Representation of the Solubility Behavior of Amino Acids in Water

Expressions suitable for the description of the behavior of small biological zwitterions in aqueous systems are developed. The validity and usefulness of the expressions are tested by correlating data for the activity coefficients of several amino acids in water. Development of rational expressions for the solid-liquid equilibria (SLE) of amino acids, in terms of the heat of solution at infinite dilution, is also presented. The solubility data of phenylalanine, tyrosine and diiodotyrosine as a function of pH are successfully correlated using the approach developed here. In addition, simultaneous correlation of the pH and temperature dependence of solubility data for diiodotyrosine in water gives very good representation. The expressions described here should prove useful for the design and optimization of separation processes, such as crystallization, with minimal experimental information.

Kathryn K. Nass
Life Sciences Group
Research Laboratories
Eastman Kodak Company
Rochester, NY 14650

Introduction

Design and scale-up of reliable unit operations for the biotechnology industry should begin with a thorough understanding of the complex mixtures generated by living systems. Species as diverse in size and molecular behavior as enzymes, nucleic acids, metabolites, electrolytes and water exist together in cellular extracts and fermentation broths; modeling their interactions and phase equilibria behavior presents a unique challenge. Standard thermodynamic approaches to mixtures could provide new ways to model these solutions and aid in refinement of current recovery and purification techniques. However, little progress has been made to date on such applications.

Recent interest in the solution behavior of aqueous electrolyte systems has led to the development of several models for the excess Gibbs energy (G^E) and electrolyte activity coefficients (γ_i), including those of Pitzer (1973), Cruz and Renon (1978), Chen et al. (1982), Ball et al. (1985), Chen (1985), and most recently Chen and Evans (1986). Applications of these models have generally been restricted to salt/water mixtures, however, and little attention has been given to the types of electrolyte systems encountered in biochemical processes (e.g., systems for which the extent of dissociation of a weak electrolyte varies with changes in pH). Earlier work by Cohn and Edsall (1965) and Greenberg (1951), among others, provides the framework for describing the behavior of biochemical compounds as a function of pH, but gives no representation of physical interactions and no firm connection to phase equilibria.

This paper outlines the development of expressions for describing the behavior of small biological ampholytes in aqueous systems. The simplicity of this physically based approach makes these expressions particularly suitable for industrial applications. We assume that all interactions may be classified as either chemical or physical, and we express G^E (and thus ln γ_i) as the sum of two separate terms. The equilibrium equations provided by chemical theory define the chemical contribution and allow the introduction of pH through "reactions" of the solute with hydrogen ions. Variations in the local environment of the solute are accounted for by a form of the Wilson equation (Wilson, 1964); other contributions, such as those from long-range electrostatic interactions, can be represented by the addition of other physically based terms, as described at the end of this paper.

We examine the utility of our model by first correlating activity coefficient data for several simple amino acids. In addition, we develop rational expressions for the solid-liquid equilibria of amino acids in terms of the heat of solution at infinite dilution, rather than the usual heat of fusion; such expressions are more appropriate for amino acids, which decompose upon melting. Solution of the SLE equation using the activity coefficient expressions provides a method for correlating amino acid solubility data as a function of pH and temperature.

Model Development

We begin by assuming that G^E (or equivalently, the dimensionless excess Gibbs energy, G^E/RT) is a linear combination of two independent terms, one arising from chemical ("specific")

interactions and the second from physical ("nonspecific") interactions. The concept of expressing G^E/RT as a linear sum of terms has been applied successfully to modeling alcohol/hydrocarbon systems by many investigators, including Redlich and Kister (1947), Kretschmer and Wiebe (1954), Renon and Prausnitz (1967), and more recently by Nagata (1985). The success of this approach for systems, where one component associates, suggests that the same approach might be useful for systems where only one component participates in chemical interactions. A parametric study on the validity of the linear model for several different kinds of binary systems is described by Nass (1986). Although the assumption of linearity may be disputed for the systems considered in this paper, the excellent representation of data that we obtain with the linear model supports its use.

The partial molar property relationship between G^E/RT and the logarithm of the activity coefficient of species i ($\ln \gamma_i$) allows us to write

$$\ln \gamma_i = \ln \gamma_i \text{ (chemical)} + \ln \gamma_i \text{ (physical)}$$

The chemical term is modeled by chemical theory; physical interactions are represented by one form of the Wilson equation. Thus,

$$\ln \gamma_i = \ln \gamma_i \text{ (chem)} + \ln \gamma_i \text{ (Wilson)}$$

Chemical Theory

A general development of the equations associated with chemical theory and their relationship to liquid-phase activity coefficients has been given by Nass (1986). The resulting expressions for the reaction equilibria, the true mole fractions and the apparent activity coefficients appear in Figure 1. The flowsheet illustrates the iterative procedure for simultaneous

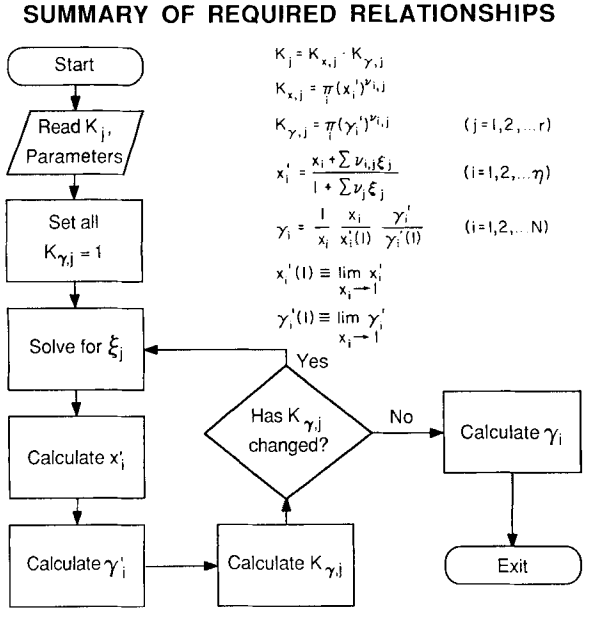


Figure 1. Overall scheme to calculate γ_i from chemical theory.

