Charge regulation of natural amino acids in aqueous solutions

Journal:	*Journal of Chemical & Engineering Data*
Manuscript ID	je-2020-00625d.R2
Manuscript Type:	Article
Date Submitted by the Author:	26-Oct-2020
Complete List of Authors:	Gallegos, Alejandro ; University of California Riverside, Department of Chemical and Environmental Engineering Wu, Jianzhong; University of California Riverside, Department of Chemical and Environmental Engineering

Charge regulation of natural amino acids in aqueous solutions

Alejandro Gallegos and Jianzhong Wu*

Department of Chemical and Environmental Engineering, University of California, Riverside, CA 92521, USA

Abstract

The aim of this work is to develop a predictive model for describing the electrostatic behavior of twenty natural amino acids under diverse solution conditions. A coarse-grained model is proposed to account for the key ingredients of thermodynamic nonideality arising from the interaction of amino acids with various solvated ions in an aqueous solution including the molecular volume exclusion effects, solvent-mediated electrostatic interactions, van der Waals attraction, and short-ranged hydrophobic and hydration forces. With a small set of parameters characterizing the intermolecular interactions and dissociation equilibrium, the thermodynamic model is able to correlate extensive experimental data for the activity coefficients and solubility of amino acids in pure water and in aqueous sodium chloride solutions. Moreover, it predicts apparent equilibrium constants for the charge regulation of all amino acids in excellent agreement with experimental data. Importantly, the thermodynamic model allows for the extrapolation of the intrinsic equilibrium constants for amino-acids conversion among different charge states (*viz*., negative, neutral and positive charges) thereby enabling the prediction of the charge behavior for all natural amino acids under arbitrary solution conditions.

* To whom all correspondence should be addressed. Email:jwu@engr.ucr.edu

1. Introduction

Natural amino acids may exist as cations, zwitterions or anions in an aqueous solution depending on pH and other thermodynamic conditions including the ionic strength and solution composition. The charge regulation, *viz*., a precise control of the electrostatic states of amino acids, is important not only for understanding their solution behavior but it has also direct implications to a broad range of biochemical applications of polypeptides and proteins such as bioadhesives, smart materials, and anti-fouling coatings[1].

The speciation of amino acids is intrinsically affiliated with the protonation/deprotonation of the terminal amino and carboxyl groups as well as the ionizable side chains such as those from Lys, Arg and Glu. Whereas the degree of proton association/dissociation of a simple acid or base can be readily determined from the Henderson–Hasselbalch equation[2], predicting the electrostatic states of amino acids is much more complicated due to the coupling of acid-base reactions. As a matter of fact, conventional experimental methods are not able to determine the acid-base equilibrium constants for the individual functional groups of amino acids precisely. Significant variations can often be found in the literature for the experimental values of these fundamental constants[3,4]. Mostly, the uncertainties can be attributed to neglecting the thermodynamic ideality due to the interactions of amino acids with themselves and with other ionic species in the solution. For practical convenience, the thermodynamic nonideality may be lumped together leading to the definition of so-called *apparent* equilibrium constants[5].

For a weak acid or base with a single ionizable group, the apparent equilibrium constant is equivalent to the true thermodynamic equilibrium constant in the limit of infinite dilution (*viz*. in an ideal solution). However, measuring the equilibrium constant for amino acids at infinite dilution is not practical; the reported values from the literature are usually obtained at salt concentrations

as low as 10 mM. The type of experimental procedures (e.g., titration or spectroscopy) can also factor in to the difference in the equilibrium constants[6]. In principle, the equilibrium constants can be predicted from quantum-mechanical (QM) calculations. However, existing computational methods provide only order-of-magnitude estimates for the formation free energies and the numerical results are highly sensitive to the selection of approximations including the system size and specific basis sets[7]. Reliable theoretical tools are yet to be developed such that they can predict apparent equilibrium constants well correlated with experimental data at finite salt concentrations and allow for the extrapolation of the experimental results to amino acids at infinite dilutions.

For a particular acid-base equilibrium, the *apparent* equilibrium constant amalgamates the true thermodynamic equilibrium constant with the activity coefficients of all pertinent chemical species. The experimental values are closely related to the thermodynamic nonideality due to solvent-mediated interactions among ionic as well as non-ionic species. Because of the long-range nature of electrostatic correlations, the activity coefficients can vary quite significantly even at low salt concentrations. Besides, various types of solvent-mediated short-range interactions could make the *apparent* equilibrium constant highly sensitive to the identity of salt ions.

In a dilute solution, amino-acid dissociation deviates from the ideal behavior mostly dominated by their electrostatic interactions with ionic species. In this case, the Debye-Hückel (DH) theory provides a reasonable description of the activity coefficients. It predicts that, at low salt concentrations, the activity coefficients of charged species (modeled as point charges) sharply falls as the ionic strength increases[8]. Because of the strong deviation from thermodynamic ideality even at low salt concentrations, the equilibrium constants cannot be reliably extrapolated from theoretical methods that do not reproduce the DH limiting law. As the DH theory becomes problematic when the salt concentration increases, the activity coefficients can be described by

using various forms of the extended DH theory[8, 9]. The revised forms retain the strong dependence of activity coefficients on the ionic strength when the salt concentration is sufficiently small, but predict a slower decay in the activity coefficient at moderate and high salt concentrations. In comparison with experimental data, the extended DH theories are typically valid up to about 0.1 M for a monovalent electrolyte solution. Beyond which the mean activity coefficient of the electrolyte shows a minimum and increases at higher salt concentrations. In the latter case, the short-range interactions between ionic species become important, thus additional terms must be included to describe the activity coefficients quantitatively[10].

For applications to acid-base equilibrium, a conventional method to represent the activity coefficients of various ionic species at high salt concentrations is given by the specific ion interaction (SIT) model[11, 12]. In this method, the activity coefficients of individual ions are predicted by the DH theory augmented with an additional term for non-electrostatic interactions that is linearly proportional to the ionic strength. The proportionality coefficient is typically a function of the ionic strength and varies with the identity of ionic species. Despite its simplicity, SIT performs reasonably well in capturing the activity coefficients associated with acid-base equilibrium in simple electrolytes. It has been applied to predicting the charge regulation of acids and bases whereby the activity coefficient of a neutral species is assumed to be only a linear function of the ionic strength[6, 13, 14]. A slightly more rigorous approach can be taken to describe non-electrostatic interactions through the virial expansion of the non-electrostatic component of the excess Gibbs energy similar to the Pitzer's equation[15]. The modified SIT model provides a near quantitative description of the activity coefficients for aqueous solutions of both monovalent salts and more complex electrolytes. Regrettably, the quantitative performance of this model is compromised by a large number of empirical parameters that are often unavailable for different

amino acids[16]. Not surprisingly, Pitzer's equation does well to capture the charge regulation of many acids and bases if sufficient experimental data are available to fix the adjustable parameters; albeit it involves a more complex procedure for numerical fitting than that for the original SIT model[17].

Over the years, a number of thermodynamic models have been developed for electrolyte solutions beyond semi-empirical modifications of the DH theory[18, 19]. Among them, the mean-spherical approximation (MSA) represents a popular choice owing to its computational simplicity to describe the ion-size effects and electrostatic correlations[20]. Importantly, MSA provides an accurate representation of both thermodynamic and transport properties of electrolyte solutions using only the ionic diameters as the adjustable parameters[21, 22]. Importantly, it obeys the limiting law at low electrolyte concentrations similar to various empirical modifications of the DH theory[23]. Besides, further improvements can be accomplished by including the ion pairing effects[24]. MSA has been previously applied to modeling the charge regulation of acids and bases[25-27]. For example, Vilariño et al. used MSA to correlate the apparent equilibrium constants of certain amino acids with satisfactory results[9, 26]. From a practical perspective, one additional benefit that MSA has over alternative theories is that the ion diameters have an intuitive physical meaning and typically fall into the expected ranges for the size of hydrated ions. Typically, an anion is less hydrated than a cation of similar diameter and the same absolute charge, thus the Pauling radius provides a reasonable estimate of the anion diameter. In contrast, cation hydration is usually nonnegligible and its diameter must be adjusted to match the experimental data[28]. Remarkably, MSA is able to correlate both thermodynamic and transport coefficients quantitatively for many electrolyte systems up to high concentrations. Further improvement can be achieved by making the cation diameter and/or the dielectric constant concentration-dependent[21]. The concentration dependence

of the ion diameter or the dielectric constant is understandable and may be determined from independent experiments[29, 30]. On the one hand, the change in the diameter of a hydration ion can be attributed to enhanced electrostatic screening as the salt concentration increases[31]. On the other hand, the dielectric constant of an electrolyte solution usually falls as more salt is added because the solvated ions reduce the orientational polarization of water molecules[30, 32].

It should be noted that the original MSA model provides a poor correlation of the activity coefficients of amino acids in aqueous solutions. For example, MSA fails to predict the reduction of activity coefficients of both glycine and sodium chloride when more glycine is added to a sodium chloride solution[33]. Because glycine exists predominately as zwitterions near a neutral pH, the primitive model would predict an increase of the activity coefficients for both glycine and sodium chloride. Understandably, an aqueous solution of amino acids is not a simple electrolyte. The thermodynamic properties depend not only on the solute size and long-range electrostatic interactions but also on the short-ranged forces such as van der Waals (vdW) attraction, hydrogen bonding, and various forms of water-mediated interactions. Alternative models have been proposed to incorporate these interactions such that a quantitative correlation can be achieved for the activity coefficients of amino acids in pure water and in aqueous solutions[6, 34-37]. For example, a perturbation theory was proposed to correlate the activity coefficients of both amino acids and peptides in pure water[35]. With the primitive model of electrolyte solutions as the reference, two perturbation terms were introduced to account for thermodynamic properties due to short-ranged interactions: one for the dispersion forces as represented by the Lennard-Jones (LJ) potential, and the other for the angle-averaged dipole-dipole interactions between amino-acid molecules. While the dipole moments were determined from quantum-mechanical calculations, two adjustable parameters were introduced in the perturbation theory, *viz*., the size and energy parameters of the

LJ potential for each amino acid or peptide. Quantitative correlations were established for the activity coefficients of a number of amino acids in liquid water. The perturbation approach was extended to study the activity coefficients of amino acids in sodium chloride solutions[34]. In the latter case, the mean activity coefficient of the electrolyte was described by the MSA model with concentration-dependent cation diameters[31], and the perturbation terms include additional thermodynamic contributions due to the LJ attraction, the dipole-induced dipole and the charge-dipole interactions between amino acids and the electrolyte. With the model parameters previously determined for amino acids in pure water and two additional adjustable parameters, the perturbation theory was able to correlate the activity coefficients for both amino acids and sodium chloride quantitatively.

