Solubility of Amino Acids: A Group-Contribution Model Involving Phase and Chemical Equilibria

Simão P. Pinho, Carlos M. Silva, and Eugénia A. Macedo*

Laboratory of Separation and Reaction Engineering, Faculdade de Engenharia, Rua dos Bragas, 4099 Porto Codex, Portugal

A new model is proposed to represent the solubility behavior of 14 amino acids and 5 small peptides in water. The UNIFAC model is combined with a Debye–Hückel term to describe the activity coefficients of the species present in the biomolecule/water system. New groups have been defined according to the group-contribution concept, and chemical equilibrium is taken into account simultaneously with the physical equilibrium. To estimate the new interaction parameters, molal activity coefficient data from the literature were used. These parameters, in addition to solubility data, were the basis for the correlation of the solubility product of the amino acids. Using this approach, satisfactory results were obtained in the representation and prediction of the solubilities of amino acids in aqueous solutions at different conditions of temperature and pH.

Introduction

Many valuable biochemicals are produced in reactors where the product concentration in a very complex mixture is very small. Therefore, the development of efficient methods for separation, concentration, and purification of biological products is of fundamental importance.

To design, optimize, and scale-up separation processes, the application of molecular thermodynamics is a useful tool. In the case of biotechnology, it is particularly important to focus the attention on the properties of aqueous systems containing salts and large, charged molecules (Prausnitz, 1989). Although amino acids are among the simplest biochemicals, they have many similarities with more complex biomolecules such as antibiotics (Orella and Kirwan, 1991) and the study of their solubility in water is a good starting point for the understanding of biochemical systems.

The successful representation of the solubilities is directly related to the ability of correlating and predicting the activity coefficients of the amino acids in solution. In this way, several attempts have been made in the last few years: Nass (1988) has assumed the activity coefficients of amino acids to be a product of two terms due to chemical reaction equilibria and physical interactions using for this the Wilson equation (Wilson, 1964), with Bondi's volume ratios (Bondi, 1968) as pure-component liquid volume ratios. Although the correlation results are satisfactory, they are limited to a few amino acids. The number of estimated parameters varies from 3 to 10. Chen et al. (1989) added two different contributions for the calculation of the excess Gibbs energy of the system: one is the result of the long-range interactions and was represented by a Pitzer–Debye–Hückel term (Pitzer, 1980); the other is due to local interactions and was formulated by a modified form of the NRTL equation (Renon and Prausnitz, 1968) with two adjustable energy parameters for each amino acid/water pair. As the data were correlated, each amino acid separately, the results were very satisfactory, both for correlation of activity coefficients (mean deviation of 2%) and of solubilities (mean deviation < 1.8%).

More recently, Gupta and Heidemann (1990) tried to describe the activity coefficients of amino acids in water considering only short-range interactions, using the modified UNIFAC model (Larsen et al., 1987). Their definition of groups is different from the usual, as they considered very large groups. The proline molecule for instance, was considered one group. In average, the results may be considered poor both for correlation and prediction.

In this work a new model for prediction of activity coefficients of amino acid/water and peptide/water mixtures has been developed. The model combines chemical and physical equilibrium simultaneously. The original UNIFAC group-contribution method (Fredenslund et al., 1975) is used for the description of physical equilibrium. New groups have been defined and new parameters were estimated. The influence of pH and of temperature on the solubility for several amino acids has been examined with satisfactory quantitative and qualitative results.

Model Development

When an amino acid or a peptide (AA) is present in an aqueous phase, some reactions take place:

$$\mathrm{AA(s)} \overset{k_s}{\rightleftharpoons} \mathrm{AA} \tag{1}$$

$$\mathrm{AA} \overset{K_D}{\rightleftharpoons} \mathrm{AA^{\pm}} \tag{2}$$

$$\mathrm{AA^{\pm}} \overset{K_1}{\rightleftharpoons} \mathrm{AA^-} + \mathrm{H^+} \tag{3}$$

$$\mathrm{AA^{\pm}} + \mathrm{H^+} \overset{K_2}{\rightleftharpoons} \mathrm{AA^+} \tag{4}$$

$$\mathrm{H_2O} \overset{K_w}{\rightleftharpoons} \mathrm{H^+} + \mathrm{OH^-} \tag{5}$$

Greenstein and Winitz (1961) reported values of K_D for amino acids and small peptides in the range 10^5–10^6. Thus, reactions 1 and 2 can be combined as follows:

$$\mathrm{AA(s)} \overset{K_s}{\rightleftharpoons} \mathrm{AA^{\pm}} \tag{6}$$

Reaction 6 indicates the dissolution of a biomolecule, forming a neutral dipolar species called a zwitterion, carrying dual electric charges. The participation of the zwitterion in acid–base reactions to form amino acid anionic or cationic species is shown by eqs 3 and 4. It is worthwhile to mention that some amino acids have more than one cationic species, such as arginine and lysine, or more than one anionic species, like tyrosine and aspartic

* Author to whom all correspondence should be addressed.

0888-5885/94/2633-1341$04.50/0

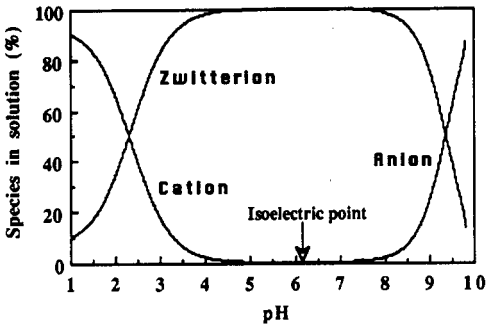

Figure 1. Percentage of the different glycine ionic species in aqueous solution at 298.15 K as a function of pH.

acid (Bohinski, 1987). However, in this study only molecules with two acid–base reactions were considered.