For r reactions, η true species, N apparent species, Lewis-Randall standard states.
Taken from Nass (1986).

solution of the equations. Note that these equations are of a general form; here, we consider the case where the $K_{\gamma,j}$ are assumed to be unity. Such an assumption follows directly from the assumption of separate and independent chemical and physical terms in the linear model, i.e., representation of physical interactions appears *only* in the physical term.

The use of the equations in Figure 1 is illustrated by the following special example, which relates the apparent activity coefficient of an uncharged amino acid (A) in aqueous solution to pH through three assumed reactions:

$$\text{I.} \quad A \rightarrow A^{\pm} \tag{1a}$$

$$\text{II.} \quad A^{\pm} \rightarrow A^- + H^+ \tag{1b}$$

$$\text{III.} \quad A^{\pm} + H^+ \rightarrow A^+ \tag{1c}$$

Here $A^{\pm}$ represents the zwitterionic form of the amino acid, and A^+ and A^- are the positive and negative forms, respectively. Note that the composition of the true solution (reactants and products) depends on the hydrogen ion (H^+) concentration, or pH. The equilibrium equations for the *ideal* true solution are given by

$$K_{\text{I}} = x'_{A^{\pm}}/x'_A = x'_3/x'_2 \tag{2a}$$

$$K_{\text{II}} = x'_{A^-}\, x'_{H^+}/x'_{A^{\pm}} = x'_4\, x'_6/x'_3 \tag{2b}$$

$$K_{\text{III}} = x'_{A^+}/x'_{A^{\pm}}\, x'_{H^+} = x'_5/x'_3\, x'_6 \tag{2c}$$

where the x'_i are the true mole fractions of the species and we have switched to numerical subscripts for convenience. The assumption of ideality for such a solution may be disputed; we prefer, then, to regard the *ratios* of true activity coefficients ($K_{\gamma,j}$) as unity. This assumption simplifies computation and results in a reasonable number of adjustable parameters.

The true mole fractions (x'_i) are related to the apparent mole fractions (x_i) through mass balance equations, written as

$$x'_1 = \frac{x_1}{1 + \xi_{\text{II}} - \xi_{\text{III}}} \tag{3a}$$

$$x'_2 = \frac{x_2 - \xi_{\text{I}}}{1 + \xi_{\text{II}} - \xi_{\text{III}}} \tag{3b}$$

$$x'_3 = \frac{\xi_{\text{I}} - \xi_{\text{II}} - \xi_{\text{III}}}{1 + \xi_{\text{II}} - \xi_{\text{III}}} \tag{3c}$$

$$x'_4 = \frac{\xi_{\text{II}}}{1 + \xi_{\text{II}} - \xi_{\text{III}}} \tag{3d}$$

$$x'_5 = \frac{\xi_{\text{III}}}{1 + \xi_{\text{II}} - \xi_{\text{III}}} \tag{3e}$$

$$x'_6 = \frac{\xi_{\text{II}} - \xi_{\text{III}}}{1 + \xi_{\text{II}} - \xi_{\text{III}}} \tag{3f}$$

The ξ_j are the dimensionless extents of reaction. The value of x'_6 is fixed by the pH of the solution, and therefore it does not

explicitly appear as part of the calculations. Simultaneous solution of Eqs. 2 and 3 yields values of the ξ_j, and thus the x'_i.

The apparent activity coefficient γ_2 of A may be related to the true quantities through the chemical theory formalism described by Prigogine and DeFay (1954). Here, we derive a general expression that allows for the choice of different standard states; we begin with the equality of the chemical potentials of the apparent and true species i,

$$\mu_i = \mu'_i$$

or written equivalently in terms of fugacities,

$$\hat{f}_i = \hat{f}'_i$$

or

$$x_i \gamma_i f_i^o = x'_i \gamma'_i (f_i^o)' \tag{4}$$

We are now free to *define* f_i^o so that the criterion for the chosen standard state is automatically satisfied. For example, if we specify a Lewis-Randall standard state, we require that

$$\lim_{x_i \to 1} \gamma_i = 1$$

Therefore, we define f_i^o as

$$f_i^o = x'_i(1)\, \gamma'_i(1)\, (f_i^o)'$$

where

$$x'_i(1) \equiv \lim_{x_i \to 1} x'_i$$

$$\gamma'_i(1) \equiv \lim_{x_i \to 1} \gamma'_i$$

This definition of f_i^o reflects the fact that an apparently pure fluid, e.g., acetic acid, may actually be a mixture of true species (monomers, dimers, etc.) formed by association. Substitution of this expression for f_i^o into Eq. 4 yields the working relationship between the apparent and true activity coefficients as

$$\gamma_i(\mathrm{LR}) = \frac{1}{x_i}\left(\frac{x'_i}{x'_i(1)}\right)\left(\frac{\gamma'_i}{\gamma'_i(1)}\right) \tag{5}$$

This expression also appears in Figure 1.

A similar expression results from the choice of an infinite-dilution (Henry's law) standard state; in this case, we require that

$$\lim_{x_i \to 0} \gamma_i = 1$$

It follows that we define f_i^o as

$$f_i^o = \gamma'_i(0)\, (f_i^o)' \tag{6}$$

where

$$\gamma'_i(0) \equiv \lim_{x_i \to 0} \gamma'_i$$

Combining this expression with Eq. 4 gives

$$\gamma_i(\mathrm{HL}) = \frac{x'_i}{x_i}\left[\frac{\gamma'_i}{\gamma'_i(0)}\right] \tag{7}$$

We now apply Eq. 7 to the example described here. The ratio of the true activity coefficient of the uncharged amino acid to the true activity coefficient at infinite dilution is assumed to be approximately unity; our final relationship then becomes

$$\gamma_2(\mathrm{chem}) = x'_2/x_2 \tag{8}$$

where we have again denoted the uncharged amino acid as species 2. Equation 8 provides one possible expression for the chemical contribution to $\ln \gamma_i$.