Whereas most electrolyte theories assume water as a dielectric continuum, a number of so-called non-primitive models have been proposed over the years[38, 39]. Held et al. considered, in addition to the explicit representation of water molecules, the anisotropic nature of amino-acid molecules by using the perturbated-chain version of the statistical associating fluid theory (PC-SAFT)[40]. With a generic model accounting for hard-sphere repulsion, dispersion attraction, and association of polar groups, PC-SAFT was able to correlate various thermodynamic properties (e.g., densities, vapor-pressure depression, activity coefficients, and solubility) for many amino acids and oligopeptides in pure water. The thermodynamic model was extended to amino acids-electrolyte aqueous systems by combining DH and the PC-SAFT equations[36]. This model captures the thermodynamic properties of electrolyte/amino acid/water solutions by using parameters determined only from the pure components and mixing rules that do not include any additional adjustable parameters. The ability to predict the mixture behavior using pure component parameters is beneficial because data for mixtures are often limited or even not available. Later,

Bang et al. studied a wide variety of these mixture systems using a combination of MSA and PC-SAFT equations[41]. It was found that the formation of ion-pair complexes is indispensable to describe the activity coefficients of both amino acids and the electrolytes. While PC-SAFT provides an excellent correlation of diverse experimental results for electrolyte systems, it involves a large set of adjustable parameters.

The purpose of this work is to develop a coarse-grained model that can be used to predict the charge behavior of all twenty natural amino acids under various solution conditions. While there has been extensive work on acid-base equilibrium and activity coefficients of electrolyte solutions, a predictive model has yet to be developed for charge regulation in aqueous solutions of natural amino acids. In order to derive reliable equilibrium constants for protonation/deprotonation reactions, such a model should well capture the key forces underlying interactions between amino acids and all ionic species in an aqueous medium. While a precise description of hydration and water-mediated forces in microscopic details is beyond the scope of the present work, their contributions to thermodynamic properties can be described with semi-empirical models such that the parameters can be fixed by comparison with experimental data. The coarse-grained approach allows us to best estimate the equilibrium constants for transition among different electrostatic states of natural amino acids.

2. Thermodynamic Model and Methods

2.1 *Electrostatic states of amino acids*

Each amino acid consists of a carboxyl group and an amine group that can be dissociated in an aqueous solution through acid-base equilibrium. These functional groups, along with possible ionizable groups in the side chain, are responsible for the titration behavior of amino acids.

Therefore, amino acids can exist in multiple states of electrostatic charge depending on pH and other solution conditions.

Conventionally, natural amino acids are divided into three types according to the charge propensity of the side chain, *viz*., neutral, acidic, and basic. Neutral amino acids contain no ionizable group in their side chains and exist predominately in charge neutrality over a broad range of solution conditions. Meanwhile, an acidic amino acid contains an acidic functional group in its side chain (e.g. glutamic acid has a carboxyl group in its side chain) which is deprotonated near a neutral pH. Similarly, a basic amino acid contains a basic functional group in the side chain and typically bears a positive charge at pH = 7.0.

Because each amino-acid molecule contains at least two ionizable groups, we can express the acid-base equilibrium in terms of multi-step protonation reactions:

$$A_{i-1} + H^+ \leftrightarrow A_i \tag{1}$$

where A_i refers to the amino acid in a given charge state. The amino-acid valence satisfies $Z_{A_i} = Z_{A_0} + i$ with Z_{A_0} being the charge for the amino acid in its fully deprotonated state. For a neutral or basic amino acid, Z_{A_0} = -1; while an acidic amino acid molecule has a valence of -2 in its fully deprotonated state. Accordingly, in the fully protonated state, each neutral or acidic amino acid has a valence of +1 and each basic amino acid has a valence of +2.

The equilibrium constant for the protonation reaction is related to the solution composition and activity coefficients

$$K_i^T = \frac{c_{A_i}}{c_{H^+} c_{A_{i-1}}} \frac{\gamma_{A_i}}{\gamma_{H^+} \gamma_{A_{i-1}}} \tag{2}$$

where K_i^T represents the equilibrium constant for the amino acid in charge state *i*, superscript *T* denotes true thermodynamic constant, c_i is the molar concentration of species *i*, and γ_i is the

corresponding activity coefficient. The equilibrium constant is a thermodynamic quantity defined by the change in the chemical potentials of reactants and products at their reference states, i.e., each species in an ideal solution at unit molar concentration. The activity coefficients account for the effect of solvent-mediated interactions among all species in the solution.

Whereas the solution composition can be monitored with various analytical tools, it is not feasible to directly measure the true thermodynamic equilibrium constant because the activity coefficients are often unknown. As a result, the equilibrium constant is conventionally expressed in terms of its apparent value in most practical applications

$$K_i' = \frac{c_{A_i}}{c_{H^+} c_{A_{i-1}}} = K_i^T \frac{\gamma_{A_{i-1}} \gamma_{H^+}}{\gamma_{A_i}} . \quad (3)$$

At infinite dilution, the activity coefficients are negligible and there is no difference between apparent and true equilibrium constants. In general, the *apparent* equilibrium constant is influenced not only by the change of the chemical potential for individual species but also by the solvent-mediated interactions.

The apparent equilibrium constant provides direct information on the speciation of the amino acid at a given solution condition. From the equilibrium constant and mathematical relations between the activity coefficients and the solution composition, Eq. (3) allows us to calculate the distributions of amino acids in different charge states. For fitting the thermodynamic model with experimental results, it is convenient to express the equilibrium constants in the logarithmic form:

$$\log K_i' = \log K_i^T + \log \gamma_{A_{i-1}} + \log \gamma_{H^+} - \log \gamma_{A_i} . \quad (4)$$

The experimental results for acid-base equilibrium are typically reported in terms of the solution concentrations and apparent equilibrium constants. Therefore, a thermodynamic model is needed to calculate the activity coefficients and extrapolate the true equilibrium constants.

2.2 *A coarse-grained model for the aqueous solutions of amino acids*

As mentioned above, a number of models can be used to represent the thermodynamic properties of amino acids in aqueous solutions, ranging from those with atomic details to coarse-grained approaches with the solvent treated implicitly as a dielectric continuum[5, 40, 42, 43]. Nonetheless, understanding the charge regulation of amino acids remains a challenge because it requires quantitative predictions of both equilibrium constants and activity coefficients over a broad range of conditions. In this work, we propose an augmented primitive model (APM) to represent amino acids in aqueous systems. The coarse-grained model describes amino acids in different charge states as hard spheres with different charge valences. Given a solution condition, the system composition can be determined from the protonation/deprotonation equilibrium self-consistently.

For simplicity, we lump all short-range interactions between amino acids and ionic species in an aqueous environment in terms of a square-well potential. While the simple model lacks atomic details and does not reflect the anisotropic nature of amino acid molecules, it provides a flexible framework to represent thermodynamic nonideality due to the transition between different electrostatic states of amino acids. As discussed above, the primitive model has been routinely used to correlate the activity coefficients of aqueous electrolyte solutions. We expect that similar results can be accomplished by adding a square-well potential to account for different types of solvent-mediated interactions. Despite extensive research, a first-principles approach to predicting water-mediated interactions remains theoretically challenging. The adoption of an effective potential allows us to capture the essential physics without evoking the microscopic details[20, 44].

With amino acids in aqueous systems represented by APM, the activity coefficient for each solute species can be decomposed into contributions from the molecular excluded-volume effects, electrostatic correlation, and water-mediated interactions

$$\ln \gamma_i = \beta\mu_i^{ex} = \beta\mu_i^{hs} + \beta\mu_i^{el} + \beta\mu_i^{sw}. \quad (5)$$

In Eq. (5), μ_i^{ex} stands for the excess chemical potential, *i.e.*, deviation from the chemical potential of species *i* in an ideal solution; $\beta = 1/(k_B T)$, k_B is the Boltzmann constant and *T* the absolute temperature, μ_i^{hs}, μ_i^{el}, and μ_i^{sw} are contributions to the excess chemical potential due to hard-sphere repulsion, electrostatic correlation, and solvent-mediated interactions, respectively. The expression for each contribution has been discussed previously and can be found in Supporting Information[23, 44, 45].

The parameters in our activity-coefficient model include the hard-sphere diameters for each amino acid in different charge states, and the energetic parameters for short-range attraction between different pairs of solute species. The determination of these parameters will be discussed in the following section. As discussed previously, the hard-sphere diameters are ascribed to the excluded volumes of hydrated molecules thereby depending on the charge states. We find that, approximately, both the hydration shell thickness and water-mediated interactions are invariant with the change in salt concentration. At a fixed temperature, all the model parameters are assumed to be independent of solution conditions.

2.3 *Evaluation of model parameters*

We calibrate the activity-coefficient model for twenty natural amino acids in aqueous solutions of sodium chloride. A similar procedure can be applied to aqueous systems containing other ionic species. The sodium chloride system is considered in this work because extensive data are available for all amino acids.

To determine the activity coefficients and the apparent equilibrium constant of acid-base equilibrium, we need the size parameters for all solute species and the interaction energies. For amino acids dissolved in a NaCl solution, the solute species include amino acids at different charge states, sodium and chloride ions, and protons or hydroxide ions. The parameters affiliated with amino acids can be fixed by fitting the APM model with their activity coefficients and solubility data in pure water and in aqueous NaCl solutions. As discussed previously, the equilibrium constants for amino-acid transition between different charge states can be determined by extrapolation of the apparent equilibrium constants to those corresponding to infinite dilution.