The equilibrium constants for these reactions are given as follows:

$$K_s = a_{AA^\pm} \tag{7}$$

$$K_1 = a_{AA^-}a_{H^+}/a_{AA^\pm} \tag{8}$$

$$K_2 = a_{AA^+}/a_{AA^\pm}a_{H^+} \tag{9}$$

$$K_w = a_{H^+}a_{OH^-} \tag{10}$$

where

$$a_i = m_i\gamma_i^* \tag{11}$$

γ_i^* is the molal unsymmetric activity coefficient, and m_i is the molality of species i. The equilibrium constant K_s is in this work considered to be an adjustable parameter. The small influence of the difference in standard states between reaction 1 (system pressure) and reaction 2 (standard pressure) is therefore incorporated into K_s. In eq 7 it is also assumed that we are in the presence of a pure solid.

As can be seen in Figure 1, in isoelectric solutions, at which the isoelectric point is defined by $pI = pH = (pK_1 + pK_2)/2$, neutral dipolar species are predominant, but for pH much higher than pI, the anionic amino acid species become predominant, while cationic species are dominant at values of pH much smaller than pI.

From the above equations it is evident that for the successful representation of solubilities and their dependence with pH and temperature it is necessary to be able to represent accurately the activity coefficients of the different species in the solution. However, this is not an easy task. Besides the powerful electric fields surrounding the charged groups that give rise to important long-range interactions and in a certain way make amino acids and peptides behave as strong electrolytes (Cohn et al., 1934), the influence of the hydrocarbon chain must also be taken into account.

To represent these interactions, a semiempirical model is proposed. The model is based on the assumption that the excess Gibbs energy of the system is a linear combination of two terms:

$$G^{*,E} = G^{*,E}_{UNIFAC} + G^{*,E}_{DH} \tag{12}$$

From eq 12 the molal scale unsymmetric activity coefficient for ionic species can be derived as

$$\ln \gamma_i^* = \ln \gamma_i^{*,C} + \ln \gamma_i^{*,R} + \ln \gamma_i^{*,DH} \tag{13}$$

The combinatorial ($\gamma_i^{*,C}$) and residual ($\gamma_i^{*,R}$) terms for short-range interactions are calculated using the original UNIFAC method, with the correction in the combinatorial term for very dilute solutions (Kikic et al., 1980). Groups already available on UNIFAC VLE parameter tables (Gmehling et al., 1982; Macedo et al., 1983; Tiegs et al., 1987; Hansen et al., 1991) have been used, and when necessary, new groups have been defined according to the group-contribution concept. As we used the unsymmetric convention and molality scale for the activity coefficients, some conversions had to be made. Thus, the calculation of activity coefficients at infinite dilution by means of UNIFAC was necessary to convert calculated values from mean rational symmetric (γ_i^x) to mean rational unsymmetric activity coefficients ($\gamma_i^{*,x}$):

$$\gamma_i^{*,x} = \gamma_i^x/\gamma^\infty \tag{14}$$

Conversion between mole fraction scale ($\gamma_i^{*,x}$) and molal scale activity coefficients (γ_i^*) was made using an expression derived by Robinson and Stokes (1965):

$$\gamma_i^{*,x} = \gamma_i^*(1 + 0.001M_A\sum m_i) \tag{15}$$

where M_A is the molecular weight of the solvent and the summation is to be made over all the solute species.

The Debye–Hückel approach presented in eq 13 as $\gamma_i^{*,DH}$ (Robinson and Stokes, 1965) was used to represent the long-range interaction forces. This term, expressed in eq 16, computes solution ionic strength (I) from the net charge of the species, and therefore, for a zwitterion this term is zero. In this case, no conversions were necessary.

$$\log \gamma_i^{*,DH} = -Az_i^2\sqrt{I}/(1 + Ba\sqrt{I}) \tag{16}$$

where

$$I = 0.5\sum m_i z_i^2 \tag{17}$$

$$A = 5.7664 \times 10^4\, ds^{0.5}/(\epsilon T)^{1.5} \tag{18}$$

$$B = 1.590 \times 10^{10}\, ds^{0.5}/(\epsilon T)^{1/2} \tag{19}$$

a is the Debye–Hückel distance of closest approach of ions, ϵ is the dielectric constant of the solvent, ds is the density of the solvent, T is the absolute temperature in Kelvin, and z is the charge number of the ion.

Correlation and Prediction of Activity Coefficients

Experimental and Physical Data. Hutchens (1976) compiled the molality scale unsymmetric activity coefficient data of amino acids and small peptides in pure water at 25 °C. The original measurements on osmotic coefficients and the subsequent conversion to activity coefficients were made by Smith and Smith (1937, 1940a,b), Hutchens et al. (1963), and Ellerton et al. (1964). Relevant information from the experimental data collected is shown in Table 1.

Dissociation reaction equilibrium constants K_1 and K_2 and their variation with the temperature have been measured by several authors. The data used in this work were found in the CRC Handbook of Chemistry and Physics (Izatt and Christensen, 1973) and King (1951).

