I&EC research
Industrial & Engineering Chemistry Research

pubs.acs.org/IECR

Article

Modeling Solubilities for Amino Acids in Water as Functions of Temperature and pH

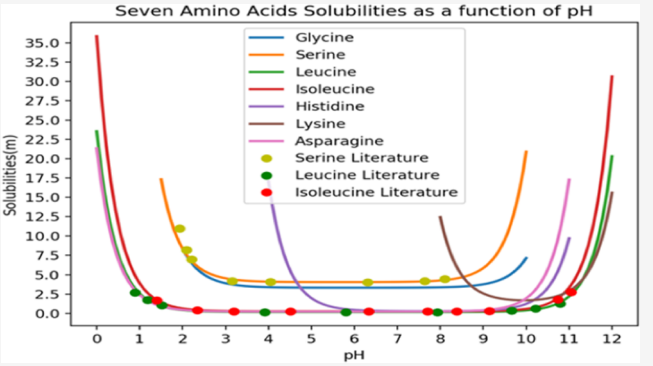
Tianxin Zhang, Ashwin Dravid, Jayanth Reddy, Lateef Aliyu, Jaeyoung Park, Zibo Wang, Kai-Wen Huang, Jinke Wu, Shuman Liu, Alan Stone, Michael Betenbaugh, and Marc Donohue*

Cite This: https://doi.org/10.1021/acs.iecr.3c00365

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Experimental results from the literature as well as from new experimental results presented here are used to determine parameters to correlate amino acid solubilities and activities in water. In particular, we have used a modified Larsen's UNIQUAC Functional-group Activity Coefficients (UNIFAC) group-contribution model to illustrate the solubility behavior of 10 amino acids in water including L-Histidine and L-Arginine that, to our knowledge, have not been published previously. New UNIFAC groups have been introduced to obtain the activity coefficients for the prediction of activities and solubilities as a function of temperature and pH. Prior models have not accounted for ionic interactions that affect pH dependent behavior and predictions. Hence, we examine whether a combination of UNIFAC and a new modified Debye−Huckel equation by Rapp et al.[1] is able to predict the activity coefficients of ionic species in solution at the high ionic strengths seen at high and low pH. We measured solubilities and fitted binary amino acid activity coefficients to estimate the new UNIFAC interaction parameters. The newly obtained UNIFAC parameters were used for the prediction of amino acid solubilities when they were predominantly charge-neutral. Then ionic interactions and pH-dependent chemical equilibria were added to calculate amino acid solubilities in aqueous solutions at different values of pH and temperature. The chemical equilibria required were calculated in a manner similar to Visual-MINTEQ[2] that will be described in a separate publication. The calculated solubilities were found to be in good agreement with our experimental measurements and with literature data.



INTRODUCTION

A biopharmaceutical is a pharmaceutical drug product manufactured in, extracted from, or partially synthesized from biological sources. The biological production of macromolecular products (primarily proteins) requires specific cell lines that meet product requirements for translational fidelity and post-translational modifications, particularly in the biopharmaceutical domain. At this time, the workhorses of biopharmaceutical production are Chinese hamster ovary (CHO) cell lines. However, CHO cells lack the genetic machinery required to produce biological, building-block molecules such as amino acids. Therefore, CHO cells need to be supplied with either media containing amino acids (for chemically defined media) or with macromolecules such as bovine serum albumin that can be catabolized into amino acids and used by the cells. Chemically defined cell-culture media often contains more than 50 compounds. Therefore, understanding the solubilities of compounds, particularly amino acids, in complex cell culture media[3] is very important for efficient biopharmaceutical production.

One of the top priorities of the biopharmaceutical industry is process intensification, i.e., creating cell cultures with high cell densities and high product titers. An important aspect of process intensification is the need to provide and maintain high nutrient concentrations to support high cell densities, high cell growth and high titer production. This requires creating new basal and feed media that have high levels of nutrients including amino acids. In many cases, process designs also include preparing media concentrates in which all the components are predissolved to ensure high bioavailability of all nutrients. Preparation of these media concentrates requires knowledge of the solubilities of each species in these complex, concentrated, multicomponent solutions.

Mammalian cell culture media can contain 50 to 100 different compounds that have complex intermolecular interactions with each other.[3] Unfavorable interactions can lead to precipitation of nutrients from cell culture, making

Received: May 4, 2024
Revised: November 12, 2024
Accepted: November 13, 2024

them unavailable for consumption. Spectroscopic analysis of these precipitates has revealed the presence of as many as 10 amino acids in the precipitate. One common strategy employed in concentrating amino acids with relatively low solubilities is to prepare a separate feed media bucket at low or high pH to increase the solubilities of the sparingly soluble compounds. Hence, this study focuses on modeling amino acid solubility behavior as a function of pH and temperature.

There are two goals of this article. The first is to present new experimental data on the solubilities of several amino acids; the second is to demonstrate that these solubilities can be correlated using a combination of an Antoine-like equation for infinite dilution activity as a function of temperature, an asymmetric activity coefficient calculated using a combination of UNIQUAC Functional-group Activity Coefficients (UNIFAC) and a newly published modified Debye−Hückel equation by Rapp et al.,[1] and solubility products, K_{sp}, to describe the variation of speciation (cation, Zwitterion, and anion) because the protonation or deprotonation levels are strong functions of pH.

Doing this required a methodology consisting of five steps. First, we measured aqueous solubilities of the α-amino acids at temperatures of 298 and 310 K.

Second, we compiled the available literature data[4−12] for the solubilities of ten amino acids as a function of temperature and pH.

Third, we compiled previously reported literature activity coefficient data[13−17] for all the amino acids, and fitted those simultaneously with a modified Larsen's UNIFAC model to get interaction parameters which were used to calculate solubilities of individual amino acids in water. While the effect of temperature on solubility can be predicted using ΔH_{fusion},[18,19] here temperature-dependent solubility data were correlated with an Antoine-like equation.

Fourth, the effects of ionic interactions were added to UNIFAC to calculate changes in speciation and, hence, solubilities considering both temperature and pH effects simultaneously.

The last step was to combine contributions to the activity coefficients from UNIFAC and ionic interactions with chemical equilibria calculated with solubility products to obtain estimates for species concentrations and activities. This follows the work of Pinho et al.[9] who also combined UNIFAC with an ionic contribution to the activity coefficient calculated using the Debye−Huckel equation and with calculations of chemical complexation equilibria to predict the behavior of 14 amino acids and 5 small peptides as a function of pH in water. The major distinction of our work from that of Pinho et al. is that our work uses a new species-independent generalization of the Davies modification to Debye−Huckel theory for the ionic contribution to the activity coefficient by Rapp et al.[1] Hence, though this model and that of Pinho et al. are structured similarly, the fitted parameters and the dependence on ionic strength are quite different. It is generally believed that the Debye−Huckel term used by Pinho et al. should not be used above 10^{-3} molar or at most not above 10^{-2} molar. Since the concentration of leucine, the least soluble amino acid considered here, is about 6×10^{-2} at its minimum (the isoelectric point), and the concentration of serine is about 4 molar, the Debye−Huckel theory is considerably outside its range of utility for these solutions. Conversely, though it is a species-independent equation, the equation by Rapp et al. gives quite accurate estimates of the ionic contribution to the activity coefficient up to 2 molal and gives reasonable estimates at higher concentrations.