In addition to amino acids, we need to fix the model parameters for salt ions, protons and hydroxide ions. Previous studies indicate that the activity coefficients of electrolyte solutions can be well correlated with the hard-sphere diameters of cations and anions within the mean spherical approximation (MSA)[21, 28]. At high salt concentrations, the performance can be further improved by making the cation diameter and/or the dielectric constant decrease with the ionic strength. In this work, we are mainly concerned with protonation/deprotonation of amino acids in aqueous electrolytes of moderate concentrations. As shown in Figure S2, the MSA model is able to reproduce the mean activity coefficient for sodium chloride from 0.0 to 1.0 M with the chloride ion diameter fixed at 3.62 Å (its Pauling diameter) and that of the sodium ion equal to 3.22 Å[46]. Throughout this work, the hard-sphere diameter for hydronium ions is fixed at 5.00 Å because it provides a reasonable estimation of the apparent equilibrium constant for natural amino acids when we ignore the activity coefficients for amino acids. This value falls within range of hydronium ion diameter previously estimated by others[9, 21]. The diameter of hydroxide ions was determined to be 3.22 Å based off comparison to the activity coefficients of sodium hydroxide solutions (see Figure S2).

Throughout this work, the parameters for amino acids were estimated by fitting experimental data using MATLAB R2018b. These parameters, along with comparison of the numerical results and experimental data, are presented in Supporting Information (Table S1). In all cases, the following objective function (OF) was minimized:

$$\mathrm{OF} = \sum [X_i^{exp} - X_i^{cal}]^2 \tag{6}$$

where X refers to the experimental data for the activity coefficients of amino acids, solubility data, or the apparent equilibrium constants. In certain cases, such as systems containing amino acids in charged states, it was necessary to combine the objective function for different data sets in order to best fit all experimental results.

3. Results and Discussion

3.1 Activity coefficients for amino acids

Given the added complexity in determining the activity coefficients of amino acids with multiple charge states, we first consider neutral amino acids in aqueous sodium chloride solutions around their isoelectric points. In this case, we can correlate the experimental data in terms of three adjustable parameters: the hard-sphere diameter for each amino acid and the energy parameters for amino acid interaction with itself and with the salt ions. These parameters are listed in Table S1.

Figure 1 presents a comparison between theoretical correlations and experimental data for the activity coefficients of four representative neutral amino acids at different salt concentrations with the solution pH equal to the respective isoelectric points. Similar comparisons for other neutral amino acids and for the activity coefficients of NaCl are presented in the SI. In all cases, our model provides a quantitative description of how the activity coefficient of amino acid is influenced by its own concentration and by the changes in the solution conditions. For each neutral

amino acid, the hard-sphere diameter is typically 10 to 20% smaller than the molecular diameter determined from the van der Waals (vdW) volume[47]. The smaller effective size may be attributed to the non-spherical shape and the zwitterionic nature of neutral amino acids[48]. Interestingly, the tendency of the hard-sphere diameter to approach that corresponding to the vdW volume increases with the hydrophobicity of the amino acids (e.g. Ala → Val → Leu). The trend may be attributed to the compression of the water layer in the solvation shell as the amino acid becomes more hydrophilic[49]. The choice of the vdW volume to describe the molecular size of the neutral amino acids provides a semi- to near quantitative fitting of the activity coefficient data, thus to a certain degree demonstrating the physical significance of our model parameters.

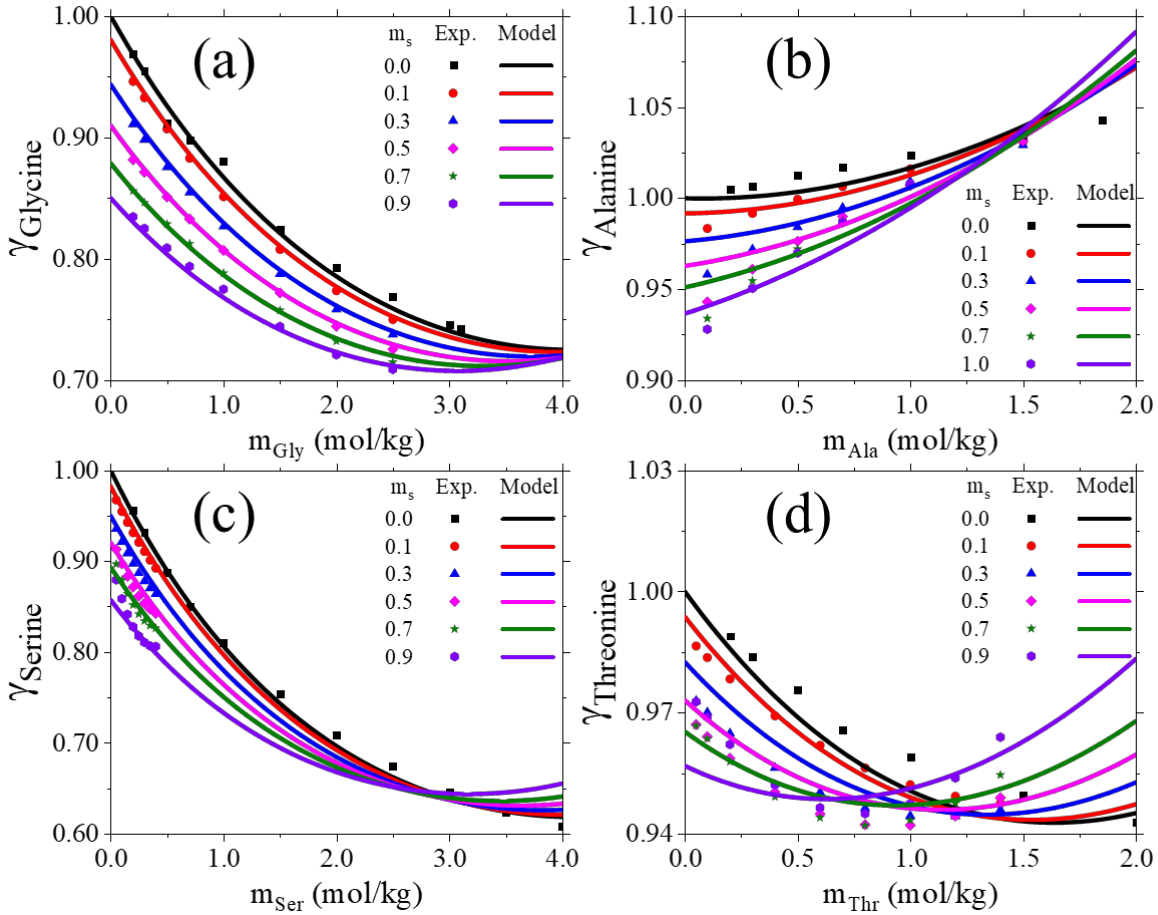

Figure 1. Activity coefficients of (a) glycine (pI = 5.98), (b) alanine (pI = 6.18), (c) serine (pI = 5.72) and (d) threonine (pI = 5.72) in NaCl aqueous solutions from experiment[34, 50-52] (symbols) and from theoretical correlations of the coarse-grained model (solid lines).

Figure 1a shows that the activity coefficient of glycine falls significantly as the concentration of amino acid or NaCl increases. The decline in the activity coefficient may be attributed to the high polarity of glycine molecules and to the salting out effect of sodium chloride. Because the amino acid is modeled as a neutral particle in this work, the negative deviation from the ideal solution behavior implies that the attractive contribution has a greater effect than the repulsion or the excluded-volume effects. At high salt molality, the different curves in Figure 1a collapse almost into a single point suggesting that the hard-sphere repulsion outcompetes the attraction contributions. The non-linear dependence is consistent with the prediction of the hard-sphere model for the excess chemical potential[45].

Figure 1b shows that the activity coefficient of alanine increases with its concentration but decreases with more added salt. In comparison to glycine, however, the dependence on the salt concentration is significantly smaller. These trends are captured in our model by using a non-electrostatic attraction energy between the amino-acid molecules lower than that for their interaction with the salt ions. The decrease in the attraction energy from Gly to Ala can be attributed to hydrophobic effect due to the addition of the methyl group in Ala's side chain, which causes the water layer surrounding the amino acid to become more structured. The higher ordering of the solvation shell weakens the interaction of the amino acid with itself and salt ions in the solution. A similar trend was observed for the change in the interaction energy from Ala to Val and from Ser to Thr as shown in Table S1. In both cases, the energy parameters falls as the carbon number of the side chain increases and the trend agrees with the findings by Khoshkhabari et al.[35].

If instead of a methyl group, a hydroxyl group is added to alanine's side chain so that it becomes serine, the activity coefficient will behave similarly to that of glycine (Figure 1c). In fact, the concentration effect on the activity coefficient of serine is much more substantial than that for glycine in terms of both the addition of amino acid and salt ions. This stronger dependence on the solute concentration is somewhat expected due to the increased capability of hydrogen bonding and dipole-dipole interactions that the additional hydroxyl group can contribute. Interestingly, if the addition to alanine were instead a weak carboxyl group like in the case of aspartic acid, the interaction strength would decrease (Table S1). While this may seem to oppose the previous statement regarding the increase of attraction due to the polar functional group, it may rather be explained by the interaction between the ionic carboxyl group and the ionic amine group, which weakens the intermolecular interactions[53]. If the carboxyl group is replaced by an amide group like that in asparagine, the attraction energy greatly increases due to the amide group not participating in ionic interactions with the amine group.

Figure 1d shows the activity coefficient of threonine, which differs from serine by one methyl group. It exhibits an initial decrease in the activity coefficient with added threonine or salt, but the trend is reversed when the threonine molality is greater than about 1.0 mol/kg. The non-monotonic trend is expected because of the reduced attractions of the amino acid with itself and with the salt in comparison to serine. While the discrepancy between the correlated and experimental data is noticeable, we consider the overall performance as satisfactory given the simplicity of this model.

For amino acids in aqueous electrolyte solutions, experimental data are often available for the activity coefficients of both amino acids and ionic species. Figure S4 compares our model correlation of the mean activity coefficients of NaCl with the experimentally determined values

for those systems considered in Figure 1. While these data were not used in the fitting, our coarse-grained model is able to describe the mean activity coefficients of NaCl in good agreement with experiment. For all four amino acids considered, the mean activity coefficient of sodium chloride decreases with an increase in the molality of the amino acid or the salt ions. The decline in the mean activity coefficients can be attributed both to the water-mediated attraction between amino acids and salt ions and to the electrostatic correlation effects. It is known experimentally that, when amino acid is added to a NaCl solution, the dielectric constant increases due to the high polarity[54]. Because the dielectric constant is inversely proportional to the strength of electrostatic correlations (see Eq. S10), the addition of amino acids reduces electrostatic attraction between salt ions thereby reducing its contribution to the mean activity coefficient of sodium chloride. Indeed, one may improve the theoretical performance by considering the dielectric behavior of the solution due to added amino acids.