New UNIFAC Groups. As discussed earlier (Figure 1) the zwitterionic species are predominant in pure water. Thus, we have defined new charged groups to describe accurately the components in the solution. A new group C=O, was also defined to represent peptide bonding. In

Table 1. Overall Representation of Experimental Data

substance	experimental data points	maximum molality	γ^* range
alanine	7	1.86	1.005–1.045
α-aminobutyric acid	7	2.00	1.011–1.165
α-aminovaleric acid	5	0.65	1.022–1.072
glycine	10	3.11	0.960–0.738
hydroxyproline	7	2.00	1.000–1.026
proline	15	7.30	1.019–2.004
serine	11	4.00	0.951–0.603
threonine	7	2.00	0.989–0.944
valine	3	0.50	1.030–1.076
alanylalanine	5	1.00	0.982–1.035
alanylglycine	5	1.00	0.931–0.855
glycylalanine	5	1.00	0.935–0.855
glycylglycine	6	1.50	0.911–0.689
triglycine	2	0.30	0.851–0.804

Table 2. New UNIFAC Group Size Parameters

group	subgroup	R_k	Q_k
COO^-	COO^-	1.3013	1.224
CNH_3^+	$CH_2NH_3^+$	1.3692	1.236
	$CHNH_3^+$	1.1417	0.924
CNH_2^+	$CH_2NH_2^+$	1.2070	0.936
CO	CO	0.7713	0.640

Table 3. Groups of the Amino Acids and Peptides on the Zwitterionic Form

substance	groups
alanine	CH_3, COO^-, $CHNH_3^+$
α-aminobutyric acid	CH_3, CH_2, COO^-, $CHNH_3^+$
α-aminovaleric acid	CH_3, $2CH_2$, COO^-, $CHNH_3^+$
glycine	COO^-, $CH_2NH_3^+$
hydroxyproline	CH_2, 2CH, OH, COO^-, $CH_2NH_2^+$
proline	$2CH_2$, CH, COO^-, $CH_2NH_2^+$
serine	CH_2, OH, COO^-, $CHNH_3^+$
threonine	CH_3, CH, OH, COO^-, $CHNH_3^+$
valine	$2CH_3$, CH_2, COO^-, $CHNH_3^+$
alanylalanine	$2CH_3$, CHNH, COO^-, $CHNH_3^+$, CO
alanylglycine	CH_3, CH_2NH, COO^-, $CHNH_3^+$, CO
glycylalanine	CH_3, CHNH, COO^-, $CH_2NH_3^+$, CO
glycylglycine	CH_2NH, COO^-, $CH_2NH_3^+$, CO
triglycine	$2CH_2NH$, COO^-, $CH_2NH_3^+$, 2 CO

the formulation of new groups, we used a similar strategy, as previously adopted for groups already available on UNIFAC tables, and tried to use those groups whenever possible. Table 2 summarizes the new groups and subgroups defined as well as the UNIFAC size parameters, which were determined using Bondi area and volume parameters (Bondi, 1968).

In the calculation of activity coefficients, solution chemistry was explicitly taken into account, since further studies on the influence of pH on solubility requires the knowledge of γ^* for the anionic and cationic species. Unfortunately, the available data do not include these conditions, and to represent local interactions, an assumption had to be made: that UNIFAC groups which constitute the anionic and cationic species are the same as those of the zwitterion. The relevant groups of all amino acids and peptides (their zwitterionic form) studied are shown in Table 3.

Parameter Estimation and Results. To estimate the model parameters, a modified Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963) was used to minimize the following objective function:

$$\mathrm{OBJ} = \sum_j (\gamma_j^{*,\mathrm{calc}} - \gamma_j^{*,\mathrm{exp}})^2 \qquad (20)$$

where γ_j^* refers to the zwitterion activity coefficient and calc and exp mean calculated with the model and experimental, respectively.

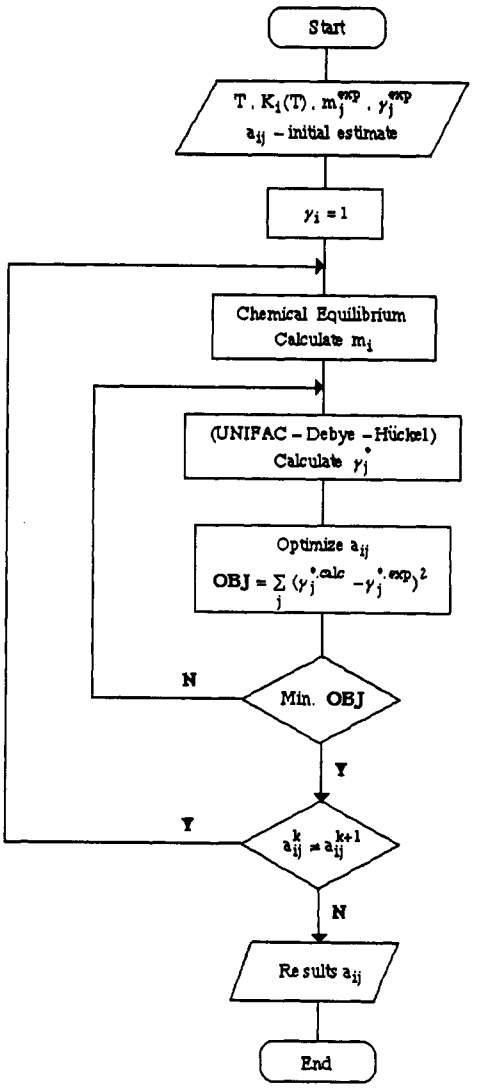

Figure 2. Algorithm used in the minimization process.