The physical contributions to the solution nonideality are represented by one version of the Wilson equation, written here for the binary mixture of water(1)/amino acid(2) as:

$$\ln \gamma_2(\mathrm{Wil}) = -\ln(x_2 + \Lambda_{21}x_1) - x_1\left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12}x_2} - \frac{\Lambda_{21}}{\Lambda_{21}x_1 + x_2}\right] \tag{9}$$

where

$$\Lambda_{ij} = (V_j/V_i)\exp\{-(\lambda_{ij} - \lambda_{ii})/RT\} \tag{10}$$

Here again we stress that the assumption of a linear model for G^E/RT implies that physical and chemical interactions are independent and separable. The Wilson equation, therefore, represents physical interactions between the uncharged amino acid and water only.

The pure-component liquid molar volumes V_i have no meaning for amino acids; instead, we set the ratio V_j/V_i equal to the ratio of the Bondi volumes (Bondi, 1968) of species i and j. This interpretation preserves the original meaning of the ratio, i.e., the ratio of the number of sites covered by species i and j on an imaginary lattice. The differences in the dimensionless interaction energy parameters, $(\lambda_{12} - \lambda_{11})/RT$ $(\equiv G_{12})$ and $(\lambda_{21} - \lambda_{22})/RT$ $(\equiv G_{21})$ are treated as adjustable quantities.

The set of expressions for $\ln \gamma_i$ developed here contains adequate flexibility to correlate data for a variety of aqueous systems, including those containing partially dissociated electrolytes and zwitterionic species. We strive to use the simplest expressions (with the fewest number of adjustable parameters) that give adequate representation of the data. The expressions include only pure-component and binary parameters, allowing rational extension to multicomponent systems. Examples of their use appear below.

Correlation of Activity Coefficient Data

We examine the validity of our expressions by correlating activity coefficient data for the L-forms of alanine, serine and threonine in water. Figure 2 illustrates the similarity in molecular structure among these three hydrophilic amino acids. A version of Marquardt's method (Marquardt, 1963) was implemented to perform the data regression and parameter estimation. The objective function is defined by

$$\mathcal{L} \equiv \Sigma\,(\gamma_i^{\mathrm{calc}} - \gamma_i^{\mathrm{exp}})^2$$

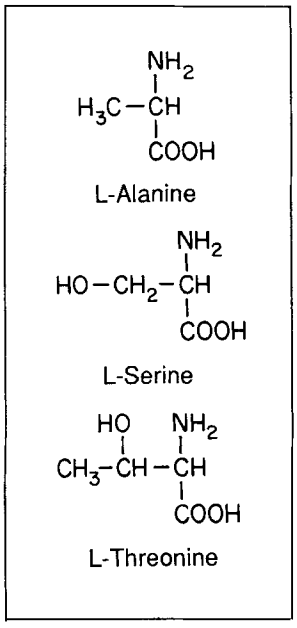


Figure 2. Molecular structures of L-alanine, L-serine and L-threonine.

Table 1 summarizes the results of the regression. The adjustable parameter K represents the equilibrium constant for the formation of a hydrogen bond between the amino acid and a water molecule. The G_{ij} are the energy parameters in the Wilson equation.

Alanine is the simplest of the three compounds; its activity coefficient is approximately linear with molality and nearly unity up to 2 m, Figure 3. The excellent fit we obtain with γ(Wil) is therefore not surprising. The introduction of a hydroxyl group on serine dramatically changes the behavior of the activity coefficient, as shown in Figure 4. We expect that negative deviations from ideality can be represented by the formation of a hydrogen-bonded complex between serine and water. The resulting expression is normalized with respect to the infinite-dilution reference state. However, the results in Table 1 suggest that the physical contribution is the dominant effect; the Wilson equation represents the data just as well as the combination of γ(chem) and γ(Wil), and gives far superior representation over γ(chem) alone. Improvement in the fit of the chemical

Table 1. Results from Fitting Activity Coefficient Data for Amino Acids in Water at 298.15 K*

Amino Acid	Maximum Molality	Terms in Model	K ($\times 10^6$)	G_{12}	G_{21}	Residual Root Mean Square (Molality) ($\times 10^2$)
Alanine	1.86	γ(Wil)	—	0.01	0.47	0.03
Serine	4.0	γ(chem)	16.4	—	—	24.0
Serine	4.0	γ(Wil)	—	0.17	1.20	0.12
Serine	4.0	γ(chem), γ(Wil)	7.17	0.17	1.20	0.12
Threonine	2.0	γ(chem)	3.21	—	—	1.85
Threonine	2.0	γ(Wil)	—	0.17	0.72	0.08
Threonine	2.0	γ(chem), γ(Wil)	3.65	0.17	0.72	0.08

*Data from Fasman (1976).

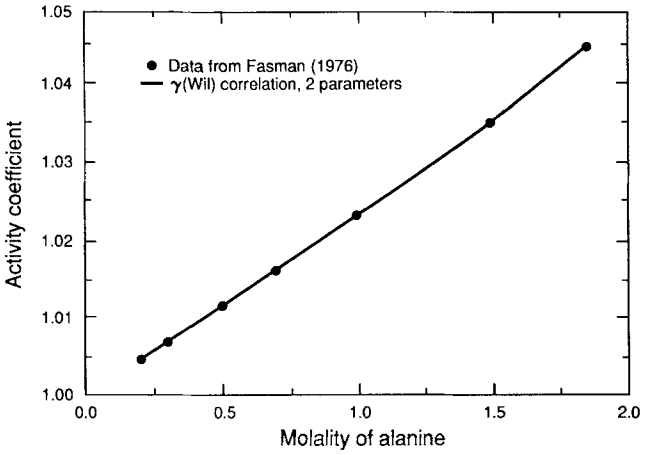


Figure 3. Experimental and calculated values for the activity coefficient of alanine in water at 298.15 K.

term may be possible by including reactions for the self-association of water. Nonspecific hydration of the amino acid leads to differences in local composition and excellent representation of the data by the Wilson equation.

Threonine differs from serine by the addition of a methyl group. Here again the Wilson equation yields an excellent fit of the data, with the chemical term having little influence on the representation of the compound's behavior. Figure 5 illustrates that the activity coefficient of threonine is closer to unity (and therefore the solution is closer to ideality) than that of serine over the same composition range. We expect that the addition of a methyl group decreases the hydrophilicity of threonine; hydration of the molecules may not be as complete as for serine, and therefore differences in the local and overall compositions may not be as great. Indeed, the value of G_{21} for threonine is smaller than the value for serine.

