559 43. Sadeghi, R., Modification of the NRTL and Wilson models for the representation of phase equilibrium
560 behavior of aqueous amino acid - electrolyte solutions. *Can. J. Chem.-Rev. Can. Chim.* **2008,** *86* (12), 1126-
561 1137;

562 44. Rivera, A.; Espinosa, A., Analysis of the NRTL parameters for amino acids in ethanol-water mixtures, using
563 the approach of relative solubility. *Afinidad* **2016,** *73* (573), 37-42.

564 45. Chen, C. C.; Crafts, P. A., Correlation and Prediction of Drug Molecule Solubility in Mixed Solvent Systems
565 with the Nonrandom Two-Liquid Segment Activity Coefficient (NRTL-SAC) Model. *Ind. Eng. Chem. Res.*
566 **2006,** *45* (13), 4816.

567 46. Jouyban, A.; Khoubnasabjafari, M.; Chan, H. K.; Acree, W. E., Jr., Mathematical representation of solubility
568 of amino acids in binary aqueous-organic solvent mixtures at various temperatures using the Jouyban-Acree
569 model. *Pharmazie* **2006,** *61* (9), 789-792.

570 47. Jouyban, A.; Shayanfar, A.; Panahi-Azar, V.; Soleymani, J.; Yousefi, B. H.; Acree, W. E.; York, P.,
571 Solubility prediction of drugs in mixed solvents using partial solubility parameters. *Journal of Pharmaceutical*
572 *Sciences* **2011,** *100* (10), 4368-4382.

573 48. Jouyban, A., Review of the cosolvency models for predicting solubility of drugs in water–cosolvent
574 mixtures. *J. Pharm. Pharm. Sci.* **2008,** *11* (1), 32;

575 49. Jouyban, A.; Shayanfar, A.; Panahi-Azar, V.; Soleymani, J.; Yousefi, B. H.; Acree, W. E.; York, P.,
576 Solubility prediction of drugs in mixed solvents using partial solubility parameters. *J. Pharm. Sci.* **2011,** *100*
577 (10), 4368;

578 50. Shayanfar, A.; Fakhree, M. A. A.; Acree, W. E.; Jouyban, A., Solubility of lamotrigine, diazepam, and
579 clonazepam in ethanol + water mixtures at 298.15 K. *J. Chem. Eng. Data* **2009,** *54* (3), 1107.

580 51. Ellegaard, M. D.; Abildskov, J.; O'Connell, J. P., Molecular Thermodynamic Modeling of Mixed Solvent
581 Solubility. *Industrial & Engineering Chemistry Research* **2010,** *49* (22), 11620-11632;

582 52. Sevillano, D. M.; van der Wielen, L. A. M.; Trifunovic, O.; Ottens, M., Model Comparison for the Prediction
583 of the Solubility of Green Tea Catechins in Ethanol/Water Mixtures. *Industrial & Engineering Chemistry*
584 *Research* **2013,** *52* (17), 6039-6048.

585 53. Cohn, E. J.; Edsall, J. T., Proteins Amino Acids and Peptides as Ions and Dipolar Ions. *J. Chem. Educ.* **1943**,
586 20 (8), 415.

587 54. Renon H., Prausnitz J. M., Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures,
588 AIChE J., **1968** 14 (1), S.135–144

589 55. Stefanis, E.; Panayiotou, C., Prediction of Hansen Solubility Parameters with a New Group-Contribution
590 Method. *Int. J. Thermophys.* **2008,** *29* (2), 568.

591 56. Prausnitz, J. M.; Lichtenthaler, R. N. ; Gomes de Azevedo, E. Molecular Thermodynamics of Fluid-Phase
592 Equilibria **1999** Prentice Hall PTR, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA 07458