In the process minimization, the distance of closest approach on the Debye–Hückel theory was set equal to 4 Å, and only new UNIFAC interaction parameters were obtained. An algorithm of the method can be seen in Figure 2, where the different convergence criteria used are shown.

Seventeen data points available from alanine and glycine (Hutchens, 1976) were used to estimate four energy parameters between the groups COO^-, CNH_3^+, CH_2, and H_2O. The determination of the interaction parameters between the new above-mentioned groups and OH was carried out using data for serine (Hutchens, 1976). In this case, we concluded that the best results were obtained when the interaction parameters for OH/COO^- and OH/CNH_3^+ were set equal to the interaction parameters for COO^-/OH and CNH_3^+/OH, respectively. Data for proline and hydroxyproline were also regressed together to estimate the parameters OH/CNH_2^+, CNH_2^+/CH_2, H_2O/CNH_2^+, and CNH_2^+/H_2O. The two last ones were set equal during the calculations. Finally, the minimization process applied on the available experimental data for alanylalanine and glycylalanine allowed the estimation of the group parameters H_2O/CO, CNH/CNH_3^+, CNH/CO, CNH_3^+/CNH, and CO/H_2O.

All the other relevant parameters were fixed. They were all given the same value (5000), because it turned out during the minimization that this value gave the best representation of the activity coefficients of the systems studied.

Table 4. Group Interaction Parameters (K)

	CH_2	OH	H_2O	CNH	COO^-	CNH_3^+	CNH_2^+	CO
CH_2	0.0	986.5[a]	1318[a]	255.7[a]	5000	5000	5000	5000
OH	156.4[a]	0.0	353.5[a]	42.70[a]	-577.0	170.6	-572.4	na[b]
H_2O	300.0[a]	-229.1[a]	0.0	168.0[a]	-1354	803.5	-114.3	93.03
CNH	65.33[a]	-150.0[a]	-448.2[a]	0.0	5000	-335.9	na	-1142
COO^-	5000	-577.0	-568.9	5000	0.0	5000	5000	5000
CNH_3^+	5000	170.6	5000	-768.4	-2041	0.0	na	5000
CNH_2^+	-536.5	5000	-114.3	na	5000	na	0.0	na
CO	5000	na	-680.0	5000	5000	5000	na	0.0

[a] Values available on UNIFAC VLE Tables (Tiegs et al., 1987). [b] na: not available.

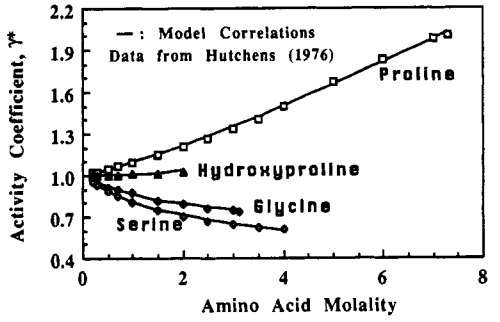

Figure 3. Experimental and calculated values for activity coefficients of amino acids in water at 298.15 K.

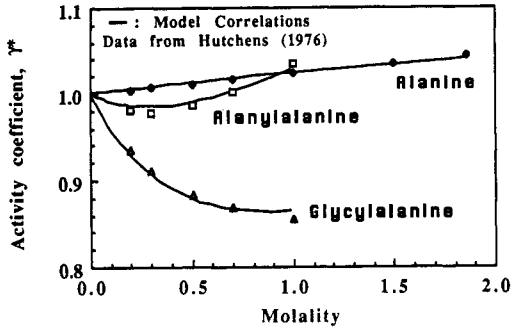

Figure 4. Experimental and calculated values for activity coefficients of alanine and some peptides in water at 298.15 K.

Table 4 presents the estimated parameters. In Figures 3 and 4 it is possible to observe the quality of the correlations. They are all of good quality. Table 5 summarizes some of the information of both correlation and prediction, and a comparison is given with the results of Chen et al. (1989) and Gupta and Heidemann (1990). The root mean square deviation (rmsd) values for correlation are quite satisfactory, and in the majority of the cases the deviations are even smaller than those calculated by Chen et al. (1989), who only correlated one amino acid or one peptide at a time. The results of Gupta and Heidemann (1990) seem to deviate more than our results even though both works are based on the UNIFAC group-contribution method and have the same number of parameters regressed.

Table 5 also shows the infinite dilution activity coefficients that arise from the minimization procedure. Unfortunately, we did not find any experimental data to compare with the estimated values; however, they seem to be quite reasonable.

The prediction results require some attention: the rmsd values calculated are much higher than the ones obtained for correlation. Nevertheless, they are about the same order of magnitude as the values obtained by Gupta and Heidemann (1990). For the prediction of the activity coefficients of valine, aminobutyric acid, and aminovaleric acid an incorrect slope was calculated for all the cases, with γ^* decreasing with increasing molality, while the data

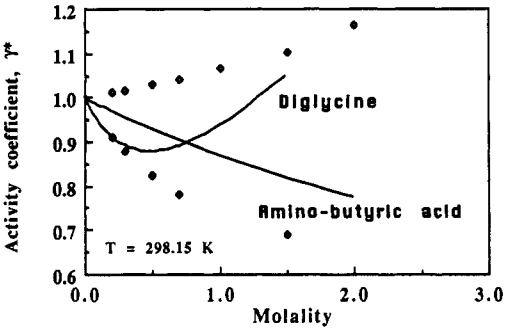

Figure 5. Comparison of the model predictions (—) with the experimental activity coefficients: ◇, aminobutyric acid; ●, diglycine. Experimental data: Hutchens (1976).

show the opposite trend. Figure 5 shows, for aminobutyric acid and glycylglycine, the predicted curves of activity coefficients.