The combination of terms for γ_2 presented in this paper is also valid for representing the liquid-phase activity coefficient of an amino acid in solid-liquid equilibrium. The next section details the development of an expression for the SLE equation in terms of the heat of solution at infinite dilution; results from the corre-

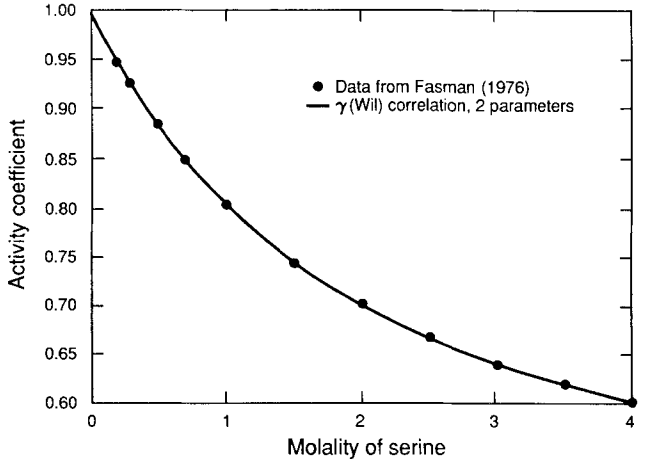


Figure 4. Experimental and calculated values for the activity coefficient of serine in water at 298.15 K.

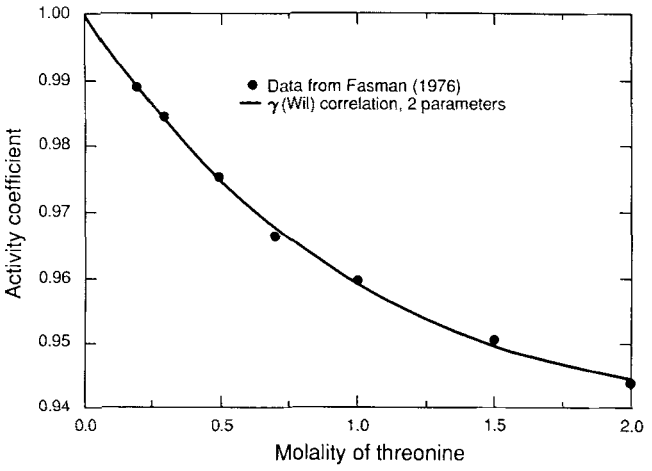


Figure 5. Experimental and calculated values for the activity coefficient of threonine in water at 298.15 K.

lation of solubility data for amino acids appear in the following section.

Formulation of Solid-Liquid Equilibrium Equation

We start with the general criterion for solid-liquid equilibrium (Prausnitz et al., 1986), written in terms of fugacities as

$$\hat{f}_i\,(\text{solid}) = \hat{f}_i\,(\text{aq}) \tag{11}$$

If we assume the formation of a pure solid, then

$$\hat{f}_i\,(\text{solid}) = f_i\,(\text{solid})$$

and by definition,

$$\hat{f}_i\,(\text{aq}) = x_i \gamma_i f_i^{\circ}$$

The quantity f_i° represents the fugacity of species i in a chosen reference state. Here, we specify the state to be that of an aqueous solution at unit molality and temperature T. Our equilibrium equation then becomes

$$x_i \gamma_i = \frac{f_i\,(\text{solid, pure, } T)}{\hat{f}_i\,(\text{aq, 1 } m, T)} \tag{12}$$

We can also write that

$$\Delta G = RT \ln\left(\frac{f_i\,(\text{solid, pure, } T)}{\hat{f}_i\,(\text{aq, 1 } m, T)}\right) \tag{13}$$

The fact that ΔG is a state function allows us to choose any convenient path for evaluation of Eq. 13; Figure 6 outlines the individual steps selected here. We can also write that

$$\Delta G = \Delta H - T\Delta S = \sum_{i=1}^{4} \Delta H_i - \sum_{i=1}^{4} T\Delta S_i \tag{14}$$

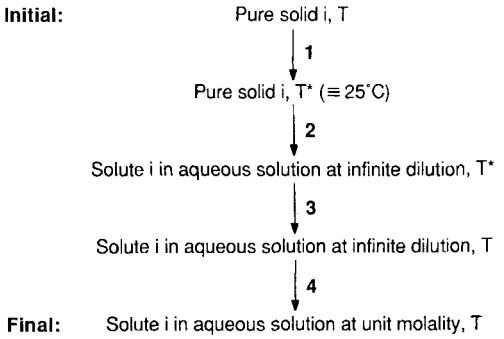


Figure 6. Hypothetical path for the evaluation of ΔG.

The enthalpic and entropic contributions for each individual step may then be expressed as

$$\Delta H_1 = \int_T^{T^*} C_p^s \, dT$$

ΔH_2 = heat of solution at infinite dilution and T^* (25°C)

$$\Delta H_3 = \int_{T^*}^{T} \overline{C}_p^{\infty} \, dT$$

$$\Delta H_4 = 0$$

$$\Delta S_1 = \int_T^{T^*} C_p^s \frac{dT}{T}$$

ΔS_2 = entropy of solution at infinite dilution and T^*

$$\Delta S_3 = \int_{T^*}^{T} \overline{C}_p^{\infty} \frac{dT}{T}$$

ΔS_4 = entropy change from infinite dilution to unit molality at T

Values for the heat capacity of the solute at infinite dilution ($\overline{C}_p^{\infty}$) may be determined experimentally or estimated by conventional methods. The accuracy of such estimates is not critical to the representation of SLE; contributions from ΔH_3 and ΔS_3 are relatively small compared to those from other steps, particularly over a small temperature range. Here, we assume (for convenience) that the heat capacity of water (C_p^w) adequately represents the heat capacity of the hydrated solute (a solute surrounded by, and interacting primarily with, a layer of water) at infinite dilution ($\overline{C}_p^{\infty}$).