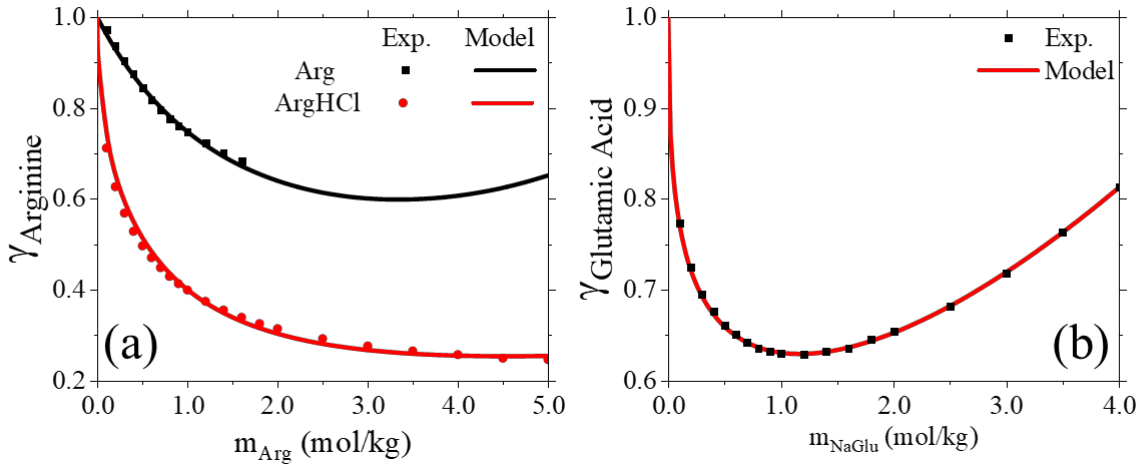

Figure 2. Activity coefficients of amino acids versus the molality for (a) arginine and its salted form (ArgHCl) in pure water and (b) for glutamic acid in salted form (NaGlu). Symbols are from experiments[53] and the lines are theoretical correlations.

The computational procedure to obtain the parameters for acidic and basic amino acids is more complicated than that for neutral amino acids. This is because, even at the isoelectric point, there are more than one charge states in coexistence for an acidic or basic amino acid in an aqueous solution. In order to determine the activity coefficients of both amino acids and salt ions, we must consider the speciation of the amino acid at each solution condition. The composition of the amino acid at different charge states are solved from the condition of chemical equilibrium in couple with the mass and charge balance conditions. The activity coefficient for the amino acid is then given by a partial contribution from each charge state (see Eq. S16). For example, at its isoelectric point (pI = 3.24), approximately 88% of glutamic acid exists in the neutral state, and 6% in the positively and negatively charged states, respectively. While the neutral state has a dominant composition, the contribution from the charged states to the activity coefficient of glutamic acid cannot be neglected due to the importance of electrostatic interactions at low salt concentrations. For acidic and basic amino acids, the thermodynamic properties are often reported in the salted form (e.g. sodium glutamate, NaGlu). In that case, the amino acid exists predominately in the charged state (e.g. Glu^-) and the activity coefficient measured experimentally has an equal contribution from the amino acid and the salt ion (see Eq. S25).

In Figure 2, we present a comparison of our model correlation with the experimental data for the activity coefficients of arginine and glutamic acid. One noticeable difference between the amino acid in its salted and pure forms is that the salted form exhibits a much smaller activity coefficient due to the electrostatic contribution. The excellent agreement exemplifies the power of the MSA model to account for electrostatic correlations over the entire concentration range considered in this work. Interestingly, for the activity coefficient of arginine hydrochloride (ArgHCl) shown in Figure 2a, it does not display the characteristic increase due to the hard-sphere

repulsion, even up to 5 mol/kg of amino acid. The stark difference may be attributed to short-range attractions between large organic compounds absent in simple electrolyte solutions. As shown in Table S1, glutamic acid has a self-attraction energy significantly lower than that of arginine. As a result, it shows a minimum in the activity coefficient curve similar to that for simple electrolytes. While the effects of hard-sphere repulsion and electrostatic correlations are well accounted for within the MSA model, inclusion of short-range attractions is essential to properly capture the thermodynamic behavior of amino acids in aqueous electrolyte solutions.

3.2 Apparent equilibrium constants

Unlike the true equilibrium constant, the apparent constants depend on pH, and the concentrations of amino acids and salt. These two quantities are equivalent only for amino acids at infinite dilutions. Nevertheless, the apparent equilibrium constants provide insight into the charge regulation of amino acids in response to the changes of the solution conditions. Qualitatively, they can be used to replace the true thermodynamic equilibrium constants of protonation/deprotonation. Because of the thermodynamic nonideality as represented by the activity coefficients of all pertinent species, we expect that apparent equilibrium constants would lead to good quantitative performance only over a narrow range of solution conditions.

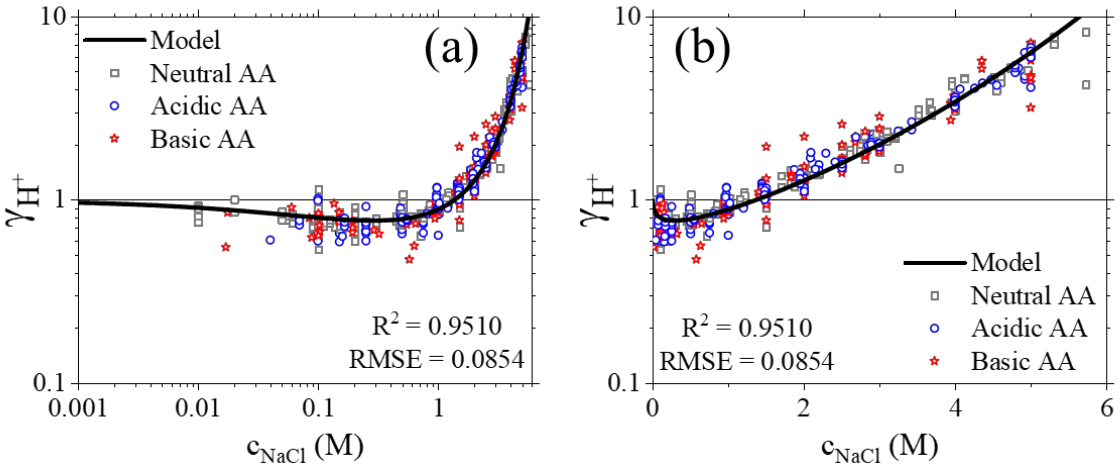

Figure 3. The activity coefficient of protons (hydronium ions) extracted from the apparent equilibrium constants of natural amino acids in NaCl solutions (Eq. 7). Here the solid line corresponds to the MSA prediction with the hard-sphere diameter for hydronium ions equal to 5.00 Å, and the symbols are based on the experimental data for the apparent equilibrium constants with the activity coefficients of amino acids calculated self-consistently.

To facilitate a connection between apparent and true equilibrium constants, we may express their ratio in terms of the activity coefficient of protons (*viz*., hydronium ions):

$$\gamma_{H^+} = \frac{K_i'}{K_i^T} \frac{\gamma_{A_i}}{\gamma_{A_{i-1}}}. \tag{7}$$

For an amino acid at low concentration, the ratio of the activity coefficients in the protonated and deprotonated state is approximately a constant, mostly determined by the change of the valence due to protonation/deprotonation. Meanwhile, the proton activity is mainly determined by its interaction with salt ions, insensitive to the types of amino acids at low concentration. Figure 3 shows a comparison of the theoretical prediction with the experimental results for apparent equilibrium constants with the activity coefficients of the amino acids predicted self-consistently. Approximately, the experimental data points collapse into a single curve for all natural amino acids. The scattering of data around the line for the proton activity coefficient shows experimental errors and non-specific salt effects not considered in the model. As the salt concentration approaches zero, the curve approaches the ideal limit (i.e. $\gamma_i = 1$). In this case, the apparent equilibrium constant would be equivalent to the true thermodynamic equilibrium constant. Most experimental results are obtained at sodium chloride concentrations at or greater than 0.01 M. Because experimental measurements of the apparent equilibrium constants become more difficult

at lower salt concentrations, a reliable theoretical model is needed to extrapolate an accurate value of the true equilibrium constant for the protonation/deprotonation equilibrium of each amino acid.

We have compared how well the model describes the apparent equilibrium constants as a function of the salt concentration for different types of amino acids. Figure 4 presents the results for alanine, a neutral amino acid. For each protonation step, the apparent equilibrium constant depends on the sodium chloride concentration. Since the size of alanine in its neutral state was determined previously by comparison with experimental activity coefficient data, there is only one adjustable parameter for each equilibrium reaction, *viz*., the size of the amino acid in its respective charge state. The parameters are shown in Table S1. As expected, alanine in its neutral state has the smallest volume, which agrees with the partial molar volume data for neutral amino acids in different charge states[55].

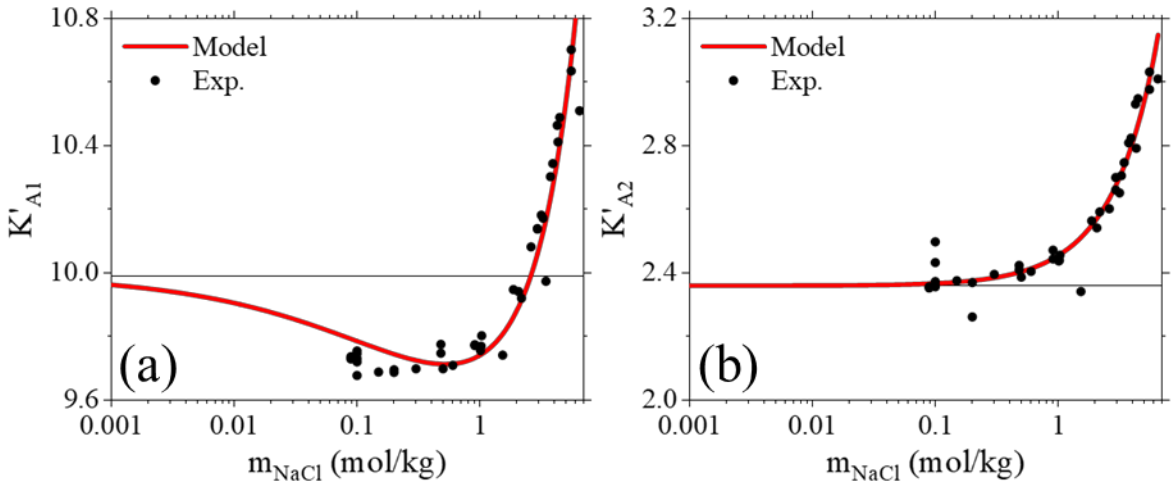

Figure 4. Apparent equilibrium constants for alanine versus NaCl concentration according to experiments[3, 17] (symbols) and our model (lines). (a) $Ala^{-1} + H^{+} \rightarrow Ala^{0}$ and (b) $Ala^{0} + H^{+} \rightarrow Ala^{+1}$.