Interactions between methyl groups and the new charged groups are not important, since all parameters have high values. The three amino acids already mentioned are obtained from alanine by addition of methyl groups (see Table 3), and the introduction of CH_2/H_2O interaction parameters is very important, suggesting the study of the effects of the addition of methyl groups on the activity coefficients of hypothetical mixtures of *n*-alkanes and water. Surprisingly, a deep decrease of γ^* with molality is observed. In some way, these results tend to show that the interaction parameters between CH_2 and H_2O (Tiegs et al., 1987) are not the best to represent the behavior of solutions of this kind. Moreover, difficulties arise in getting good correlations using amino acids with a long hydrocarbon chain, which support our point. However, as we want to maintain the group-contribution concept, no reevaluation of that pair of parameters was tried for our specific case. The same observations are valid for the other predictions.

Temperature Dependence of the Solubilities

In this work the solubility data is represented by regressing the thermodynamic solubility constant of eq 7 in the form:

$$\ln K_s = a + \frac{b}{T} + c \ln T \qquad (21)$$

where a, b, and c are parameters.

Hutchens (1976) compiled the solubilities of amino acids between 0 and 100 °C. However, original experimental data from Dalton and Schmidt (1933) and Dunn et al. (1933) cover only the temperature range between 0 and 75 °C. The known values of solubility above this temperature were obtained by extrapolation of the correlated curve.

To regress the coefficients of eq 21, only experimental and interpolated values of solubilities between 0 and 75 °C were used. The objective function used for this purpose

Table 5. Relevant Information from Minimization and Comparisons between Different Approaches

substance	γ^{∞}_{pred} 298.15 K	no. regressed parameters	rmsd (%) Chen et al. (1989)	Gupta and Heidemann (1990)	this work
alanine	0.85	4	0.04	8.97	0.19
glycine	1.67	4	2.07	4.20	0.60
hydroxyproline	0.02	3	0.36	0.06	0.39
proline	0.03	3	1.30	3.01	1.21
serine	2.20	2	2.80	3.32	0.24
alanylalanine	0.85	5	2.04		0.52
glycylalanine	0.88	5	2.44		0.59
α-aminobutyric acid	2.74		0.32	17.34	17.87
α-aminovaleric acid	10.19				16.40
threonine	5.68		0.70	13.78	26.12
valine	13.48		4.47	12.15	16.43
alanylglycine	0.19		2.92		28.22
glycylglycine	0.41		3.25		17.06
triglycine	0.06		0.67		10.00

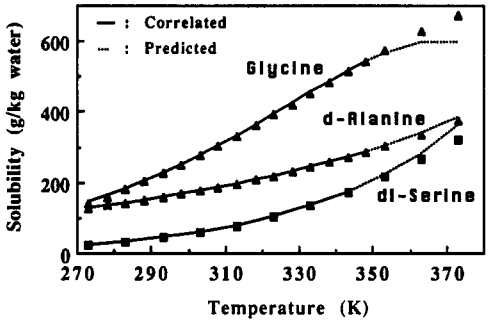

Figure 6. Amino acid solubility in water: correlation and prediction curves. Comparison with experimental data from Dalton and Schmidt (1933) and the extrapolated values from Hutchens (1976).

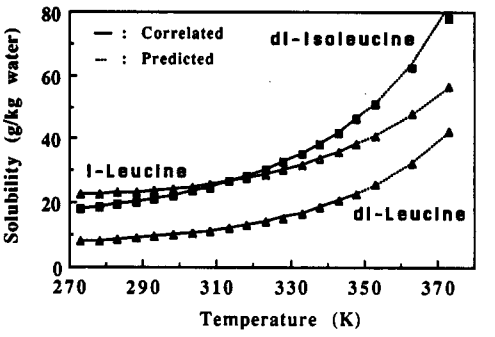

Figure 7. Amino acid solubility in water: correlation and prediction curves. Comparison with experimental data from Dalton and Schmidt (1933) and the extrapolated values from Hutchens (1976).

Table 6. Solubility Constants of Amino Acids in Water

substance	maximum molality	a	b	c	rmsd (%)
d-alanine	3.20	−25.32	1006	3.968	0.35
dl-alanine	3.58	−13.60	318.2	2.318	0.40
glycine	7.25	114.8	−5558	−16.72	1.62
l-hydroxyproline	4.12	−43.90	2520	6.410	0.44
l-isoleucine	0.29	−44.31	1663	6.531	0.01
dl-isoleucine	0.35	−121.7	4709	18.26	0.01
l-leucine	0.29	−117.5	4886	17.42	0.03
dl-leucine	0.17	−170.8	6687	25.57	0.02
dl-norleucine	0.22	−135.6	4934	20.47	0.01
l-proline	21.79	−182.3	9771	26.91	1.68
l-serine	5.63	664.4	−30250	−98.65	8.90
dl-serine	1.63	118.7	−7265	−16.71	0.39
l-valine	0.54	−55.99	2287	8.319	0.02
dl-valine	1.00	−62.73	2486	9.412	0.26

was

$$F_{min} = \sum_j (S_j^{calc} - S_j^{exp})^2 \qquad (22)$$

with S being the solubility in g/kg of water.