Several versions of UNIFAC[17,20,21] have been proposed to date. These are used widely to represent the activity coefficients for a very wide variety of substances including amino acids in aqueous solutions. In addition to the work of Pinho et al, Gupta and Heidemann correlated the activity coefficients of amino acids in water with just Larsen's UNIFAC model.[11] Other correlations of binary amino acid solubilities (1 amino acid + water) have been made with other thermodynamic models such as non-random two-liquid model (NRTL),[7] perturbed chain-statistical associating fluid theory (PC-SAFT)[8,22,23] and Wilson's equation.[13] Unfortunately, compared to the extensive data set used by Pinho et al., the experimental data measured and used to determine parameters for most of these models are quite limited.

In addition to the thermodynamic effects covered in this publication, there are other phenomena that need to be considered. For example, in addition to the changes in speciation that occur with pH, many amino acids form salts in aqueous solution that can precipitate. Two examples are sodium glutamate and sodium aspartate. The complexes may have solubilities that are lower than those of the un-ionized amino acids at some conditions and therefore these salts may precipitate at a lower amino acid concentration than the pure amino acid itself. Therefore, inclusion of the contribution of ionic interactions to the activity coefficient model and inclusion of chemical complexation serve as key points for predicting solubility of amino acids in complex pharmaceutical media.

UNIFAC is used to calculate the nonionic contributions to the liquid-phase activity coefficients of solutes. As UNIFAC is a group-contribution method, a large number of compounds can be modeled using a combination of a smaller number of group−group interaction parameters provided the compounds modeled contain common UNIFAC groups. Additionally, if all the chemicals share most groups, the data required for prediction is simpler to obtain experimentally since only a small number of ternary systems (2 amino acids in water) need to be measured to ensure that ternary and multicomponent behavior is reasonably accurate. The UNIFAC groups need to be defined judiciously in order to balance the interests of scale and model flexibility. For amino acids in pharmaceutical media, Gupta and Heidemann[11] as well as Pinho et al.[13] used the UNIFAC models to correlate activities of amino acids in water using a particular set of UNIFAC groups. The particular UNIFAC group sets used were small, and were the minimally required sets to describe a small subset of biologically relevant amino acids studied in these papers. These could therefore only predict activity coefficients for a few amino acids and peptides. In this paper, we define several new UNIFAC groups not previously reported in the literature like the arginine and histidine side chains.

MATERIALS AND METHODS

Materials. Ten amino acids: Glycine,[24] L-Alanine, L-Serine, L-Threonine, L-Valine, L-Leucine, L-Isoleucine, L-Histidine, L-Lysine, and L-Arginine were obtained from Sigma-Aldrich. These all had purities greater than >99.5% (mole fraction). Solutions were prepared using purified water from a Millipore filter system (18MΩ).

Experimental Procedure. Saturated solutions were made by mixing an amount of amino acid powder to Millipore water

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Table 1. Collected[25,26] and Calculated Bondi Volumes and Surface Areas for Functional Groups

	H_2O	α-CH_2	α-CH	NH_2	COOH	Sc CH_3	Sc CH_2	Sc CH	OH	hist-ring	arg-triN
R	0.92	0.6744	0.4469	0.6948	1.3013	0.9011	0.6744	0.4469	1	1.4515	2.1354
Q	1.4	0.54	0.228	1.15	1.224	0.848	0.54	0.228	1.2	1.248	1.87164

in excess of that for saturation. Amino acid powders were weighed using an analytical balance with an accuracy of ±0.1 mg. The weighed amino acids were suspended in 20 mL water in 50 mL conical tubes. These solutions were equilibrated in a thermostated orbital mixer for a minimum of 24 h. Following equilibration, the saturated amino acid solutions were filtered using 0.45 μm poly(vinylidene fluoride) (PVDF) filters to obtain saturated solutions devoid of solid particles. The saturated amino acid solutions then were diluted significantly so that concentrations were within the range where the high-performance liquid chromatography (HPLC) ultraviolet (UV) detector gave linear readings for peak area versus concentration in solutions of known compositions. Calibration lines for each amino acid were made using the origin and at least three known concentrations within the linear range of detection. Solubility measurements were conducted at 20 and 37 °C.

HPLC Analysis. Concentration measurements were conducted using high performance liquid chromatography (HPLC) equipped with a diode array detector (DAD) that allowed for measuring UV absorption at 338 nm. These concentration measurements were performed using an automated precolumn derivatization HPLC method. Briefly, amino acid samples were first neutralized in borate buffer to ensure that the amino terminus is neutralized. Subsequently, primary amines on the amino acids were reacted with ortho-phthaldehyde (OPA). The derivatized product enables the separation of different amino acids using a reverse-phase mode and it enables sensitive UV detection at 338 nm. For each condition, two saturated solutions were prepared separately and each of these two samples was measured with at least three injections using the HPLC amino acid method. The amino acid concentration was calculated from the diluted samples, and the solubility limit was determined by multiplying the measured concentration by the dilution factor used for the sample. The average solubility measurements of these amino acid solutions are reported, along with the standard deviations (Supporting Table S1).

Derivatized amino acids were separated on a reverse phase Agilent Poroshell HPH-C18 column. This column possesses a stationary phase suitable for retaining hydrophobic molecules. Mobile phase A consisted of an aqueous buffer containing 10 mM sodium phosphate, 10 mM sodium tetraborate, and 5 mM sodium azide at pH 8.2. Mobile phase B was an organic buffer composed of 45% methanol, 45% acetonitrile, and 10% water. The two mobile phases were used to generate a gradient elution from 98 to 43% A in 14 min. Initially, all derivatized amino acids are retained on the column at 98% A. As the composition of mobile phase B increases, amino acids begin to elute from the column. The increasing concentration of the organic solvent reduced hydrophobic interactions between the stationary phase and the molecules, facilitating their elution. Due to variations in their side chains, each amino acid exhibited different degrees of hydrophobicity, requiring different concentrations of organic solvent for their elution.

THERMODYNAMIC MODEL: THEORY AND RESULTS

Activity coefficient data were collected from the literature.[13−17] The minimization function is the sum of squares of the unweighted differences between the data and calculations as

$$\mathrm{OBJ} = \sum_j (\gamma_j^{\mathrm{calc}} - \gamma_j^{\mathrm{data}})^2 \quad (1)$$

The variables γ_j^{calc} and γ_j^{data} represent the activity coefficients calculated by the model and those determined experimentally, respectively. The Levenberg−Marquardt algorithm optimizes the model parameters by minimizing the objective function OBJ, which quantifies the difference between calculated and observed activity coefficients. This optimization directly enhances the model's accuracy and predictive capability, with a lower OBJ indicating a closer match to experimental data

$$\ln(\gamma_i^*) = \ln(\gamma_i^{*,\mathrm{C}}) + \ln(\gamma_i^{*,\mathrm{R}}) + \ln(\gamma_i^{*,\mathrm{ionic}}) \quad (2)$$

Where the subscripts C and R refer to the combinatorial and residual contributions to the LUNIFAC activity coefficients.[20] A new modified Debye−Huckel equation by Rapp et al.[1] is used to calculate the long-range ionic contribution to the activity coefficient. This new equation has the advantage that it gives much more reasonable predictions at ionic strengths of 0.5−2.0 mol/kg water without needing species-specific parameters. Other extended Debye−Hückel equations typically are limited to ionic strengths of less than 0.5 mol/kg.