The heat of solution ($\Delta H_2 \equiv \Delta H^{\text{sol}}$) at infinite dilution and T^* can be obtained from calorimetric measurements. The quantity ΔH_4 is identically zero, since we require that

$$\lim_{x^w \to 1} \gamma_{\text{solute}} = 1$$

at all temperatures (Robinson and Stokes, 1970). Values for the sum $\Delta S_2 + \Delta S_4$ ($\equiv \Delta S_{2+4}$) may be obtained as part of the regression procedure, or they may be arbitrarily set equal to zero, if appropriate. The heat capacities may be safely assumed to be independent of temperature over a small range.

Table 2. Results from Fitting Solubility Data for Phenylalanine in Water at 323.15 K*

Terms in Model	No. of Parameters	pK_I	pK_{II}	pK_{III}	$\Delta S_{2+4}/R$	G_{12}	G_{21}	Residual Root Mean Square
γ(chem)	1	0.64	†	†	—	—	—	58.8
γ(chem)	2	−0.16	†	†	1.55	—	—	25.3
γ(chem)	3	2.47	7.24	0.02	—	—	—	53.3
γ(chem)	4	−7.35	9.36	0.07	17.6	—	—	32.6
γ(chem), γ(Wil)	4	−1.86	†	†	8.29	1.00	−3.47	16.5

*Data from Jaffari and Thompson (1987).
†Values fixed during regression: $pK_{II} = 9.13$, $pK_{III} = 1.83$ (Greenberg, 1951).

Summation over all steps and substitution into Eq. 14 yields

$$\Delta G = C_p^s (T^* - T) + \Delta H^{sol} + C_p^w (T - T^*) - T\{C_p^s \ln (T^*/T) + \Delta S_{2+4} + C_p^w \ln (T/T^*)\} \quad (15)$$

The ratio of fugacities may now be evaluated by substituting Eq. 15 into Eq. 13 to give

$$\ln\left[\frac{f_i\,(\text{solid, pure, } T)}{\hat{f}_i\,(\text{aq, 1 } m, T)}\right] = -\left\{\left(\frac{C_p^s - C_p^w}{RT}\right)(T^* - T) + \frac{\Delta H^{sol}}{RT} - 1/R\,[(C_p^s - C_p^w)\ln (T^*/T) + \Delta S_{2+4}]\right\}$$

Our complete expression for describing the solid-liquid equilibrium of an amino acid in water then becomes

$$x_i\gamma_i = \exp\left\{\left(\frac{C_p^w - C_p^s}{RT}\right)(T^* - T) - \frac{\Delta H^{sol}}{RT} - 1/R\,[(C_p^w - C_p^s)\ln (T^*/T) - \Delta S_{2+4}]\right\} \quad (16)$$

The use of Eq. 16 as a tool for correlating solubility data is illustrated in the next section.

Correlation of Solubility Data

Tables 2, 3 and 4 summarize the results from correlating solubility data for L-phenylalanine, L-tyrosine and L-diiodotyrosine as a function of pH. The chemical term represents the reactions 1a – 1c given previously as an example; the parameters are the pK_j of the reactions, defined by

$$pK_j \equiv -\log_{10} K_j$$

The values for pK_{II} and pK_{III} were taken from Greenberg (1951), when those parameters were treated as fixed. Additional flexibility of the model obtained when all three pK_j were considered to be adjustable quantities. Fasman (1976) provided values for the heats of solution at infinite dilution and the heat capacities of the solids. The quantity $\Delta S_{2+4}/R$, which appears in the SLE equation, was either fixed at zero, or determined by the regression procedure. Addition of the Wilson equation (Eqs. 9 and 10) provided two more adjustable parameters (G_{12} and G_{21}). Solution of the SLE equation for x_2 (i.e., solubility) in this case required an iterative procedure; we used a simple Newton-Raphson method to determine the value of x_2. The objective function here was defined by

$$\mathcal{L} \equiv \Sigma (x_2^{calc} - x_2^{exp})^2$$

Figure 7 illustrates the regression results from correlating solubility data for phenylalanine with different combinations of adjustable parameters. A single parameter [pK_I in γ(chem)] gives the correct qualitative behavior, as shown in Figure 7a, but the model is unable to represent the magnitude of the solubility curve. Addition of the entropy parameter greatly improves the goodness of fit; inclusion of this parameter is necessary for quantitative reproduction of results, even when other adjustable parameters are included, Figures 7c and 7d. The best overall fit obtains with the combination of chemical and physical terms containing four adjustable parameters. This combination also yields the most physically realistic values for the parameters, as given in Table 2.

Tyrosine is a hydrophobic amino acid similar in structure to phenylalanine, except that it contains a hydroxyl group on the aromatic ring. Figure 8 shows the model representation of the data with different numbers of adjustable parameters. Here again the chemical term with a single parameter yields good qualitative representation. Note that, unlike the phenylalanine/

Table 3. Results from Fitting Solubility Data for Tyrosine in Water at 298.15 K*

Terms in Model	No. of Parameters	pK_I	pK_{II}	pK_{III}	$\Delta S_{2+4}/R$	G_{12}	G_{21}	Residual Root Mean Square
γ(chem)	1	0.10	†	†	—	—	—	0.36
γ(chem)	2	−9.36	†	†	21.7	—	—	0.48
γ(chem)	3	0.33	8.87	2.60	—	—	—	0.32
γ(chem)	4	−9.57	9.37	2.11	21.7	—	—	0.33
γ(chem), γ(Wil)	4	−10.0	†	†	55.3	−1.58	4.61	0.16
γ(chem), γ(Wil)	6	−7.86	9.39	1.84	58.1	−1.79	6.71	0.10

*Data from Hitchcock (1924) and Nass (1987).
†Values fixed during regression: $pK_{II} = 9.11$, $pK_{III} = 2.20$ (Greenberg, 1951).