The fitted parameters are listed in Table 6, and the quality of the regression for some amino acids can be seen in Figures 6 and 7. These figures also show the predicted solubilities with our model at temperatures above 75 °C and the extrapolated solubilities from the experimental data given by Hutchens (1976). Some results related with those predictions are shown in Table 7. It is possible to verify big discrepancies between predicted and extrapolated values for glycine, *dl*-serine, and *dl*-valine at 100 °C, but on average the errors are not large (<5.9%). For glycine the model predicts a maximum solubility (Figure 6) while extrapolated values show linear increase in the solubility up to 100 °C. It would be nice to see what is the experimental curve in that zone.

Table 7. Comparison between the Extrapolated (S^{ext}) and Predicted Solubilities (S^{calc}) in g/kg of Water at 373.15 K

amino acid	S^{ext}	S^{calc}	error[a] (%)
d-alanine	373.0	383.8	−2.92
dl-alanine	440.4	454.9	−3.29
glycine	671.7	597.9	10.99
l-hydroxyproline	706.9	722.1	−2.15
dl-isoleucine	78.02	81.93	−5.01
l-leucine	56.38	55.83	0.97
dl-leucine	42.06	41.73	0.78
dl-norleucine	52.29	53.46	−2.24
l-proline	3355	3409	−1.61
dl-serine	322.4	363.0	−12.59
dl-valine	188.1	227.2	−20.79

[a] Error (%) = $100(S^{ext} - S^{calc})/S^{ext}$.

Although the results are good, some aspects deserve attention: (i) Extrapolations had to be made both on temperature and composition. The temperature range is 0–100 °C, while the UNIFAC parameters were estimated from experimental data at 25 °C only. As regards composition, for almost all amino acids, maximum molality is now much higher than the maximum molality of experimental data on activity coefficients (Table 1). (ii) The model was used indifferently to *d*, *l*, and *dl* forms of amino acids, since UNIFAC does not differentiate optical isomers.

Finally, it is important to mention that the reaction equilibrium constants of eqs 8–10 were considered functions of temperature (King, 1951; Robinson and Stokes, 1965; Izatt and Christensen, 1973).

pH Influence on the Solubility of Amino Acids

Studies of solubility dependence on pH are intimately related with the consideration of equilibrium reactions 3–6 in the solution. Thus, the activity coefficient model

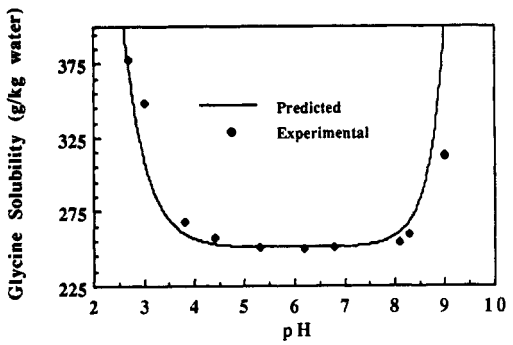

Figure 8. Experimental and calculated values for the solubility of glycine at 298.15 K. Experimental data: Needham et al. (1971).

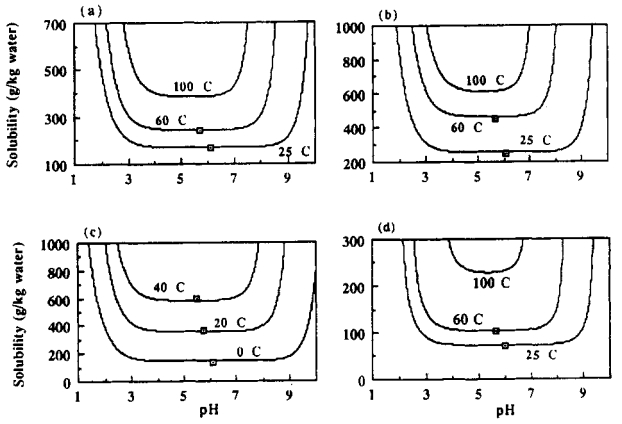

Figure 9. Predicted solubility of different amino acids as function of pH and temperature: (a) *d*-alanine; (b) glycine; (c) *l*-serine; (d) *dl*-valine.

and the results from solubility correlation were used to estimate solubilities at different conditions of pH.

In Figure 8 a comparison between solubilities of glycine at 25 °C at different pH values predicted by the model and from experimental data (Needham et al., 1971) is given. The quality of the predictions shown are better than expected: In the experimental work it was necessary to add a strong electrolyte (acid or basic) to fix the pH at the required value, and of course the electrolyte gives rise to new important interactions between the molecules, as can be seen in the works of Schrier and Robinson (1974) and Briggs et al. (1974).

Figure 9 shows predicted curves of solubility as a function of pH at different temperatures for some amino acids. They are similar to the experimental curves obtained by Dalton et al. (1930) for diiodotyrosine at three temperatures, showing the same type of behavior as we change from one temperature to the other. It is also possible to identify the predicted minimum solubility at the isoelectric point with the invariant solubility bands on both sides of it. The predicted bands at 25 °C are inside the range of 2–3 pH units on either side of the isoelectric point, as indicated in the works of Needham et al. (1971) and Zumstein and Rousseau (1989).