UNIFAC Prediction and Activity Coefficients Model. Activity coefficients for ionic species in aqueous solution are defined asymmetrically, i.e., the activity coefficient is defined as unity at infinite dilution. Precipitation of a particular species from solution takes place when the activity of that species in solution in solution drops below the ratio of standard state fugacities. This ratio is a function of temperature only and was fitted to the temperature-dependent solubility data and activity coefficients.

Throughout this project, the assumption is made that the precipitated solid is a pure, crystalline amino acid and effects of phenomenon such as ion hydration and adsorption of aqueous species onto solid surfaces are neglected. Finally, in addition to the group−group interaction parameters determined by fitting experimental data, UNIFAC requires structural group-specific parameters (*R* and *Q*). Most of the structural parameters used here were taken from prior calculations in literature, but others, such as those for the arginine side chain, reported in Table 1, were calculated using Bondi's method.

Group−group interaction parameters, reported in Table 2, were obtained by fitting amino acid activity coefficient data from binary mixtures[7,8,10,16] (one amino acid and water) to the UNIFAC model using the lsqnonlinear function (the Levenberg−Marquardt algorithm) in MATLAB. For systems where no experimental data were available but thermodynamic parameters determined by fitting data to other models, particularly for the PC-SAFT model,[14,22,23] these parameters were used to generate "data" to fit UNIFAC parameters. In addition, there are a number of sources of osmotic coefficients

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Table 2. Regressed Interaction Parameter Matrix for Functional Groups

molecular group	H_2O	α-CH_2	α-CH	NH_2	COOH	Sc CH_3	Sc CH_2	Sc CH	OH	hist-ring	arg-triN
H_2O	0	1074.27647	1074.27647	589.294118	−39.914758	687.633445	687.633445	687.633445	−46.058912	137.45157	162.809473
α-CH_2	−373.28877	0	0	917.124675	−423.9838	1630.77474	1630.77474	1630.77474	−53.56731	−6.4378521	−6.0012592
α-CH	−373.28877	0	0	917.124675	−423.9838	1630.77474	1630.77474	1630.77474	−53.56731	−6.4378521	−6.0012592
NH_2	−143.0016	1847.92589	1847.92589	0	1420.414	147.936251	147.936251	147.936251	−450.59515	−33.407015	−30.791652
COOH	35.8428469	1731.80301	1731.80301	1811.67115	0	1955.85366	1955.85366	1955.85366	−509.3219	−35.691281	−32.758066
Sc CH_3	−445.72264	1852.94817	1852.94817	−85.764749	1310.07747	0	0	0	1318.79517	−15.226986	−43.354016
Sc CH_2	−445.72264	1852.94817	1852.94817	−85.764749	1310.07747	0	0	0	1318.79517	−15.226986	−43.354016
Sc CH	−445.72264	1852.94817	1852.94817	−85.764749	1310.07747	0	0	0	1318.79517	−15.226986	−43.354016
OH	−106.04071	798.479899	798.479899	1603.11386	−669.14955	1989.07367	1989.07367	1989.07367	0	0	0
hist-ring	104.657343	563.370669	563.370669	580.31348	−178.01733	495.100866	495.100866	495.100866	0	0	0
arg-triN	587.673713	884.650126	884.650126	1419.84002	−527.85625	1604.45467	1604.45467	1604.45467	0	0	0

for amino acids in water were measured and reported for binary mixture;[14,23] the amino acid activity coefficients were obtained from this data using the Gibbs−Duhem equation.[15] Table 1 contains the functional groups and their Bondi parameters for the ten amino acids studied here. Table 3 describes the UNIFAC groups that each amino acid is made of. Isoleucine is an isomer of leucine as both amino acids have the same molecular formula and UNIFAC groups but differ in structure, and thus UNIFAC predicts the same activity coefficients for both molecules.

Table 3. Functional Groups for 10 Amino Acids

amino acid name	groups
glycine	−α-CH_2, −NH_2, −COOH
alanine	−α-CH, −NH_2, − COOH, −Sc CH_3
serine	−α-CH, −NH_2, −COOH, −OH
threonine	−α-CH, −NH_2, −Sc CH_3, −Sc CH
valine	−α-CH, −NH_2, −COOH, −Sc CH_3 (*2), −Sc CH
leucine	−α-CH, −NH_2, −Sc CH_3 (*2), −Sc CH_2, −Sc CH
isoleucine	−α-CH, −NH_2, −Sc CH_3 (*2), −Sc CH_2, −Sc CH
histidine	−α-CH, −NH_2, −COOH, −Sc CH_2, −histidine ring
lysine	−α-CH, −NH_2 (*2), −COOH, − Sc CH_2 (*4)
arginine	−α-CH, −NH_2, −COOH, − Sc CH_2, −arginine triN

The interaction parameters for ten amino acids were obtained after fitting binary system activity coefficient data simultaneously as described earlier.

Figure 1 compares experimental the calculated activity coefficients. Activity coefficients are shown for concentrations ranging from 0 molal to the amino acid solubility. As Leucine and Isoleucine have the same functional groups, the Leucine line in the diagram represents the behavior of both Leucine and Isoleucine (Table 4).

Temperature Dependence of Amino Acid Solubility. Although the activity coefficient is a function of both temperature and composition, the ratio of solid and liquid standard state fugacities[16] is a function of temperature alone. The temperature dependence of the ratio of standard state fugacities has been fitted using an Antoine-like equation

$$\ln\left(f_i^{*s}/f_i^{*}l\right) = A_i + \frac{B_i}{T} + C_i \ln\left(T\right) \text{ (where } T \text{ is in Kelvin)} \quad (3)$$

where i is the component index, the activity is in a molal concentration scale, m_i is the molality of the i^{th} component, A_i, B_i and C_i are component specific temperature dependence parameters and T is the absolute temperature. This is an empirical method for calculating the solubility variation with temperature based on solubility data since the heat of dissolution for amino acids is difficult to obtain experimentally.

Experimental solubilities at various temperatures were multiplied with activity coefficients at those temperatures[6−9,11,15] to obtain ratio of solid to liquid standard state fugacities. Experimental measurements existed for the solubilities at different temperatures, but activity coefficient measurements generally were not available as functions of temperature. The activity coefficients were calculated from UNIFAC using interaction parameters fit from binary activity coefficient data. Additionally, as UNIFAC interaction parameters are temperature independent; temperature is used purely as an input and thus activity coefficients for a mixture at any