Table 4. Results from Fitting Solubility Data for Diiodotyrosine in Water at 298.15 K*

Terms in Model	No. of Parameters	pK_I	pK_{II}	pK_{III}	$\Delta S_{2+4}/R$	G_{12}	G_{21}	Residual Root Mean Square ($\times 10^2$)
γ(chem)	1	−1.32	†	†	—	—	—	4.23
γ(chem)	2	−9.44	†	†	18.7	—	—	4.27
γ(chem)	3	−0.56	5.64	2.93	—	—	—	3.66
γ(chem)	4	−0.70	5.67	2.90	0.26	—	—	3.76
γ(chem), γ(Wil)	4	−4.55	†	†	27.7	−0.35	4.72	2.40
γ(chem), γ(Wil)	6	−4.63	7.02	1.58	28.0	−0.40	5.58	2.26

*Data from Dalton et al. (1930).
†Values fixed during regression: $pK_{II} = 6.48$, $pK_{III} = 2.12$ (Greenberg, 1951).

water system, inclusion of the entropy parameter is not critical to the goodness of fit. However, more realistic values for pK_I (negative quantities) obtain when ΔS_{2+4} *is* included in the regression. Differences in the degree of hydrate formation of tyrosine vs. phenylalanine (Fasman, 1976) may be reflected in the relative importance of ΔS_{2+4} to the precision of the correlation. Once again the combination of chemical and physical terms with six adjustable parameters gives the best fit over the whole pH range.

Table 4 presents the results from correlating solubility data of diiodotyrosine in water at 298.15 K. Halogenation of the aromatic ring decreases the solubility of the compound in the isoelectric zone, and shortens the flat region of minimum solubility, as noted by Dalton et al. (1930). Mörner (1913) states that the presence of iodine in the hydroxyphenyl ring markedly increases its acidic properties; our results also show that the regressed values for pK_{II} of tyrosine are much larger (and therefore the equilibrium constants K_{II} are much smaller) than the values of pK_{II} obtained from diiodotyrosine. Note that the values of pK_{III} for the two compounds are approximately the same, however. Such comparisons lend validity to the model (and to the values of the regressed parameters), and suggest that extension of the model to more complex, multicomponent mixtures could be successful.

Simultaneous Correlation of Solubility Data as a Function of pH and Temperature

Correlating expressions should be capable of simultaneously representing the pH and temperature behavior of a system, if the expression is to be useful for the design of biochemical unit operations. Temperature appears naturally in the relationship of the equilibrium constants to the standard Gibbs energy change of reaction j (Smith and Van Ness, 1975):

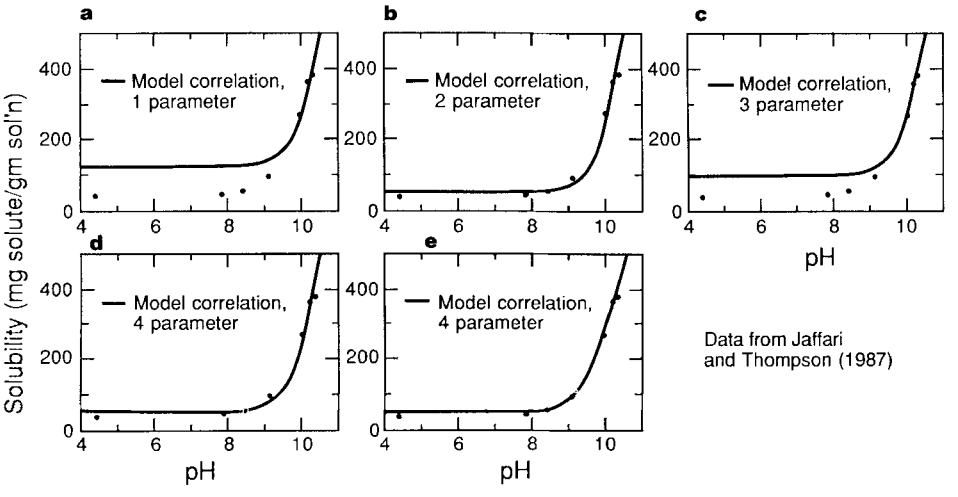
$$\ln K_j = -\frac{\Delta G_j^o}{RT} \tag{17}$$

Application of the Gibbs-Helmholtz equation yields

$$\frac{d \ln K_j}{dT} = \frac{\Delta H_j^o}{RT^2} \tag{18}$$

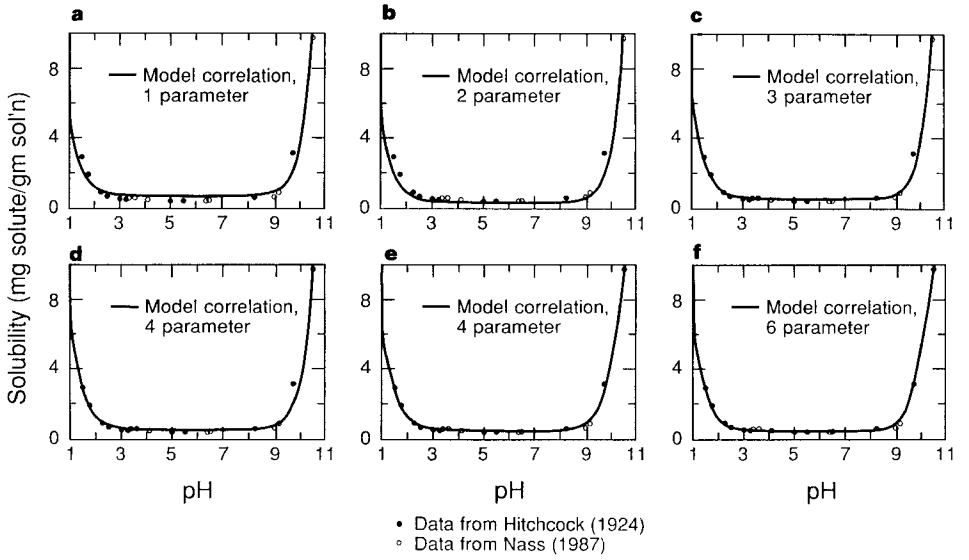
where ΔH_j^o is the standard enthalpy change of reaction j. We assume that ΔH_j^o is independent of temperature (usually a good assumption over a small range); integration of Eq. 18 then gives

$$\ln [K_j(T)/K_j(T^*)] = \frac{\Delta H_j^o}{R}\left(\frac{1}{T^*} - \frac{1}{T}\right) \tag{19}$$



(a)–(d): γ(chem) only
(e): γ(chem) and γ(Wil).

Figure 7. Experimental and calculated values for the solubility of phenylalanine in water at 323.15 K.