As shown in Figure 4, the apparent equilibrium constant for the first protonation step ($Ala^{-1} + H^{+} \rightarrow Ala^{0}$) exhibits a minimum as the salt concentration increases. However, such a minimum disappears for the second protonation step ($Ala^{0} + H^{+} \rightarrow Ala^{+1}$). Instead, the apparent equilibrium

constant increases monotonically with more added salt. The different trends arise from the imbalance between the absolute charge of reactants and products (*i.e.*, A^- and H^+ versus A^0) for the first protonation step while the second step has the same total absolute charge on both sides of the reaction. In the former case, there is an electrostatic contribution to the apparent equilibrium constant which would favor the direction toward more charged species. As a result, the apparent equilibrium constant decreases with an increase in the salt concentration in the dilute regime. For the second protonation step, there is equal charge for the reactants and the product, leading to the cancelation of the electrostatic contributions. At higher salt concentrations, both steps exhibit a characteristic increase in the apparent equilibrium constants with the ionic strength. This can be explained by the larger contribution of the hard-sphere repulsion to the activity coefficients of the chemical species. The equilibrium shifts towards the side that has the least volume occupancy, i.e., the product side in both protonation steps. In other words, the excluded volume effects lead to the increase in the apparent equilibrium constant. A similar trend can be observed for other neutral amino acids as shown in Figures S11-14. Clearly, our model is able to well account for how the apparent equilibrium constant changes with the sodium chloride concentration for all neutral amino acids. It thus enables extrapolation of the true equilibrium constants from the experimental data.

Next, we consider how the apparent equilibrium constants for acidic amino acids change with the sodium chloride concentration by taking glutamic acid as an example. For an acidic amino acid, the protonation reaction takes place in three steps with the valence increasing from -2 to 1. Figure 5 shows a comparison between our model and the experimental data. For glutamic acid, the second and third protonation steps are similar to the two protonation steps for a neutral amino acid (e.g., alanine). The appearance of a minimum in the apparent equilibrium constant also

highlights the importance of the charge distribution on both sides of the chemical reaction. In a dilute solution, the electrostatic interaction is the governing factor to determine how the apparent equilibrium constant changes with the salt concentration. By contrast, the trend is little variant with the specific functional groups being protonated. In all cases, we see a sharp increase in the apparent equilibrium constant at high salt concentrations. Meanwhile, the behavior at low salt concentrations is mainly determined by electrostatic effects on protonation reactions. The electrostatic effects also explain why the apparent equilibrium constant for the first protonation step is significantly reduced in compared to that for the second step for glutamic acid and other acidic amino acids. Because the electrostatic component of the activity coefficient of an ionic species scales with the squared valence, glutamic acid in the divalent anion state is significantly more difficult to protonate at low salt concentrations (i.e. the equilibrium reaction favors the A^{2-} species) compared to that for the monovalent anion state (Figure 5b). Overall, the agreement of our model correlation with the experimental data is quite satisfactory.

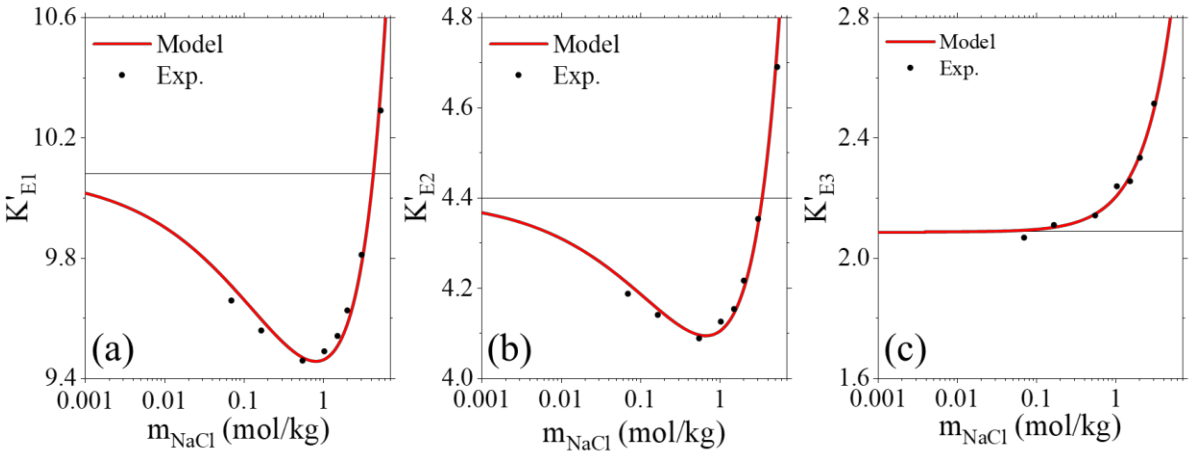

Figure 5. Apparent equilibrium constants of protonation for glutamic acid from experiments[13] (symbols) and model (lines). (a) $Glu^{-2} + H^{+} \rightarrow Glu^{-1}$, (b) $Glu^{-1} + H^{+} \rightarrow Glu^{0}$, and (c) $Glu^{0} + H^{+} \rightarrow Glu^{+1}$.

Lastly, we consider the protonation of the basic amino acids in aqueous sodium chloride solution. Figure 6 shows the apparent equilibrium constants for the protonation of histidine from negative charge to divalent positive charges. The dependence of the apparent equilibrium constants on the salt concentration looks quite similar to that of alanine for the first two steps of protonation. However, histidine has one additional protonation step leading to a divalent cation state. By a comparison of Figure 6b and 6c, we see that both apparent equilibrium constants exhibit a monotonic increase with the salt concentration. The third protonation step shows a more rapid increase in the apparent equilibrium constant than the second protonation step at low salt concentrations. As discussed earlier, the different effect results from the electrostatic contribution which favors the divalent cation form similar to the case with glutamic acid. Unlike a divalent anion in the glutamic acid case, however, the divalent cation form appears on the product side of the chemical equilibrium leading to a more rapid increase in the apparent equilibrium constant. The electrostatic behavior is well captured by our model demonstrating its capability in describing all steps of protonation for different amino acids.

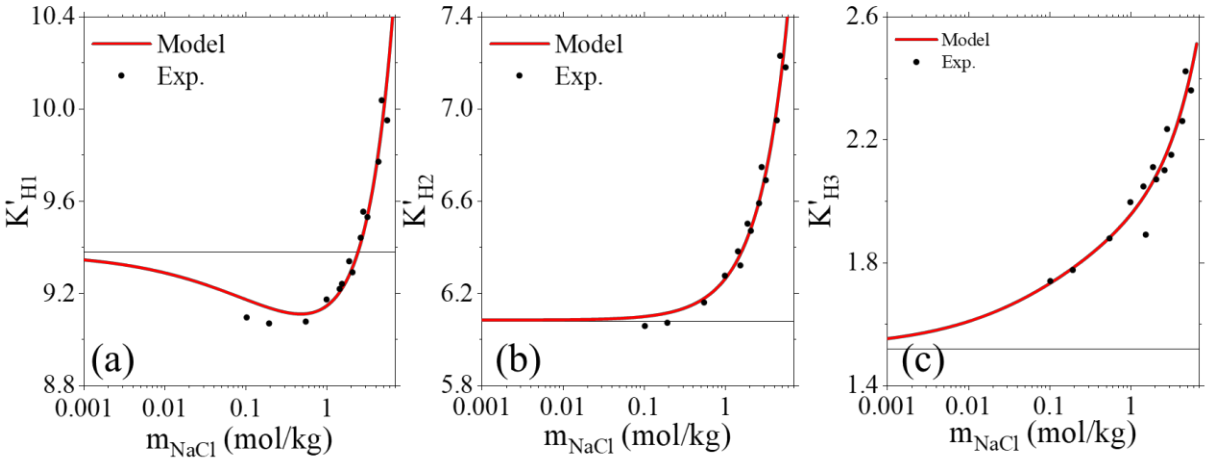

Figure 6. Apparent equilibrium constants of protonation for histidine from experiments[13] (symbols) and model (lines). ((a) $His^{-1} + H^{+} \rightarrow His^{0}$, (b) $His^{0} + H^{+} \rightarrow His^{+1}$ and (c) $His^{+1} + H^{+} \rightarrow His^{+2}$.

As mentioned earlier, the hard-sphere diameter of an amino acid depends on its charge state. The exact value is obtained by fitting the experimental data for apparent equilibrium constants for different types of protonation. We use only one adjustable parameter for each protonation step because the size of the amino acid in its neutral state can be determined by comparison with the activity coefficient data. For example, the hard-sphere diameter for alanine is 4.91 Å in its neutral state. By correlation with the experimental data for the apparent equilibrium constants, we found that the diameters of alanine in the monovalent cation and anion states to be 5.76 Å and 5.44 Å, respectively. Qualitatively, the neutral state is the smallest in size in good agreement with the experimental data for the partial molar volumes of amino acids in different charge states [55, 56].