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

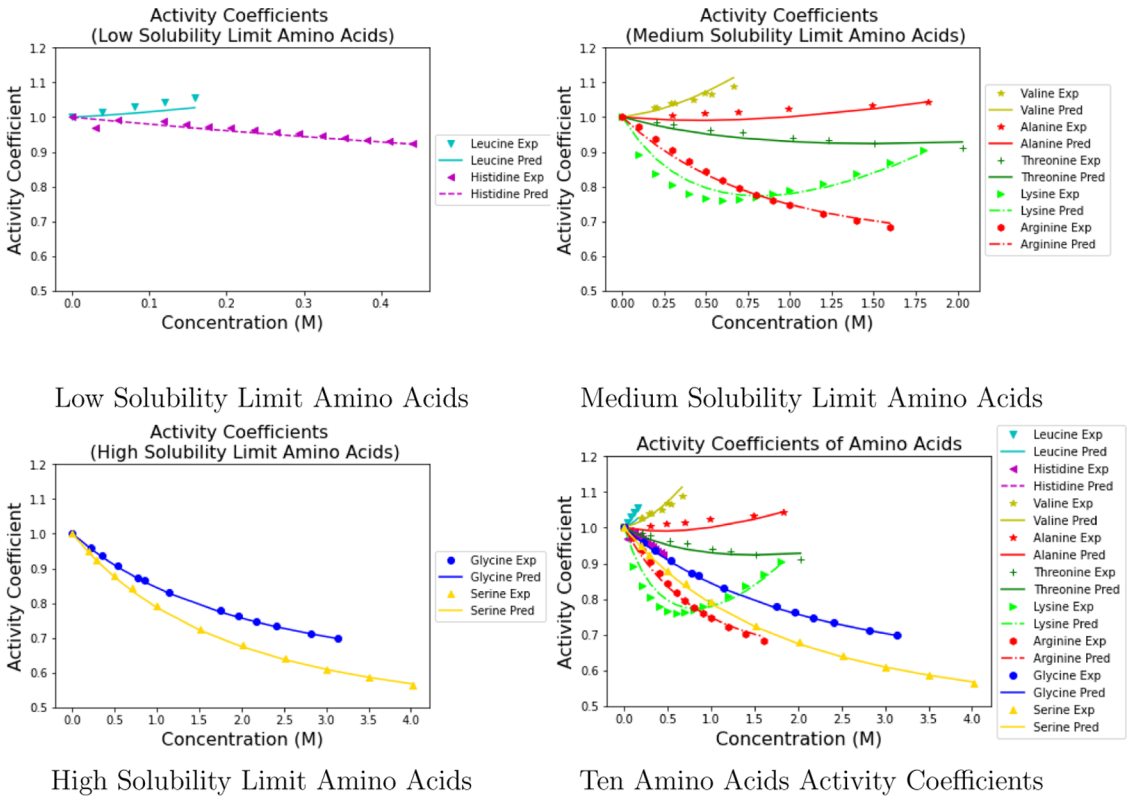


Figure 1. Predictions compared with literature data[4,5,7,9−11] of amino acids activity coefficients for all ten amino acids.

Table 4. Mean Difference Between Experimental[4,5,7,9−11] and Calculated Amino Acid Activity Coefficients

amino acid	mean percentage error (%)
glycine	0.37
alanine	1.25
serine	0.49
threonine	1.03
valine	2.54
leucine	1.61
isoleucine	1.61
histidine	0.72
lysine	2.34
arginine	0.87

Table 5. Mean Percentage Error for Binary Amino Acid Solubility Calculations Compared With Literature and Experimental Measurements

amino acid	mean percentage error (%)
glycine	0.23
alanine	0.44
serine	0.00
threonine	0.16
valine	0.22
leucine	0.55
isoleucine	0.06
histidine	0.00
lysine	0.00
arginine	0.00

temperature can be obtained. The objective function used for the regression was a simple least-squares function of the form.

$$\mathrm{OBJ} = \sum_j (a_j^{\mathrm{cal}} - a_j^{\mathrm{exp}})^2 \tag{4}$$

Antoine parameters are given in Table 6, and the mean percentage error values in Table 5. These calculations are used as activity thresholds and combined with chemical equilibria as described in the section describing the incorporation of pH effects.[10] The results are satisfactory both quantitatively and qualitatively, as has been visualized in Figure 2.

pH Dependence of Amino Acid Solubilities. Amino acid solubility varies significantly with pH, and by altering the pH of a solution, the solubility of an amino acid can be greatly increased. In cell culture feed media, it is crucial to minimize the volume of water added while maximizing nutrient concentration. Therefore, feed media often contain amino acids at much higher concentrations than those found in basal media. For highly soluble amino acids, these concentrations can be achieved directly. However, for sparingly soluble amino acids, adjusting the pH of the feed media can enhance their solubility, achieving the necessary concentrations without causing precipitation. Understanding how amino acid solubility varies with pH is essential for designing high-concentration feed media. Most amino acids show substantial increases in solubility at extreme pH values due to changes in protonation levels. These protonation levels are crucial because only the neutral species can precipitate. At their isoelectric point (pI), amino acids predominantly exist as zwitterions with no net charge, making them prone to precipitation by being in equilibrium with their neutral (molecular) form. Before discussing the impact of pH on protonation levels, it is important to consider all the species present in solution.

$$\mathrm{HAA\ (solid)} \overset{K_{D1}}{\rightleftharpoons} \mathrm{HAA\ (dissolved\ molecular)} \overset{K_{D2}}{\rightleftharpoons} \mathrm{HAA^{\pm}}$$

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

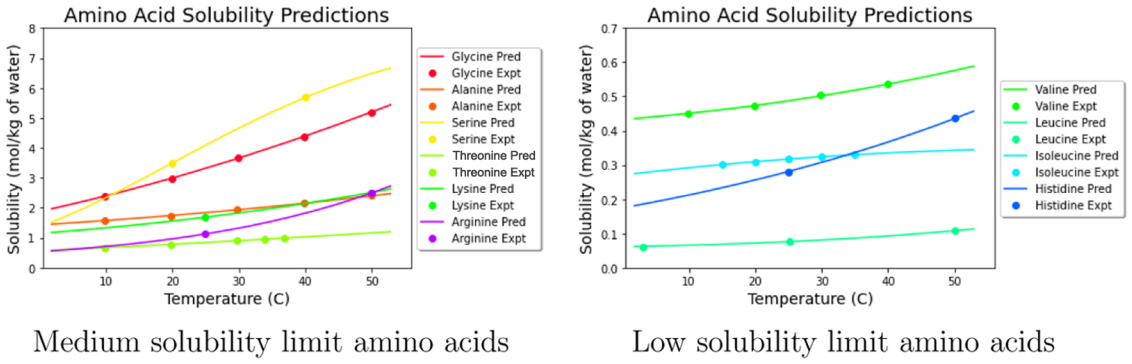


Figure 2. Predictions of solubilities of binary amino acids compared with data[7,8,10,16] as a function of temperature.

Where HAA is the amino acid and $HAA^{\pm}$ is the zwitterionic form.

Most amino acids exist in the zwitterionic form as K_{D2} is quite large (Greenstine and Winitz, 1961). Thus, for simplicity, we assume that the solid phase of amino acid can directly generate zwitterionic amino acid ($HAA^{\pm}$) and ignore the intermediate dissolved molecular species.[9] This also has the advantage of being able to use the solubility values previously developed. Strictly speaking, it is the uncharged species that precipitates and not the zwitterion (which is charged but has a net charge of zero). If the dissolved molecular species concentration were then to be considered to be the solubility, all of the previous solubility values obtained, which were obtained for total aqueous concentrations for a species, would need to be divided by $(1/(1 + K_{D2}))$ to obtain the values for uncharged species. It is important to note that this simplification is only possible because the zwitterion and the uncharged dissolved molecular species always exist in a fixed ratio. Henceforth, both the zwitterionic and uncharged aqueous species shall be referred to jointly as HAA (Table 6).