(a)–(d): γ(chem) only
(e), (f): γ(chem) and γ(Wil).

Figure 8. Experimental and calculated values for the solubility of tyrosine in water at 298.15 K.

Both $K_j(T^*)$ and ΔH_j^o are now treated as adjustable parameters.

The temperature dependence of the physical contribution arises in the definition of Λ_{ij}, given in Eq. 10 as

$$\Lambda_{ij} \equiv (V_j/V_i) \exp\{-(\lambda_{ij} - \lambda_{ii})/RT\} \qquad (10)$$

If the ratio V_j/V_i and the difference $\lambda_{ij} - \lambda_{ii}$ may be treated as only weak functions of temperature, then Eq. 10 provides an expression that relates temperature to the parameters in the Wilson equation. The regression parameters are (for a binary system) $\lambda_{12} - \lambda_{11}$ and $\lambda_{21} - \lambda_{22}$; if additional flexibility is required, the quantity V_2/V_1 may also be considered adjustable.

Figure 9 illustrates the ability of our expressions to represent simultaneously the pH and temperature dependence of the solubility behavior of diiodotyrosine in water. Data at 273.15, 298.15 and 313.15 K were simultaneously regressed to obtain a single set of parameters; the expressions and resulting values generated the three solid curves shown in Figure 9. The fit of the data is within expected experimental error, except for portions of the curve representing 298.15 K.

We also tested the ability of the expressions to interpolate between two temperatures by first regressing solubility data at 273.15 and 313.15 K, and then calculating values for the solubility at 298.15 K. The results are displayed in Figure 10, and the values of these adjustable parameters, as well as those values corresponding to the curves in Figure 9, are given in Table 5. The expressions successfully predict the solubility behavior of diiodotyrosine at 298.15 K solely from data at other temperatures.

Application of our expressions to biochemical systems containing salts will require the addition of terms to represent

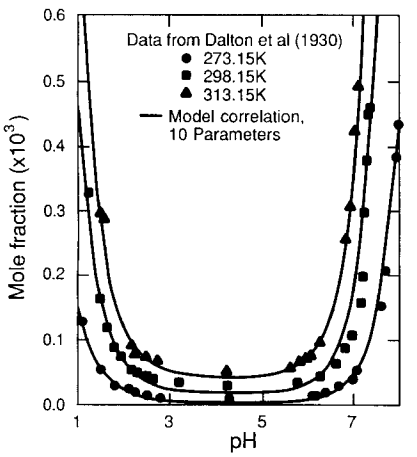


Figure 9. Experimental and calculated values for the solubility of diiodotyrosine in water.

At 273.15 K, 298.15 K and 313.15 K; residual root mean square = 2.68×10^{-2}.

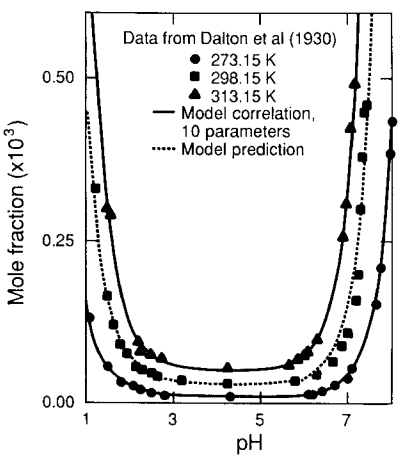


Figure 10. Experimental and calculated values for the solubility of diiodotyrosine in water.

At 273.15, 298.15 and 313.15 K; dotted line represents prediction of behavior at 298.15 K from regression of data at 273.15 K and 313.15 K; residual root mean square = 1.04×10^{-2}.

Table 5. Results from Fitting Solubility Data for Diiodotyrosine in Water

	Fit to Data at	
	273.15, 298.15 and 313.15 K	273.15 and 313.15 K
pK_I	3.00	2.54
pK_{II}	3.11	3.86
pK_{III}	5.43	4.70
$\Delta S_{2+4}/R$	−4.63	−3.74
$\lambda_{12}-\lambda_{11}$*	7.66×10^2	-3.55×10^3
$\lambda_{21}-\lambda_{22}$*	8.95×10^3	1.27×10^4
ΔH_1^{o*}*	-3.25×10^4	-2.90×10^4
ΔH_2^{o*}*	2.47×10^4	4.03×10^4
ΔH_3^{o*}*	1.98×10^4	3.34×10^4
V_2/V_1	1.05×10^{-1}	1.31
Residual Root Mean Square ($\times 10^2$)	2.68	1.04

*J/mol.

long-range ion-ion interactions. We propose to introduce these contributions through one form of the familiar Debye-Hückel expression, given here as

$$\ln \gamma_i(DH) = -\sum_j \frac{z_j^2 e^2}{2\epsilon_o DkT}\left(\frac{\kappa x_j'}{1+\kappa a_j}\right)\left(\frac{1}{\sum_m x_m'}\right) \qquad (20)$$

Here, z_j is the valency of ion j, e is the charge of a proton, ϵ_o is the vacuum permittivity, D is the dielectric constant of the medium, k is Boltzmann's constant, a_j is the distance of closest approach of ion j and κ is a quantity defined by

$$\kappa \equiv \left(\frac{4\pi e^2 \sum_i c_i z_i^2}{\epsilon_o DkT}\right)^{1/2} \qquad (21)$$

The a_j may be fixed or adjustable, depending on the required precision of the correlation. Further testing of the extended model on data for other systems, particularly multicomponent systems containing salts, is needed to verify its usefulness as a design tool.

Acknowledgment

We thank M. Jaffari and P. Thompson for providing solubility data for phenylalanine in water. Discussions with M. M. Abbott, K. Hellyar, and P. J. Makarewicz were helpful and appreciated. S. Baron provided excellent technical assistance with the measurement of tyrosine solubility data.