For a neutral amino acid, we can intuitively understand the increase of the hydration diameter as it becomes charged because of the appearance of ion-dipole interaction between the solute and the solvent. However, this trend does not hold for acidic or basic amino acids. We found that glutamic acid has the smallest hard-sphere diameter in the monovalent anion form. Similarly, histidine has the smallest hard-sphere diameter in the monovalent cation form. The change in the molecular size with the valence also agrees with the experimental data for the partial molar volumes of amino acids referenced earlier. For these cases, amino acid hydration in different charge states cannot be simply explained in terms of the electrostatic interactions. The hard-sphere diameter reflects an interplay between the side chain charge (and amino acid as a whole) and the hydrogen bonding capability of the functional group in the protonated and deprotonated forms. It

has been found[57] that the carboxyl group decreases its partial molar volume upon deprotonating. The trend is opposite for an amine group which leads to a higher partial molar volume when it is deprotonated. The first protonation step of an amino acid involves the protonation of the amine group thus the hard-sphere diameter will decrease according to the change in the partial molar volume. A basic amino acid has an additional amine group in comparison to alanine and its protonation from neutral to a positive charge reduces the size. As the protonation of the carboxyl group increases the size, an acidic amino acid should have a minimum size in the monovalent negatively charged state.

3.3 Thermodynamic equilibrium constants

In order to properly characterize the charge regulation of amino acids at arbitrary solution conditions, we need the true thermodynamic equilibrium constants for all protonation reactions, which are intrinsic properties for each amino acid depending only on temperature. These fundamental constants provide a basis to predict when amino acids will switch from one state to another due to the protonation of their functional groups. As mentioned earlier, these thermodynamic equilibrium constants are quite difficult to accurately measure and the reported values rely on extrapolation techniques with the activity coefficients correlated with various theoretical models. While conventional methods like Pitzer's equation and SIT are useful to correlate the activity coefficients, these models involve a large number of empirical parameters that are unavailable for many amino acids. By contrast, our model uses only a small number of parameters to account for the size effects and electrostatic correlations. All these parameters have clear physical significance and can be estimated from alternative experiments or obtained by correlation with a relatively small amount of experimental data. Therefore, our model allows us to

extrapolate the true equilibrium constants of protonation for natural amino acids at room temperature

We demonstrated in the previous sections that our coarse-grained model well captures the electrostatic behavior of amino acids in sodium chloride solutions. Based on the calibration of our model with apparent equilibrium constants from the literature, we are able to obtain a set of thermodynamic equilibrium constants for protonation of all twenty amino acids (Table 1). These constants provided the best fit to the experimental data when they are extrapolated to the infinite dilution limit. While the results were extracted from experiments in NaCl solutions, the true thermodynamic equilibrium constants are independent of the properties of NaCl thus useful to determine the charge states of amino acids under other electrolyte conditions.

Table 1. Recommended equilibrium constants for the protonation of natural amino acids.

Amino Acid	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$
Ala	9.99	2.36	-
Arg	12.65[a]	9.01	1.75
Asn	9.02	2.13	-
Asp	10.17	3.97[a]	1.93
Cys	10.65	8.34[a]	1.85
Gln	9.24	2.15	-
Glu	10.08	4.40[a]	2.09
Gly	9.73	2.24	-
His	9.38	6.08[a]	1.52
Ile	9.82	2.34	-
Leu	9.82	2.34	-
Lys	10.95[a]	9.09	1.82
Met	9.41	2.15	-
Phe	9.27	2.14	-
Pro	10.73	1.95	-

Ser	9.29	2.16	-
Thr	9.20	2.23	-
Trp	9.60	2.30	-
Tyr	10.63[a]	9.30	2.10
Val	9.76	2.30	-

[a]The pK_a value for the side chain.

Tables S2-S4 compare the equilibrium constants obtained from this work with those extrapolated by using other methods like Pitzer's equation and SIT. In Table S2, we compare the predictions for certain neutral amino acids using different theoretical methods. We see that there is not much variation between the different methods for the two protonation steps. This is expected because all methods provide a good fitting of the activity coefficients. One key advantage of our model is that all parameters have clear physical meanings and can be fitted with a small amount of data. A similar finding is found for those amino acids with acidic side chains (Table S3) and those with basic side chains (Table S4). The general agreement across different thermodynamic models further highlights the benefit of our model in terms of its simplicity.

3.4 Speciation of amino acids in pure water and in sodium chloride solution

As discussed before, amino acids in water can exist in terms of multiple charge states due to the protonation of the functional groups. The change between different charge states is governed by the thermodynamic equilibrium constant. In section 3.3, we proposed a set of thermodynamic equilibrium constants for all twenty amino acids based off the correlations of our model to the apparent equilibrium constants measured by experiments. These constants can be employed under the assumption that the solution is ideal to study how the solution pH influences the speciation of amino acids. Alternatively, the apparent equilibrium constants measured experimentally can be used instead to demonstrate the effect that the non-ideality on the speciation. This is appropriate if the amino acid concentration is dilute compared to the electrolyte such that the protonation to

its different states does not impact the solution conditions, particularly the ionic strength. This is also true when the solution pH is adjusted since there will be noticeable changes in the ionic strength at very low or very high pH. With the true equilibrium constants and the thermodynamic model to account for interaction between solute species, we can now predict the speciation of amino acids in pure water and in NaCl solution at any concentrations.

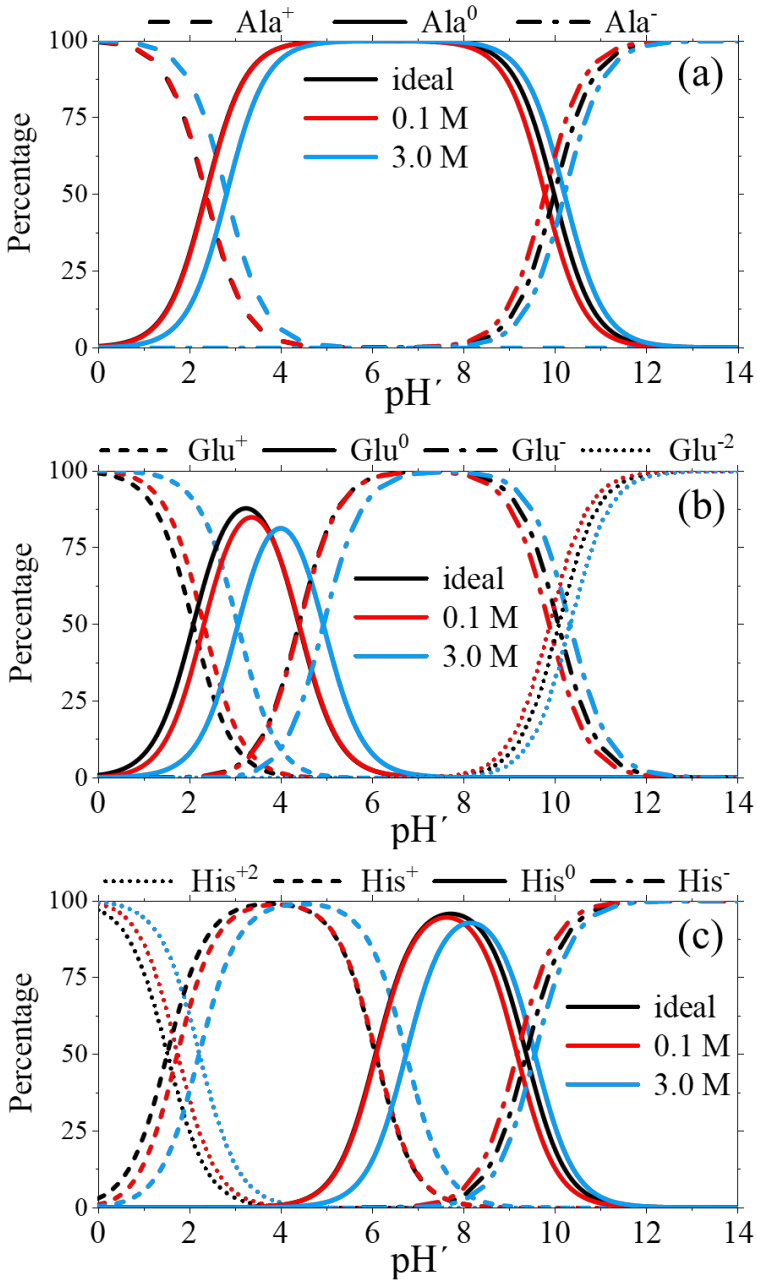

Figure 7. Speciation diagrams for three types of amino acids: (a) neutral, (b) acidic, and (c) basic as represented by alanine, glutamic acid, and histidine, respectively. Shown here are the percentages of amino acids in different charge states at ideal condition and in 0.1 and 3.0 M sodium chloride concentrations.

We may elucidate the speciation of amino acids by considering three representative amino acids in different NaCl concentrations. Figure 7 shows the speciation diagrams of alanine, glutamic acid, and histidine as the representatives of the neutral, acidic, and basic amino acids. The speciation is plotted in terms of $pH' = -\log[H^+]$, i.e. in terms of the logarithm of the concentration of free protons in the solution. A smooth transition between different states of amino acids is observed as pH′ increases. Figure 7a shows that alanine can speciate into three different charge states (i.e. Z = -1, 0, +1). In the pH range between 4 and 8, alanine exists entirely in the neutral state. At either lower or higher pH′, alanine exists in only two of the three possible charge states. Near the isoelectric point (pI = 6.18), we only need to consider alanine as a neutral species. During the transition from one charge state to another, there is an intersection between the two lines which reflects the apparent equilibrium constant at these conditions. Clearly, the apparent equilibrium constant changes with the salt concentration. In the first protonation step (i.e. $Ala^{-1} \rightarrow Ala^{0}$), the pH′ required to protonate 50% of alanine to its neutral state shifts to a lower value as the salt concentration increases from the ideal-solution condition to 0.1 M, which reflects the decrease in the apparent equilibrium constant as shown in Figure 7a. However, a further increase in the salt concentration results in a higher pH′ (i.e., a lower hydrogen concentration) to reach the same level of protonation. As discussed previously, this reflects the shift towards the product side of the first protonation step in order to reduce the volume of the chemical species. For the second protonation step, there is no significant difference between the ideal and 0.1 M case due to the hard-sphere

contribution being relatively insignificant at low salt concentrations. A further increase in the salt concentration leads to the second protonation of alanine occurring at higher pH′ values.