Table 6. Solubility Constant Parameters A, B, C Fitted for Amino Acids of Interest

amino acid name	A	B	C
glycine	−13.12	1112.79	3.11
alanine	−74.80	−2398.99	11.83
serine	285.11	14967.51	−41.07
threonine	−0.53	1180.84	0.75
valine	−82.02	−3175.39	12.41
leucine	−180.96	−7085.65	27.16
isoleucine	58.89	2991.11	−8.76
histidine	−35.59	−155.47	5.92
lysine	−142.91	−4832.61	22.31
arginine	−178.45	−5777.16	27.89

The effect of protonation levels on amino acid solubilities can be explained through the following example. Consider a neutral amino acid with single amine and acid groups (such as Serine) denoted as HAA. This amino acid can undergo additional protonation by protonation of the COO^- functional group or deprotonation through deprotonation of the NH^{3+} group. Thus, the amino acid can exist in AA^- and $HAAH^+$ forms.

$$H^+ + AA^- \overset{K_1}{\rightleftharpoons} HAA^{\pm}$$

$$HAA^{\pm} + H^+ \overset{K_2}{\rightleftharpoons} HAAH^+$$

Simultaneously, another equilibrium for water governing the pH exists.

$$H_2O \overset{K_w}{\rightleftharpoons} H^+ + OH^-$$

K_w is the equilibrium constant of water where K_w equals to 1.0×10^{-14} at 25 °C. The pH is additionally assumed to be changed by the addition of either NaOH or HCl as required, which are assumed to dissociate completely. At pH values near the isoelectric point, the neutral zwitterionic forms of the amino acids are the dominant species present, and thus are in equilibrium with the dissolved molecular species which can precipitate. At pH values that differ significantly from the pI, most of the aqueous amino acid is either additionally protonated (at low pHs) or deprotonated (at high pH), and only a very small proportion is present in the neutral zwitterionic form. Thus, the total aqueous concentration of amino acid necessary for the concentration of the dissolved molecular form to reach the solubility limit is very high, and thus a much larger amount of amino acid can be dissolved in solution. For each amino acid, the sum of percentages of all species (protonated, zwitterionic, deprotonated) at any pH must be 100. For amino acids with additional acidic or basic functional groups in their side chains, an additional equilibrium equation would be present to account for the potential protonation or deprotonation of the side chain.[12] For example, aspartic acid and glutamic acid have more than one anionic species and lysine will have 2 cationic species. K_1 and K_2 are the equilibrium constants for generating zwitterionic form and protonated form of amino acids.[8] K_1 and K_2 are also used to determine the presence of a form at specific pH. Table 7 displayed the ten amino acids that were predicted for their pKa1, pKa2, and pKx values, in addition to the pI value, to facilitate the calculation of the pH dependence of solubility, as illustrated in workflow Figure 3. Furthermore, it provided insight into the amino acid percentage as a function of pH, as demonstrated in Figure 4. The identity of the acid or base utilized to affect the pH change is not specified, and thus ion-specific interactions from the counterion are not incorporated. In principle, the short-range interactions that are specific to each ion also can be incorporated by introducing it as a functional group in UNIFAC. Additionally, any complexes formed with the counterion and their impact on solubility are not included here, but will be presented in a forthcoming paper (details are in the Discussion section). The algorithmic loop of calculating the solubility given a particular pH value is depicted in Figure 3.

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Table 7. Ten Amino Acid pK and pI Values, Where pK_1 Is the Negative of the Logarithm of the Dissociation Constant for the −COOH Group, pK_2 Is the Negative of the Logarithm of the Dissociation Constant for the $-NH_3$ Group, and pK_x Is the Negative of the Logarithm of the Dissociation Constant for Any Other Group in the Molecule, and pI Is the pH at the Isoelectric Point

amino acid	pK_1	pK_2	pK_x	pI
glycine	2.34	9.6		5.97
alanine	2.344	9.868		6.106
serine	2.21	9.15		5.68
threonine	2.088	9.1		5.594
valine	2.286	9.719		6.003
leucine	2.33	9.74		6.04
isoleucine	2.32	9.76		6.04
histidine	1.82	9.17	6.00	7.59
lysine	2.18	8.95	10.53	9.74
arginine	2.17	9.04	12.48	10.76

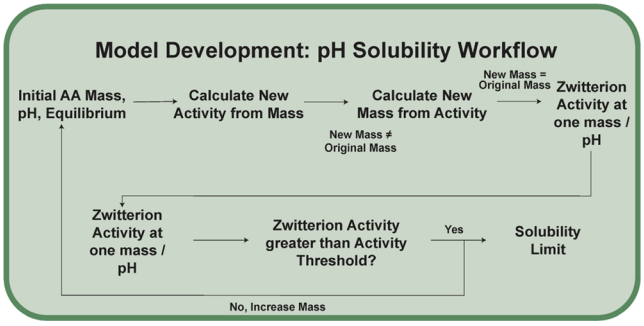


Figure 3. pH solubility workflow for binary amino acids.

Figure 4 shows L-Serine and L-Lysine percentage species distributions as a function of pH at 25 °C. L-Lysine is a basic amino acid (an amino acid with a basic functional group in its side chain), and it has 2 cationic forms and 3 pK_a values. L-Serine is a neutral amino acid (amino acid with neutral groups in the side chain that do not participate in protonation and deprotonation) and has only two pK values. This difference introduces an additional protonation level for L-Lysine and causes the difference in pI as seen above.

Figure 5 illustrates the calculated L-Serine solubilities at 0 and 20C solubilities as a function of pH. The temperature dependence of solubilities was incorporated using the Antione-like equation and the ionic component of activity coefficients was included as described earlier. Comparison of the predictions with literature values exhibits a good match, providing evidence for the ability to utilize the temperature and pH effect frameworks simultaneously. Increasing pH will cause the deprotonated amino acid to be predominant, while decreasing the pH causes the protonated form to be predominant in solution. In either case, the solubility increases as the pH departs from the isoelectric point and approaches extreme values.

Figure 6 shows L-Serine and L-Leucine solubility as a function of pH at 25 °C. For both amino acids, the subfigure on the left portrays solubility predictions in molality on a linear scale and that on the right portrays the same on a semilog scale.

Our correlations are not accurate at very high or very low pH because the assumption that the ionic contribution to the activity coefficient is species independent breaks down at these conditions.[1] The assumption of a species-independent ionic activity coefficient is reasonable when the pH is not significantly different from the pI of the amino acids considered. However, this assumption breaks down at extreme pH (below 2 and above 11 for most amino acids).

Figure 7 shows solubility data (in molality) for solutions containing 10 different amino acids as a function of pH at 25 °C on a linear and semilog scale. Our calculations are in good agreement with data from the literature for L-Serine, L-Leucine and L-Isoleucine.

DISCUSSION

Variations in solubility and activity of ten amino acids with temperature and pH were successfully correlated. The ability to calculate solubilities as a function of pH will help basal and feed media development, especially for feed media containing amino acids with low solubilities at their isoelectric points. The ability to calculate solubilities as a function of temperature will help to ensure that precipitation does not occur when the media is stored at low temperatures in certain parts of the supply chain.