Notation

A = uncharged amino acid
A^+ = positively charged amino acid
A^- = negatively charged amino acid
$A^\pm$ = zwitterionic amino acid
C_p = constant pressure heat capacity, J/mol · K
$\bar{C}_p$ = constant pressure heat capacity of solute i in solution, J/mol · K
D = dielectric constant of medium
G_{ij} = interaction energy parameter for species i and j
G^E = molar excess Gibbs energy, J/mol
H^+ = hydrogen ion
K_j = equilibrium constant for reaction j
N = number of apparent species
R = gas constant, 8.314 J/mol · K
s = objective function
T = temperature, K
V = Bondi volume, m^3/mol
a = distance of closest approach, m
c = concentration, mol/m^3
e = charge of a proton, 1.60219×10^{-19} C
f = fugacity, Pa
k = Boltzmann's constant, 1.38066×10^{-23} J/K
m = molality, mol/kg solvent
pK = dissociation constant, $\equiv -\log_{10} K$
r = number of reactions
x = apparent mole fraction
x' = true mole fraction
z_j = valency of ion j

Greek letters

γ = apparent activity coefficient
γ' = true activity coefficient
γ(chem) = activity coefficient from chemical theory
$\gamma(DH)$ = activity coefficient from Debye-Hückel equation
γ(Wil) = activity coefficient from Wilson equation
ΔG = change in Gibbs free energy, J/mol
ΔH = change in enthalpy, J/mol
ΔS = change in entropy, J/mol · K
ϵ_o = vacuum permittivity, 8.854188×10^{-12} C^2/J · m
κ = reciprocal Debye length, 1/m
Λ = energy parameter in Wilson equation
λ_{ij} = interaction energy parameter, J/mol
ν_{ij} = stoichiometric coefficient of species i in reaction j
ξ = extent of reaction
η = number of true species

Superscripts

calc = calculated
exp = experimental
o = standard or reference state
s = solid
sol = solution
w = water
* = reference temperature, 298.15 K
Λ = property in solution
∞ = infinite dilution

Subscripts

1 = solvent
2 = solute
2 + 4 = Steps 2 and 4
i, j = any species
I, II, III = reaction I, II or III
m = all true species

Literature Cited

Ball, F. X., W. Fürst, and H. Renon, "An NRTL Model for Representation and Prediction of Deviation from Ideality in Electrolyte Solutions Compared to the Models of Chen (1982) and Pitzer (1973)," *AIChE J.*, **31**, 392 (1985).

Bondi, A., *Physical Properties of Molecular Crystals, Liquids and Glasses*, Ch. 14, Wiley, New York (1968).

Chen, C-C., H. I. Britt, J. F. Boston, and L. B. Evans, "Local Composition Model for Excess Gibbs Energy of Electrolyte Systems," *AIChE J.*, **28**, 588 (1982).

Chen, C-C., "Representation of Solid-Liquid Equilibrium of Aqueous Electrolytes With the Electrolyte NRTL Model," CODATA Symp. on Crit. Eval. and Pred. Phase Equilib. in Multicomponent Systems, Paris (1985).

Chen, C-C., and L. B. Evans, "A Local Composition Model for the Excess Gibbs Energy of Aqueous Electrolyte Systems," *AIChE J.*, **32**, 444 (1986).

Cohn, E. J., and J. T. Edsall, *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Hafner Publishing, New York (1965).
Cruz, J.-L., and H. Renon, "A New Thermodynamic Representation of Binary Electrolyte Solutions Nonideality in the Whole Range of Concentrations," *AIChE J.*, **24**, 817 (1978).
Dalton, J. B., P. L. Kirk, and C. L. A. Schmidt, "The Apparent Dissociation Constants of Diiodotyrosine, Its Heat of Solution, and Its Apparent Heat of Ionization," *J. Biol. Chem.*, **88**, 589 (1930).
Fasman, G. D., ed., *Handbook of Biochemistry and Molecular Biology*, 3rd ed., Vol I, 104, CRC Press (1976).
Greenberg, D. M., *Amino Acids and Proteins, Theory, Methods, Application*, Charles C. Thomas, Publisher (1951).
Hitchcock, D. I., "The Solubility of Tyrosine in Acid and in Alkali," *J. Gen. Physiol.*, **6**, 747 (1924).
Jaffari, M., and P. Thompson, unpublished (1987).
Kretschmer, C., and R. Wiebe, "Thermodynamics of Alcohol-Hydrocarbon Mixtures," *J. Chem. Phys.*, **22**, 1697 (1954).
Marquardt, D., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. Ind. Appl. Math.*, **11**, 431 (1963).
Mörner, C. T., *Z. Physiol. Chem.*, **88**, 124 (1913) [given in Dalton et al. (1930)].
Nagata, I., "On the Thermodynamics of Alcohol Solutions. Phase Equilibria of Binary and Ternary Mixtures Containing any Number of Alcohols," *Fluid Phase Equilibria*, **19**, 153 (1985).
Nass, K. K., "Precise Reduction of Vapor/Liquid Equilibrium Data by Molecularly-Inspired Expressions for the Excess Gibbs Energy," Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, NY (1986).
Nass, K. K., unpublished (1987).
Pitzer, K. S., "Thermodynamics of Electrolytes: I. Theoretical Basis and General Equations," *J. Phys. Chem.*, **77**, 268 (1973).
Prausnitz, J. M., R. Lichtenthaler, and E. de Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibria*, 2nd ed., Ch. 9, Prentice-Hall (1986).
Prigogine, I., and R. Defay (translated by D. H. Everett), *Chemical Thermodynamics*, Ch. XXVI, Longmans, Green and Co., London (1954).
Redlich, O., and A. Kister, "On the Thermodynamics of Nonelectrolyte Solutions and its Technical Applications: III. Systems with Associated Components," *J. Chem. Phys.*, **15**, 849 (1947).
Renon, H., and J. Prausnitz, "On the Thermodynamics of Alcohol-Hydrocarbon Solutions," *Chem. Eng. Sci.*, **22**, 299 (1967); errata **22**, 1891 (1967).
Robinson, R. A., and R. H. Stokes, *Electrolyte Solutions*, 2nd ed., Butterworths, London (1970).
Smith, J. M., and H. C. Van Ness, *Introduction to Chemical Engineering Thermodynamics*, 3rd ed., Ch. 9, McGraw-Hill, New York (1975).
Wilson, G., "Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing," *J. Am. Chem. Soc.*, **86**, 127 (1964).

Manuscript received Dec. 21, 1987, and revision received Apr. 26, 1988.