Conclusions

A new model combining chemical equilibria with a UNIFAC–Debye–Hückel approach to describe physical equilibria has been developed for the correlation and prediction of activity coefficients. New charged groups have been defined, taking into account the charges in the zwitterionic, the anionic, and the cationic species of amino acids. The results for correlation are in a very good agreement with experimental data, while for predictions the model must be used with caution, since the average rmsd is large.

The model provides good results for the correlation of the solubility, and a comparison between calculated solubilities with the model and the extrapolated values from the experimental curve shows good agreement.

Since studies of pH influence on the solubility are extremely dependent on the correlation of solubilities, care must be taken at temperatures higher than 75 °C. However, estimated bands of constant solubility at different temperatures seem to be very reasonable for all amino acids studied.

Finally, we want to stress a point already emphasized in this work: the available data are scarce and old, which may introduce some doubts in the obtained results. Therefore, experimental work to verify and improve the results is most welcome.

Acknowledgment

The authors are very grateful to Professor Peter Rasmussen (Technical University of Denmark, Lyngby, Denmark) for many helpful discussions. We recognize with thanks the financial support of JNICT (Portugal), for a research grant to S.P.P.

Nomenclature

A = solvent
A = Debye–Hückel parameter ($mol^{-0.5}$ $kg^{0.5}$)
AA = amino acid or peptide
AA^+ = amino acid or peptide cation
AA^- = amino acid or peptide anion
$AA^\pm$ = zwitterion
a = solubility constant in eq 21
a = activity; UNIFAC binary interaction parameter (K)
a = Debye–Hückel distance of closest approach of ions (m)
B = Debye–Hückel parameter ($mol^{-0.5}$ $kg^{0.5}$ m^{-1})
b = solubility constant in eq 21, K
c = solubility constant in eq 21
F_{min} = objective function (eq 22)
$G^{*,E}$ = unsymmetric excess Gibbs energy
I = ionic strength (mol/kg of solvent)
K = thermodynamic chemical equilibrium constant
K_D = ratio of zwitterion to uncharged forms of amino acids
K_w = ionic product of water
K_s = thermodynamic solubility constant for zwitterion
k_s = thermodynamic solubility constant for amino acid
M = molecular weight (g/mol)
m = molality (mol/kg of solvent)
OBJ = objective function (eq 20)
Q = group area parameter
R = group volume parameter
S = solubility (g/kg of solvent)
s = solid phase
T = absolute temperature (K)
z = charge number of the ion

Greek Letters

γ = activity coefficient
ϵ = dielectric constant

Subscripts

DH = long-range contribution represented by Debye–Hückel equation
i = component i; group i; reaction i
j = data point j; group j
k = group k
UNIFAC = local contribution by UNIFAC model

Superscripts

C = combinatorial term from UNIFAC
calc = calculated
DH = long-range contribution by Debye–Hückel equation
exp = experimental
ext = extrapolated from experimental curve
k = iteration
R = residual term from UNIFAC
x = mean rational scale
* = unsymmetric convention
∞ = infinite dilution

Literature Cited

Bohinski, R. C. *Modern Concepts in Biochemistry*, 5th ed.; Allyn and Bacon: Newton, MA, 1987; Chapter 3.

Bondi, A. *Physical Properties of Molecular Crystals, Liquids and Glasses*; Wiley: New York, 1968; Chapter 14.

Briggs, C. C.; Lilley, T. H.; Rutherford, J.; Woodhead, S. The Activity of Calcium Chloride in Aqueous Solutions of Some Amino Acids at 25 °C. *J. Solution Chem.* **1974**, *3*, 649–658.

Chen, C. C.; Zhu, Y.; Evans, L. B. Phase Partioning of Biomolecules: Solubilities of Amino Acids. *Biotechnol. Prog.* **1989**, *5*, 111–118.

Cohn, E. J.; McMeekin, T. L.; Edsall, T. J.; Weare, J. H. Studies in the Physical Chemistry of Amino Acids, Peptides and Related Substances. II. The solubility of α-Amino Acids in Water and in Alcohol-Water Mixtures. *J. Am. Chem. Soc.* **1934**, *56*, 2270–2282.

Dalton, J. B.; Schmidt, C. L. A. The Solubilities of Certain Amino Acids in Water, the Densities of their Solutions at Twenty Five Degrees, and the Calculated Heats of Solution and the Partial Molal Volumes. *J. Biol. Chem.* **1933**, *103*, 549–578.

Dalton, J. B.; Kirk, P. L.; Schmidt, C. L. A. The Apparent Dissociation Constants of Diiodotyrosine, its Heat of Solution, and its Apparent Heat of Ionization. *J. Biol. Chem.* **1930**, *88*, 589.

Dunn, M. S.; Ross, F. J.; Read, L. S. The Solubility of the Amino Acids in Water. *J. Biol. Chem.* **1933**, *103*, 579–595.

Ellerton, H. D.; Reinfelds, G.; Mulcahy, D. E.; Dunlop, P. J. Activity, Density, and Relative Viscosity Data for Several Amino Acids, Lactamide, and Raffinose in Aqueous Solution at 25 °C. *J. Phys. Chem.* **1964**, *68*, 398–402.

Fredenslund, Aa.; Jones, R. L.; Prausnitz, J. M. Group-Contribution Estimation of Activity Coefficients in Non Ideal Liquid Mixtures. *AIChE J.* **1975**, *21*, 1086–1099.

Gmehling, J.; Rasmussen, P.; Fredenslund, Aa. Vapor-Liquid Equilibria by UNIFAC Group Contribution. Revision and Extension 2. *Ind. Eng. Chem. Process Des. Dev.* **1982**, *21*, 118–127.