Deviations between experimental and calculated activities, activity coefficients, and solubilities are lower than those in prior publications. Table 4 shows the mean error values for amino acid activity coefficient calculations compared with literature data. All calculated error values were less than 2.6% and were in good agreement with literature values. It was

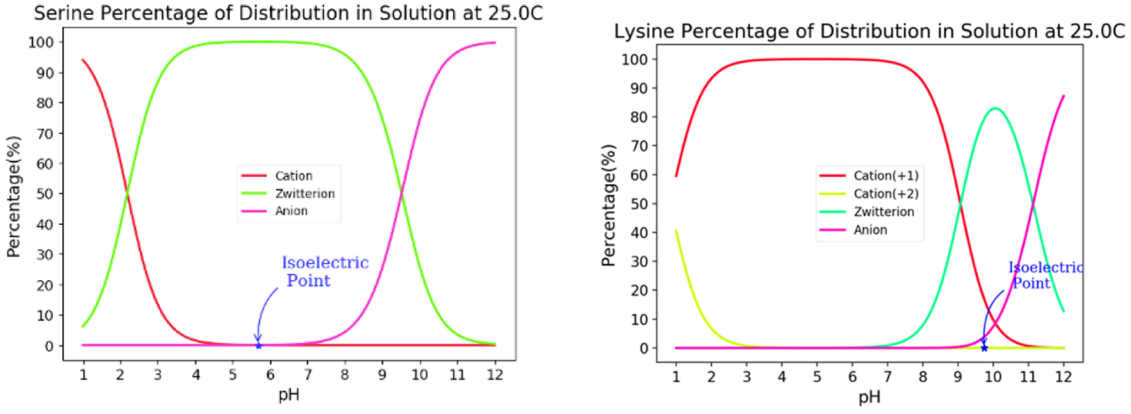


Figure 4. Example of L-Serine and L-Lysine ionic species distribution in aqueous solution at 298.15 K as a function of pH.

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

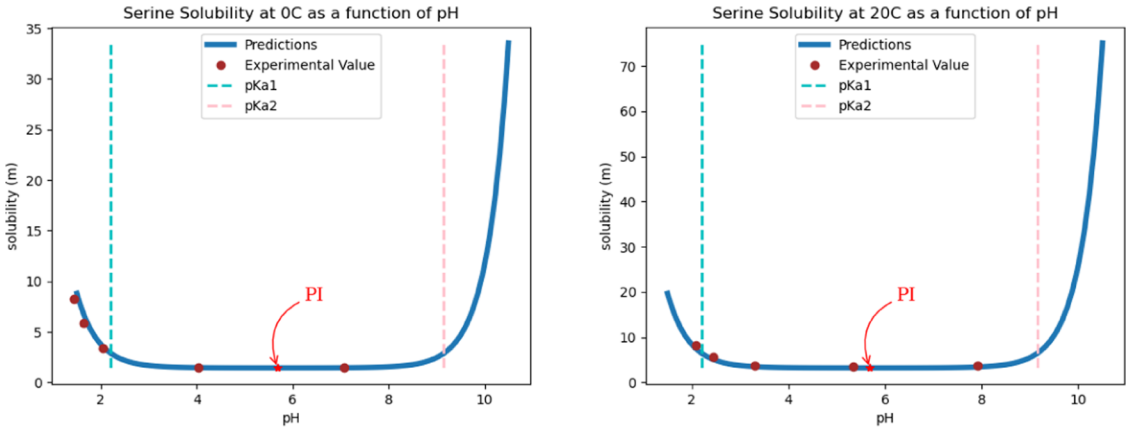


Figure 5. L-Serine solubilities as a function of pH at 0 and 20 °C.

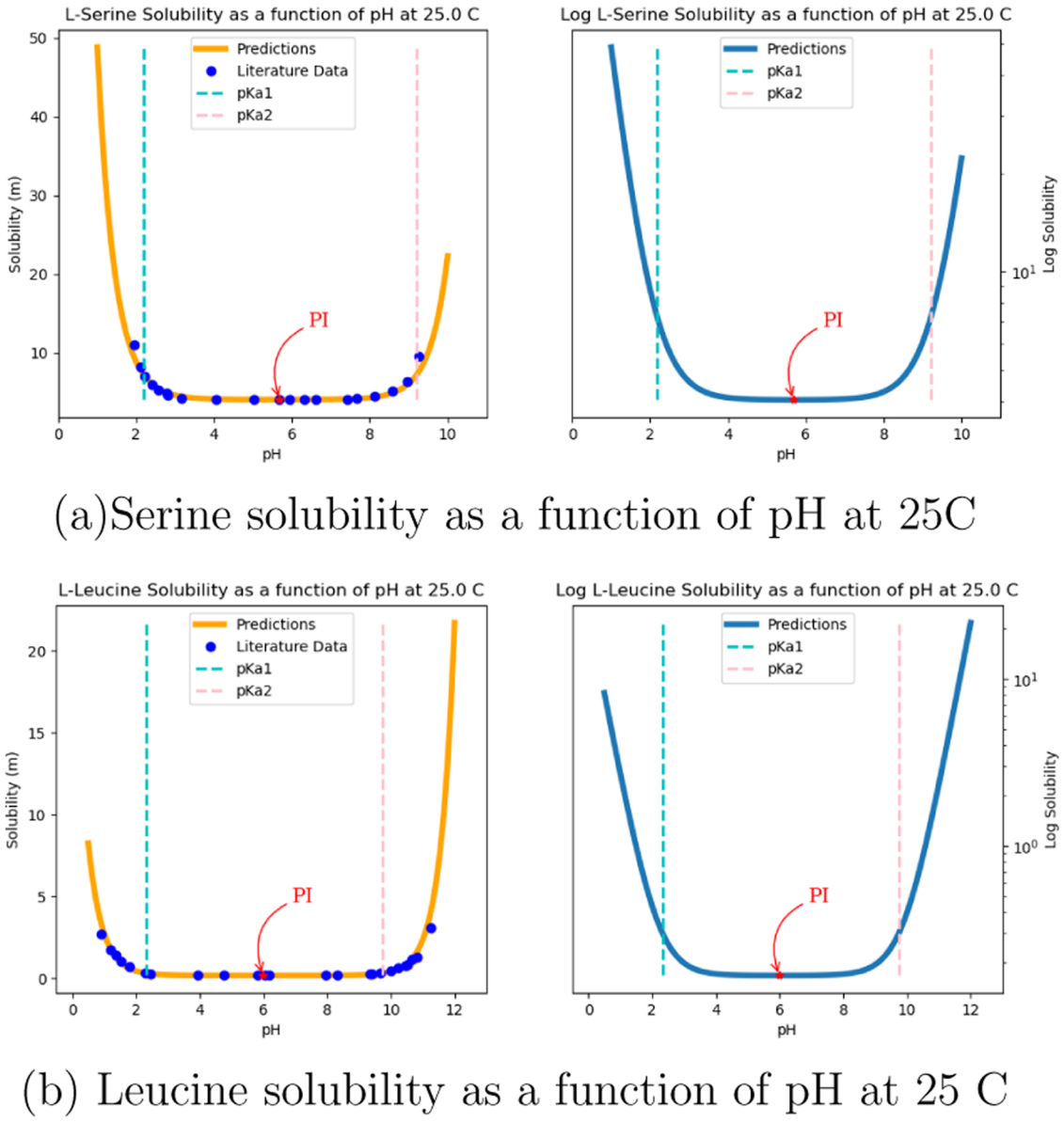


Figure 6. Example calculations compared with literature data[7,8] for pH dependence of solubility of L-Serine and L-Leucine in linear and semilog scales.