As shown in Figure 7b and c, glutamic acid and histidine exist in four different charge states when they are dissolved in to an aqueous solution. In the region where the pH′ is between 6 to 8, glutamic acid exists in its monovalent anionic state which has a narrower pH′ window in comparison to that for alanine. On the other hand, the neutral state only reaches about 88% of the speciation of glutamic acid around its isoelectric point (i.e. pH′ = 3.24). This is the reason why it was necessary to consider the speciation of glutamic acid in calculating the activity coefficients. For acidic or basic amino acids, we could not consider only the predominate state like that used for the neutral amino acids. It is also interesting to note the large shift in the position of the isoelectric point and the decrease in the percentage that exists in the neutral state as the salt concentration increases. Figure 6c shows the speciation of histidine in pure water and two NaCl solutions. It exhibits a higher presence of the neutral state than that for glutamic acid because the protonation steps are not as close together with each other. As was seen for glutamic acid, there can be significant shifts in the speciation as a result of the increase in the salt concentration.

Another way to look at the speciation of amino acids is the pH effect on the average charge of the amino acid. Figure 8 shows how the average charge of alanine, glutamic acid, and histidine changes with pH′ for the ideal case and two NaCl concentrations. As expected, we see clear influence of the salt concentration on the net charge of the amino acids. For alanine, the second protonation step shows the onset of charge at higher pH′ values than that for the salt-free case. Glutamic acid, on the other hand, shows a noticeable change in the net charge curve when the salt concentration increases (Figure 8b). At higher salt concentrations, we can observe how the narrowing and shift of the pH′ window for the glutamic acid in its neutral state results in a

significant change in the net charge in the range of pH′ between 2 to 6. While histidine does not exhibit as close of protonation steps as glutamic acid, the increase in salt concentration still makes noticeable changes in the net charge in the transition regions (Figure 8c).

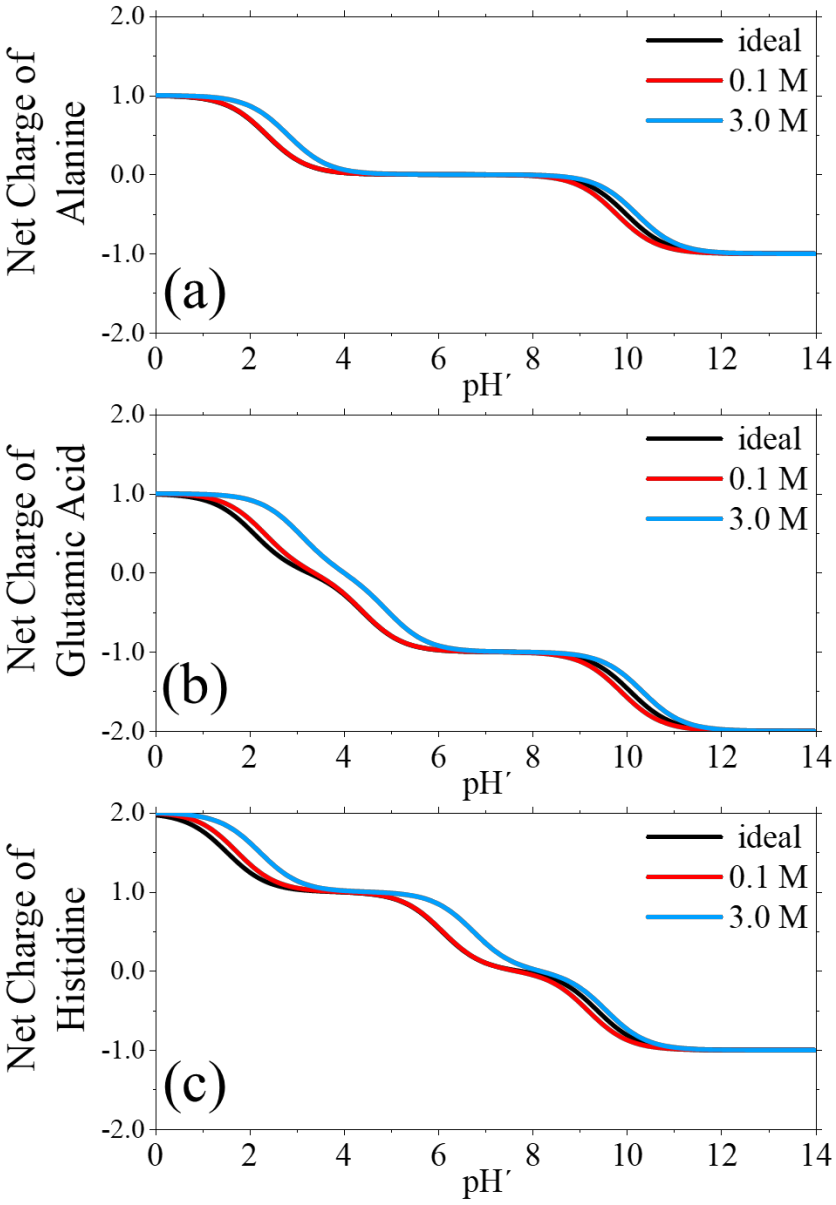

Figure 8. The net charge of (a) neutral, (b) acidic, and (c) basic amino acids versus pH′ (i.e. the hydrogen concentration) at three solution conditions: pure water, 0.1 M and 3.0 M sodium chloride solutions.

4. Conclusion

In this work, we have developed an augmented primitive model (APM) that captures the thermodynamic properties of all twenty natural amino acids in aqueous sodium chloride solutions. In addition to the volume exclusion effects due to the finite size of the solute species and electrostatic correlations, APM accounts for the water-mediated attraction of amino acids with themselves and with salt ions. The activity coefficients are described in terms of the mean-spherical approximation (MSA) for the primitive model and a mean-field approximation for the short-range attraction represented by the square-well potential.

We first demonstrated that APM is able to describe the activity coefficients of amino acids in pure water and in aqueous sodium chloride solutions. This was done by correlating the size parameters of each amino acid in the neutral state and the interaction energy with itself and with the salt ions. The model is able to quantitatively capture the experimental data for all natural amino acids. More importantly, the parameters determined through the fitting procedure maintain physical significance (e.g. the size of the amino acid in its neutral state is consistent with that determined from the van der Waals volume). We find that the short-range attraction tends to outcompete the repulsive excluded volume contribution to the activity coefficient leading to a reduction in the activity coefficient upon addition of either amino acid or salt. In addition, the addition of polar groups results in an increase in the interaction energy while hydrophobic groups diminish the interaction energy. Through the use of MSA to account for the finite size effects and electrostatic correlations, our model can well capture the change in the activity coefficient between the neutral state and charged state of the amino acid by comparison to the experimental data for amino acids in their natural and salted (e.g. ArgHCl) forms.

Lastly, we considered the speciation of amino acids by comparison of the theoretical predictions with experimental data for the apparent equilibrium constants in aqueous sodium chloride solutions. Our model is able to well describe the apparent equilibrium constants for all protonation steps of neutral, acidic, and basic amino acids by correlating the size of the amino acid in its different charge states and the thermodynamic equilibrium constants (i.e., the apparent equilibrium constants in the limit of infinite dilution). The change in size between the different charge states of amino acids is in good agreement with the partial molar volume data determined from experiments. The simplicity of our coarse-grained model and its ability to accurately capture the charge regulation of all amino acids provide confidence in our proposed set of the equilibrium constants of different protonation reactions for all natural amino acids. These fundamental constants will be valuable in determining the charge states of amino acids and understanding their biochemical properties under complicated solution conditions.

Supporting Information:

The supporting information contains theoretical details including the expressions for activity coefficient, osmotic coefficient, solubility, and mass/charge balance conditions. We also list all the correlated values (sizes and interaction energies) within our model determined by fitting the experimental data. In addition, it includes additional figures that compare the theoretical predictions with the experimental data for the activity coefficients, solubility, and apparent equilibrium constants and the references for experimental data that are used in calibration of our model. Lastly, we have also included a statistical comparison of all the data calibrated in our model as well as a tabulated form of the data presented in this work.

Acknowledgements: This work is financially supported by the National Science Foundation Harnessing the Data Revolution Big Idea under Grant No. NSF 1940118. Additional support is provided by the NSF Graduate Research Fellowship under Grant No. DGE-1326120.

References

1. Gitlin, I.; Carbeck, J. D.; Whitesides, G. M., Why are proteins charged? Networks of charge-charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angew Chem Int Ed* **2006,** 45, 3022-3060.
2. Atkins, P. W.; De Paula, J.; Keeler, J., *Atkins' physical chemistry molecular thermodynamics and kinetics*. Eleventh edition. ed.; Oxford University Press: Oxford, United Kingdom ; New York, NY, 2019; p xix, 445 pages.
3. Bretti, C.; Giuffrè, O.; Lando, G.; Sammartano, S., Modeling solubility and acid-base properties of some amino acids in aqueous NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures. *SpringerPlus* **2016,** 5, 928-928.
4. Berthon, G., Critical evaluation of the stability constants of metal complexes of amino acids with polar side chains (Technical Report). *Pure Appl Chem* **1995,** 67, 1117-1240.
5. De Stefano, C.; Foti, C.; Gianguzza, A.; Rigano, C.; Sammartano, S., Chemical speciation of amino acids in electrolyte solutions containing major components of natural fluids. *Chem Spec Bioavailab* **1995,** 7, 1-8.
6. Bretti, C.; Crea, F.; De Stefano, C.; Sammartano, S.; Vianelli, G., Some thermodynamic properties of dl-Tyrosine and dl-Tryptophan. Effect of the ionic medium, ionic strength and temperature on the solubility and acid–base properties. *Fluid Phase Equilib* **2012,** 314, 185-197.
7. Gómez-Bombarelli, R.; González-Pérez, M.; Pérez-Prior, M. T.; Calle, E.; Casado, J., Computational Calculation of Equilibrium Constants: Addition to Carbonyl Compounds. *J. Phys. Chem. A* **2009,** 113, 11423-11428.
8. Bickmore, B. R.; Wander, M. C. F., Activity and Activity Coefficients. In *Encyclopedia of Geochemistry: A Comprehensive Reference Source on the Chemistry of the Earth*, White, W. M., Ed. Springer International Publishing: Cham, 2018; pp 21-23.
9. Vilariño, T.; Fiol, S.; L. Armesto, X.; Brandariz, I.; E. Sastre de Vicente, M., Effect of ionic strength on the protonation of various amino acids analysed by the mean spherical approximation. *J Chem Soc Faraday Trans* **1997,** 93, 413-417.
10. Elizalde, M. P.; Aparicio, J. L., Current theories in the calculation of activity coefficients—II. Specific interaction theories applied to some equilibria studies in solution chemistry. *Talanta* **1995,** 42, 395-400.
11. Guggenheim, E. A.; Turgeon, J. C., Specific interaction of ions. *J Chem Soc Faraday Trans* **1955,** 51, 747-761.
12. Bretti, C.; Foti, C.; Sammartano, S., A new approach in the use of SIT in determining the dependence on ionic strength of activity coefficients. Application to some chloride salts of interest in the speciation of natural fluids. *Chem Spec Bioavailab* **2004,** 16, 105-110.
13. Bretti, C.; Cigala, R. M.; Giuffrè, O.; Lando, G.; Sammartano, S., Modeling solubility and acid-base properties of some polar side chain amino acids in NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures. *Fluid Phase Equilib* **2018,** 459, 51-64.