Greenstein, J. P.; Winitz, M. *Chemistry of the Amino Acids*; John Wiley & Sons: New York, 1961; Vol. 1, pp 447.

Gupta, R. B.; Heidemann, R. A. Solubility Models for Amino Acids and Antibiotics. *AIChE J.* **1990**, *36* (3), 333–341.

Hansen, H. K.; Rasmussen, P.; Fredenslund, Aa.; Schiller, M.; Gmehling, J. Vapor–Liquid Equilibria by UNIFAC Group Contribution. Revision and Extension 5. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352–2355.

Hutchens, J. O. In *Handbook of Biochemistry and Molecular Biology*, 3rd ed.; Fasman, G. D., Ed.; CRC Press: Cleveland, OH, 1976; Vol. 1, pp 104–125.

Hutchens, J. O.; Figlio, K. M.; Granito, S. M. An Isopiestic Comparison Method for Activities. *J. Biol. Chem.* **1963**, *238*, 1419–1422.

Izatt, R. M.; Christensen, J. J. In *CRC Handbook of Chemistry and Physics*, 53rd ed.; Weast, R. C., Ed.; The Chemical Rubber Co.: Cleveland, OH, 1973; Chapter J.

Kikic, I.; Alessi, P.; Rasmussen, P.; Fredenslund, Aa. On the Combinatorial Part of the UNIFAC and UNIQUAC Models. *Can. J. Chem. Eng.* **1980**, *58*, 253–258.

King, E. J. The Ionization Constants of Glycine and the Effect of Sodium Chloride Upon its Second Ionization. *J. Am. Chem. Soc.* **1951**, *73*, 155–159.

Larsen, B. L.; Rasmussen, P.; Fredenslund, Aa. A Modified UNIFAC Group-Contribution Model for Prediction of Phase Equilibria and Heat of Mixing. *Ind. Eng. Chem. Res.* **1987**, *26*, 2274–2286.

Levenberg, K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Q. Appl. Math.* **1944**, *2*, 164–168.

Macedo, E. A.; Weidlich, U.; Gmehling, J.; Rasmussen, P. Vapor–Liquid Equilibria by UNIFAC Group Contribution. Revision and Extension 3. *Ind. Eng. Chem. Process Des. Dev.* **1983**, *22*, 676–678.

Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431.

Nass, K. K. Representation of the Solubility Behavior of Amino Acids in Water. *AIChE J.* **1988**, *34*, 1257–1266.

Needham, T. E.; Paruta, A. N.; Gerraughty, R. J. Solubility of Amino Acids in Pure Solvent Systems. *J. Pharm. Sci.* **1971**, *60*, 565–567.

Orella, C. J.; Kirwan, D. J. Correlation of Amino Acid Solubilities in Aqueous Aliphatic Alcohol Solutions. *Ind. Eng. Chem. Res.* **1991**, *30*, 1040–1045.

Pitzer, K. S. Electrolytes. From Dilute Solutions to Fused Salts. *J. Am. Chem. Soc.* **1980**, *102*, 2902–2906.

Prausnitz, J. M. Biotechnology: A New Frontier for Molecular Thermodynamics. *Fluid Phase Equilib.* **1989**, *53*, 439–451.

Renon, H.; Prausnitz, J. M. Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures. *AIChE J.* **1968**, *14*, 135–144.

Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*, revd ed.; Butterworths: London, 1965; Chapter 2.

Schrier, E. E.; Robinson, R. A. Free Energy Relationships in Aqueous Amino Acid and Peptide Solutions Containing Sodium Chloride. *J. Solution Chem.* **1974**, *3*, 493–501.

Smith, P. K.; Smith, E. R. B. Thermodynamic Properties of Solutions of Amino Acids and Related Substances. II. The Activity of Aliphatic Amino Acids in Aqueous Solution at Twenty Five Degrees. *J. Biol. Chem.* **1937**, *121*, 607–613.

Smith, P. K.; Smith, E. R. B. Thermodynamic Properties of Solutions of Amino Acids and Related Substances. V. The Activities of Some Hydroxy- and *n*-Methylamino Acids and Proline in Aqueous Solution at Twenty Five Degrees. *J. Biol. Chem.* **1940a**, *132*, 57–64.

Smith, P. K.; Smith, E. R. B. Thermodynamic Properties of Solutions of Amino Acids and Related Substances. VI. The Activities of Some Peptides in Aqueous Solution at Twenty Five Degrees. *J. Biol. Chem.* **1940b**, *135*, 273.

Tiegs, D.; Gmehling, J.; Rasmussen, P.; Fredenslund, Aa. Vapor–Liquid Equilibria by UNIFAC Group Contribution. Revision and Extension 4. *Ind. Eng. Chem. Res.* **1987**, *26*, 159–161.

Wilson, G. M. Vapor–Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing. *J. Am. Chem. Soc.* **1964**, *86*, 127–130.

Zumstein, R. C.; Rousseau, R. W. Solubility of *l*-Isoleucine in and Recovery of *l*-Isoleucine from Neutral and Acidic Aqueous Solutions. *Ind. Eng. Chem. Res.* **1989**, *28*, 1226–1231.

Received for review September 20, 1993
Revised manuscript received January 21, 1994
Accepted February 1, 1994[⊕]

[⊕] Abstract published in *Advance ACS Abstracts,* March 15, 1994.