found that simultaneously fitting the interaction parameter matrix from multiple literature sources for amino acids improved the accuracy of the calculations. Table 5 shows the mean error for calculated binary amino acid solubilities

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

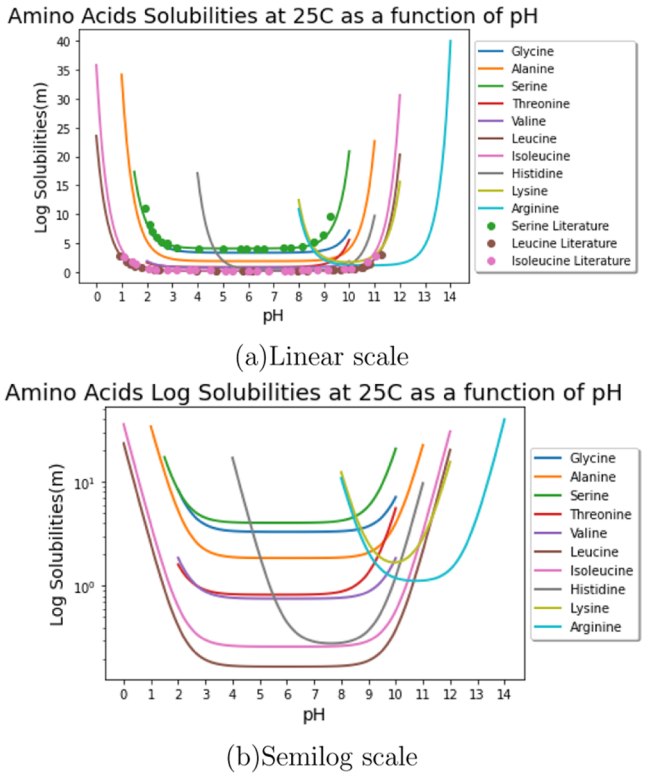


Figure 7. Solubility predictions and experimental data[7] of amino acids as a function of pH from 0 to 14 at 25 °C.

compared to experimental values. All are less than 0.6%. Table 8 summarizes the L-Serine solubility as a function of pH for

Table 8. Serine Solubility Prediction Errors at Different Temperatures as a Function of pH

serine solubilities (°C)	mean percentage error (%)
0	8.02
20	8.18

temperatures of 0 and 20 °C. The mean error is around 8%. The larger error is attributable to values at extreme pH where the ionic activity coefficient model may not be applicable. Table 9 shows the mean error for amino acid solubilities at various pHs at 25 °C. The errors obtained range from 3 to 10%.

Table 9. Mean Percentage Error Calculation for Amino Acids with Literature Data[7]

amino acid solubilities	mean percentage error (%)
Serine	3.22
Leucine	9.90
Isoleucine	9.67

Because UNIFAC is a group contribution method and because the ionic contribution to the activity coefficient has no ion-specific parameters, the model presented here can be extended to cover all 20 amino acids relevant to mammalian cell cultures with a limited amount of additional data. Similarly, the model can be extended to dipeptides by adding only one additional group for −CONH-. Thus, in principle, fitting activity coefficient data for a small number of dipeptides would be sufficient to obtain activity coefficient predictions for all relevant dipeptides. Solubility data, as well as data about protonation effects with pH would still be required.

Mixtures of amino acids containing three or more components (2 or more amino acids in water) also can be calculated. The effects of pH may be neglected for amino acids with similar pIs. Furthermore, ionic contributions to the activity coefficient may also be neglected since almost all of the amino acids exist as a zwitterion that is in equilibrium with the uncharged dissolved amino acid molecule. Some ternary solutions, such as serine-glycine-water, introduce no new additional interactions among groups that were not already present in serine-water and glycine-water systems. Therefore, solubility calculations for such ternary systems can be made directly from binary activity coefficient and solubility data. Some others, such as tryptophan-serine, will introduce new interactions in a ternary solution. This is because the ternary solution for these would contain the tryptophan side chain group from tryptophan and the hydroxyl group from serine in the solution simultaneously. Since these groups are not present simultaneously in binary solutions of these or other amino acids. Hence, ternary solubility data would be required to fit the value for the interaction parameter between the new groups.

Additionally, ternary, quarternary or other multicomponent systems which contain amino acids with large differences in pI would require the consideration of all of their equilibria simultaneously. This is because the addition of an amino acid with a neutral side chain to a solution containing an amino acid with an acidic or basic side chain will result in the solution being at a pH between both of their pIs, thus increasing solubility for both, the extent of which for each amino acid is based on the difference between the pH of the solution and the pI for the respective amino acids. Thus, implementation of the model developed here would need to be generalized to work with multiple amino acid equilibria simultaneously in order to consider such solutions. Additionally, if these ternary systems introduce new interactions that were not present in binary systems, the new UNIFAC interaction parameters would need to be fit along with the pH prediction, a problem that is substantially challenging. Therefore, a general extension of the model to ternary or higher systems is challenging and shall be addressed, at least in part, in future publications.

Additionally, the current model calculates solubility changes with pH based on the assumption that the compound that precipitates is the neutral amino acid. However, inducing a pH change requires the addition of an acid or a base, both of which, along with H^+ or OH^- ions, also contain counterions. Salts formed between a protonated amino acid and an anion, or a deprotonated amino acid and a cation may precipitate before the solubility limit for the amino acid itself is reached. For example, Glutamic Acid is an acidic amino acid that has its pI at acidic pH around 3, adding large amount of NaOH to the system would result in the precipitation of Na-Glutamic Acid complex near neutral pH.

Additionally, pharmaceutical media contain diverse components in addition to amino acids such as sugars, and salts with various ions. These components can interact with each other through multiple reactions, forming a large number of ionic and nonionic equilibria. This is the case for formation of ion-amino acid complexes, for which a large amount of data is already publicly available in aquatic chemistry databases. As an example, in a solution of Glutamic Acid, NaCl and $CuSO_4$ in

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

water, both Cu^{2+} and Na^{+} will complex with Glutamic Acid to form salts such as Na-Glutate and Cu-Glutate. The larger problem consisting of multiple salts, amino acids and pH change agents such as acids, bases and buffers would require consideration of all possible equilibria from reactions involving any subset of the species added to solution. This is an extension or generalization of the 2−3 equilibria included for binary solutions of each amino acid, and is a promising venue for future work. Predicting this requires considering reactions leading to the formation of salts and complexes, robust database with equilibria constants (association or dissociation constants) for the complexes formed. It also will require an algorithm to calculate multiple equilibria simultaneously.

CONCLUSIONS

Activity coefficients and solubilities for multiple amino acid solutions from the literature and new experimental data have been correlated with a model that combines short-range and long-range interactions. The short-range interactions are calculated using UNIFAC and the long-range ionic interactions are calculated using a species-independent extended Debye−Huckel equation by Rapp et al.[1] New UNIFAC functional groups were introduced for parts of the side chains in L-Histidine and L-Argnine. Parameters required for activity coefficient predictions in UNIFAC were fitted to all available data simultaneously and errors lower than 3% were obtained. This model was combined with a chemical equilibrium model similar to Visual-MINTEQ[2] to extend the applicability of the model to different pH values. A pH solubility workflow diagram was generated in order to make these calculations. The workflow, which was successful in modeling the pH dependence of amino acids in binary solutions, can in principle be extended to more complex systems involving multiple amino acids, salts, fatty acids, sugars, acids and bases that better resemble cell growth media.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.3c00365.