14. Bretti, C.; Giuffrè, O.; Lando, G.; Sammartano, S., Modeling solubility and acid–base properties of some amino acids in aqueous NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures. *SpringerPlus* **2016,** 5, 928.
15. Pitzer, K. S., *Activity Coefficients in Electrolyte Solutions*. 2018.
16. Kim, H. T.; Frederick, W. J., Evaluation of Pitzer ion interaction parameters of aqueous electrolytes at 25.degree.C. 1. Single salt parameters. *J Chem Eng Data* **1988,** 33, 177-184.
17. De Stefano, C.; Foti, C.; Gianguzza, A.; Sammartano, S., The interaction of amino acids with the major constituents of natural waters at different ionic strengths. *Mar Chem* **2000,** 72, 61-76.
18. Lee, L. L., *Molecular thermodynamics of electrolyte solutions*. World Scientific: Singapore ; Hackensack, NJ, 2008; p xii, 251 p.
19. Hossain, N.; Ravichandran, A.; Khare, R.; Chen, C. C., Revisiting electrolyte thermodynamic models: Insights from molecular simulations. *AlChE J* **2018,** 64, 3728-3734.
20. Blum, L., Mean spherical model for asymmetric electrolytes. *Mol Phys* **1975,** 30, 1529-1535.
21. Simonin, J.-P.; Blum, L.; Turq, P., Real Ionic Solutions in the Mean Spherical Approximation. 1. Simple Salts in the Primitive Model. *J Phys Chem* **1996,** 100, 7704-7709.
22. Simonin, J.-P.; Bernard, O., Organic electrolyte solutions: Modeling of deviations from ideality within the binding mean spherical approximation. *Fluid Phase Equilib* **2018,** 468, 58-69.
23. Maribo-Mogensen, B.; Kontogeorgis, G. M.; Thomsen, K., Comparison of the Debye–Hückel and the Mean Spherical Approximation Theories for Electrolyte Solutions. *Ind Eng Chem Res* **2012,** 51, 5353-5363.
24. Bernard, O.; Blum, L., Binding mean spherical approximation for pairing ions: An exponential approximation and thermodynamics. *J Chem Phys* **1996,** 104, 4746-4754.
25. Vilarino, T.; de Vicente, M. E. S., Theoretical calculations of the ionic strength dependence of the ionic product of water based on a mean spherical approximation. *J Solution Chem* **1997,** 26, 833-846.
26. Vilariño, T.; Sastre de Vicente, M. E., Protonation of Glycine in Saline Media: Evaluation of the Effect of Ionic Strength by Use of the Mean Spherical Approximation. *J Phys Chem* **1996,** 100, 16378-16384.
27. Vilariño, T.; Barriada, J. L.; Sastre de Vicente, M. E., The mean spherical approximation and the prediction of the size of the species involved in an ionization equilibrium in saline media. *Phys Chem Chem Phys* **2001,** 3, 1053-1056.
28. Triolo, R.; Grigera, J. R.; Blum, L., Simple electrolytes in the mean spherical approximation. *J Phys Chem* **1976,** 80, 1858-1861.
29. Waluyo, I.; Huang, C.; Nordlund, D.; Bergmann, U.; Weiss, T. M.; Pettersson, L. G. M.; Nilsson, A., The structure of water in the hydration shell of cations from x-ray Raman and small angle x-ray scattering measurements. *J Chem Phys* **2011,** 134, 064513-064513.
30. Nörtemann, K.; Hilland, J.; Kaatze, U., Dielectric Properties of Aqueous NaCl Solutions at Microwave Frequencies. *J Phys Chem A* **1997,** 101, 6864-6869.
31. Lu, J.-F.; Yu, Y.-X.; Li, Y.-G., Modification and application of the mean spherical approximation method. *Fluid Phase Equilib* **1993,** 85, 81-100.
32. Gavish, N.; Promislow, K., Dependence of the dielectric constant of electrolyte solutions on ionic concentration: A microfield approach. *Phys Rev E* **2016,** 94, 012611.

33. Soto-Campos, A. M.; Khoshkbarchi, M. K.; Vera, J. H., Activity coefficients of the electrolyte and the amino acid in water + NaNO3 + glycine and water + NaCl + dl-methionine systems at 298.15 K. *Biophys Chem* **1997,** 67, 97-105.
34. Khoshkbarchi, M. K.; Vera, J. H., A Perturbed Hard-Sphere Model with Mean Spherical Approximation for the Activity Coefficients of Amino Acids in Aqueous Electrolyte Solutions. *Ind Eng Chem Res* **1996,** 35, 4755-4766.
35. Khoshkbarchi, M. K.; Vera, J. H., A Simplified Perturbed Hard-Sphere Model for the Activity Coefficients of Amino Acids and Peptides in Aqueous Solutions. *Ind Eng Chem Res* **1996,** 35, 4319-4327.
36. Held, C.; Reschke, T.; Müller, R.; Kunz, W.; Sadowski, G., Measuring and modeling aqueous electrolyte/amino-acid solutions with ePC-SAFT. *J Chem Thermodyn* **2014,** 68, 1-12.
37. Breil, M. P.; Mollerup, J. In *Thermodynamics, Experimental, and Modelling of Aqueous Electrolyte and Amino Acid Solutions*, 2001; 2001.
38. Theiss, M.; Gross, J., Nonprimitive Model Electrolyte Solutions: Comprehensive Data from Monte Carlo Simulations. *J Chem Eng Data* **2020,** 65, 634-639.
39. Smith, W. R.; Nezbeda, I.; Kolafa, J.; Moucka, F., Recent progress in the molecular simulation of thermodynamic properties of aqueous electrolyte solutions. *Fluid Phase Equilib* **2018,** 466, 19-30.
40. 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.
41. Bang, C.-H.; Choi, H.-K.; Lee, B.-S., Modeling of activity coefficients of amino acid and electrolyte in aqueous solutions. *J Mol Liq* **2016,** 223, 1-9.
42. Conway, O.; An, Y.; Bejagam, K. K.; Deshmukh, S. A., Development of transferable coarse-grained models of amino acids. *Mol Sys. Des Eng* **2020,** 5, 675-685.
43. Dasetty, S.; Barrows, J. K.; Sarupria, S., Adsorption of amino acids on graphene: assessment of current force fields. *Soft Matter* **2019,** 15, 2359-2372.
44. Jin, Z.; Tang, Y.; Wu, J., A perturbative density functional theory for square-well fluids. *J Chem Phys* **2011,** 134, 174702.
45. Mansoori, G. A.; Carnahan, N. F.; Starling, K. E.; Leland, T. W., Equilibrium Thermodynamic Properties of the Mixture of Hard Spheres. *J Chem Phys* **1971,** 54, 1523-1525.
46. Robinson, R. A.; Stokes, R. H., Tables of osmotic and activity coefficients of electrolytes in aqueous solution at 25° C. *J Chem Soc Faraday Trans* **1949,** 45, 612-624.
47. Zhao, Y. H.; Abraham, M. H.; Zissimos, A. M., Fast Calculation of van der Waals Volume as a Sum of Atomic and Bond Contributions and Its Application to Drug Compounds. *J Phys Org Chem* **2003,** 68, 7368-7373.
48. Edward, J. T.; Farrell, P. G., Relation between van der Waals and Partial Molal Volumes of Organic Molecules in Water. *Can J Chem* **1975,** 53, 2965-2970.
49. Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C.; Wilcox, W., Hydrophobic moments and protein structure. *Faraday Symp Chem Soc* **1982,** 17, 109-120.
50. Lee, B. S.; Kim, K. C., Measurements and Modeling of the Activity Coefficients and Solubilities of L-alanine in Aqueous Electrolyte Solutions. *Korean J Chem Eng* **2010,** 48, 519-533.
51. Khoshkbarchi, M. K.; Soto-Campos, A. M.; Vera, J. H., Interactions of DL-serine and L-serine with NaCl and KCl in aqueous solutions. *J Solution Chem* **1997,** 26, 941-955.

52. Soto-Campos, A. M.; Khoshkbarchi, M. K.; Vera, J. H., Interaction of DL-threonine with NaCl and NaNO3 in aqueous solutions: e.m.f. measurements with ion-selective electrodes. *J Chem Thermodyn* **1997,** 29, 609-622.
53. Bonner, O. D., Osmotic and activity coefficients of sodium and potassium glutamate at 298.15 K. *J Chem Eng Data* **1981,** 26, 147-148.
54. Wyman, J.; McMeekin, T. L., The Dielectric Constant of Solutions of Amino Acids and Peptides. *J Am Chem Soc* **1933,** 55, 908-914.
55. Rao, M. V. R.; Atreyi, M.; Rajeswari, M. R., Estimation of partial molar volumes of α-amino acids in water. *J Chem Soc Faraday Trans* **1984,** 80, 2027-2032.
56. Rao, M. V. R.; Atreyi, M.; Rajeswari, M. R., Partial molar volumes of .alpha.-amino acids with ionogenic side chains in water. *J Phys Chem* **1984,** 88, 3129-3131.
57. Mishra, A. K.; Ahluwalia, J. C., Apparent molal volumes of amino acids, N-acetylamino acids, and peptides in aqueous solutions. *J Phys Chem* **1984,** 88, 86-92.

For Table of Contents Only