Binary paper (ZIP)

Experimental solubility data at various temperatures, presented in Table S1, along with comprehensive explanations of newly defined UNIFAC group contributions, and the MATLAB code employed for structural parameter determination (PDF)

AUTHOR INFORMATION

Corresponding Author

Marc Donohue − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*; Email: mdd@jhu.edu

Authors

Tianxin Zhang − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*; orcid.org/0000-0002-3115-6468

Ashwin Dravid − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Jayanth Reddy − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Lateef Aliyu − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Jaeyoung Park − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Zibo Wang − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Kai-Wen Huang − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Jinke Wu − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Shuman Liu − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Alan Stone − *Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*; orcid.org/0000-0002-3660-1129

Michael Betenbaugh − *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States*; orcid.org/0000-0002-1237-5550

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.3c00365

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the Johns Hopkins University Chemical and Biomolecular Engineering Department and AMBIC (Advanced Mammalian Biomanufacturing Innovation Center) for their support during this project.

REFERENCES

(1) Rapp, K. M.; Betenbaugh, M. J.; Donohue, M. D. A Semi-Empirical equation modeling Long-Range contributions to the activity coefficients of individual ions at high ionic strengths. *J. Mol. Liq.* **2024**, *407*, No. 125144.

(2) Gustafsson, J. P. *Visual MINTEQ*, Version 3.1 [Software], 2020. https://vminteq.com/.

(3) Ritacco, F. V.; Wu, Y.; Khetan, A. Cell culture media for recombinant protein expression in Chinese hamster ovary (CHO) cells: History, key components, and optimization strategies. *Biotechnol. Prog.* **2018**, *34*, 1407−1426.

(4) Jin, X. Z.; Chao, K. C. Solubility of four amino acids in water and of four pairs of amino acids in their water solutions. *J. Chem. Eng. Data* **1992**, *37*, 199−203.

(5) Chen, C.-C.; Zhu, Y.; Evans, L. B. Phase partitioning of biomolecules: solubilities of amino acids. *Biotechnol. Prog.* **1989**, *5*, 111−118.

(6) Dalton, J. B.; Schmidt, C. L. A., The Solubilities of Certain Amino Acids and Related Compounds in Water, the Densities of Their Solutions at Twenty-Five Degrees, and the Calculated Heats of Solution and Partial Molal volumes.II. *J. Biol. Chem.*, **1935**. *109*. 241−248, https://www.sciencedirect.com/science/article/pii/S0021925818752344, DOI: 10.1016/s0021-9258(18)75234-4.

(7) Tseng, H.-C.; Lee, C.-Y.; Weng, W.-L.; Shiah, I.-M. Solubilities of amino acids in water at various pH values under 298.15K. *Fluid Phase Equilib.* **2009**, *285*, 90−95.
(8) Daldrup, J.-B. G.; Held, C.; Sadowski, G.; Schembecker, G. Modeling pH and Solubilities in Aqueous Multisolute Amino Acid Solutions. *Ind. Eng. Chem. Res.* **2011**, *50*, 3503−3509.
(9) Pinho, S. P.; Silva, C. M.; Macedo, E. A. Solubility of Amino Acids: A Group-Contribution Model Involving Phase and Chemical Equilibria. *Ind. Eng. Chem. Res.* **1994**, *33*, 1341−1347.
(10) Kuramochi, H.; Noritomi, H.; Hoshino, D.; Nagahama, K. Measurements of Solubilities of Two Amino Acids in Water and Prediction by the UNIFAC Model. *Biotechnol. Prog.* **1996**, *12*, 371−379.
(11) Gupta, R. B.; Heidemann, R. A. Solubility models for amino acids and antibiotics: Solubility Models for Amino Acids and Antibiotics. *AIChE J.* **1990**, *36*, 333−341.
(12) Fuchs, D.; Fischer, J.; Tumakaka, F.; Sadowski, G. Solubility of Amino Acids: Influence of the pH value and the Addition of Alcoholic Cosolvents on Aqueous Solubility. *Ind. Eng. Chem. Res.* **2006**, *45*, 6578−6584.
(13) Xu, X.; Pinho, S. P.; Macedo, E. A. Activity Coefficient and Solubility of Amino Acids in Water by the Modified Wilson Model. *Ind. Eng. Chem. Res.* **2004**, *43*, 3200−3204.
(14) Held, C.; Cameretti, L. F.; Sadowski, G. Measuring and Modeling Activity Coefficients in Aqueous Amino-Acid Solutions. *Ind. Eng. Chem. Res.* **2011**, *50*, 131−141.
(15) Bonner, O. Osmotic and activity coefficients of some amino acids and their hydrochloride salts at 298.15 K: Semantic scholar. *J. Chem. Eng. Data* **1982**, *27*, 422−423.
(16) Kuramochi, H.; Noritomi, H.; Hoshino, D.; Nagahama, K. Representation of activity coefficients of fundamental biochemicals in water by the UNIFAC model. *Fluid Phase Equilib.* **1997**, *130*, 117−132.
(17) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, *21*, 1086−1099.
(18) Chua, Y. Z.; Do, H. T.; Schick, C.; Zaitsau, D.; Held, C. New experimental melting properties as access for predicting amino-acid solubility. *RSC Adv.* **2018**, *8*, 6365−6372.
(19) Do, H. T.; Chua, Y. Z.; Kumar, A.; Pabsch, D.; Hallermann, M.; Zaitsau, D.; Schick, C.; Held, C. Melting properties of amino acids and their solubility in water. *RSC Adv.* **2020**, *10*, 44205−44215.
(20) Larsen, B. L.; Rasmussen, P.; Fredenslund, A. A modified UNIFAC group-contribution model for prediction of phase equilibria and heats of mixing. *Ind. Eng. Chem. Res.* **1987**, *26*, 2274−2286.
(21) Lohmann, J.; Joh, R.; Gmehling, J. From UNIFAC to modified UNIFAC (Dortmund). *Ind. Eng. Chem. Res.* **2001**, *40*, 957−964.
(22) Cameretti, L. F.; Sadowski, G. Modeling of aqueous amino acid and polypeptide solutions with PC-SAFT. *Chem. Eng. Process.* **2008**, *47*, 1018−1025.
(23) 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 J. Chem. Eng.* **2010**, *27* (1), 266−267.
(24) Cao, Z.; Hu, Y.; Li, J.; Kai, Y.; Yang, W. Solubility of glycine in binary system of ethanol+water solvent mixtures: Experimental data and thermodynamic modeling. *Fluid Phase Equilib.* **2013**, *360*, 156−160.
(25) Gmehling, J.; Rasmussen, P.; Fredenslund, A. Vapor-liquid equilibriums by UNIFAC group contribution. Revision and extension. 2. *Ind. Eng. Chem. Process Des. Dev.* **1982**, *21*, 118−127.
(26) Hansen, H. K.; Rasmussen, P.; Fredenslund, A.; Schiller, M.; Gmehling, J. Vapor-liquid equilibria by UNIFAC group contribution. 5. Revision and extension. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352−2355.

https://doi.org/10.1021/acs.iecr.3c00365
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX