

McMillan–Mayer solution thermodynamics for a protein in a mixed solvent

R.A. Curtis^{a,b,*}, J. Newman^{a,b}, H.W. Blanch^{a,b}, J.M. Prausnitz^{a,b}

^a *Chemical Engineering Department, University of California, Berkeley, CA 94720, USA*

^b *Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

Received 17 October 2000; accepted 10 September 2001

Abstract

McMillan–Mayer solution theory is used to calculate the thermodynamic properties of solute/solvent systems from the set of solvent-averaged interactions between solute particles. To match experimental data, the independent variables of the McMillan–Mayer framework (temperature, the set of solvent chemical potentials, and the set of solute concentrations) need to be converted to the Gibbs framework, where the independent variables are temperature, pressure, and the set of concentrations for all components. Previous conversions have been restricted to systems with a single solvent and one or more solutes. However, in the analysis of light-scattering data, it is sometimes necessary to consider multiple solvents. Here, we present the thermodynamics of McMillan–Mayer solution theory for a single protein solute and a mixed solvent of aqueous salt; we show that protein–salt preferential-interaction parameters can be readily determined from data reduction within the McMillan–Mayer framework. We provide a sample calculation of a liquid–liquid phase separation for protein solutions, wherein we convert the phase diagram from the McMillan–Mayer framework to the Gibbs framework. For the coexistence curve of the McMillan–Mayer phase diagram, the salt molality of a hypothetical outside solution is held constant, whereas for the coexistence curve for the Gibbs phase diagram, the salt molality of the light phase is held constant. The difference between the curves is determined by the preferential-interaction parameter. The two curves are identical only in the limiting case where the preferential-interaction parameter is zero, the solvent can be considered as a pseudo-one-component solvent. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Liquid–liquid equilibria; Aqueous salt solutions; Potential of mean force model

1. Introduction

Salt-induced protein precipitation and salt-induced protein crystallization are extensively used separation methods in biochemistry and biotechnology. Salt-induced protein precipitation is commonly used as a first step in the purification of proteins because of its selectivity and low cost, and salt-induced protein

* Corresponding author. Fax: +1-510-643-1228.

E-mail address: rcurtis3955@yahoo.com (R.A. Curtis).

crystallization is often used for obtaining protein crystals for X-ray diffraction studies. Optimizing these processes requires knowledge of the protein–aqueous–electrolyte phase diagrams as a function of the solution conditions, such as salt type, salt concentration, temperature, and pH. The thermodynamics of protein solutions depends crucially on the intermolecular interactions between proteins, salt ions, and water. Because these interactions are not well understood, physically realistic simplifying assumptions are needed to determine the phase diagrams.

When there are large differences in the concentrations of the components in a solution, thermodynamic properties are commonly calculated from statistical–mechanical models, where the inputs are solute–solute interactions averaged over positions of solvent molecules, where the solutes are the dilute components and the solvents are the concentrated components. The averaging process reduces the order of the effective Hamiltonian such that calculations can be made assuming a pseudo-one-component system. Models based on effective solute–solute interactions are in the framework of McMillan–Mayer solution theory, where the n -body solvent-averaged effective interaction plays the same role as the n -body potential does in imperfect-gas theory [1]. Accordingly, the theories used to calculate thermodynamic properties of gases can be used to calculate the thermodynamic properties of liquid solutions where solute molecules interact through the solvent, provided that the n -body interaction energies are replaced with the n -body potentials of mean force.

For a single protein dissolved in an aqueous electrolyte solution, because typically the salt and water are each at a much greater concentration than the protein, we consider the salt and water as a mixed solvent; the protein is the solute. However, determining the effective protein–protein interactions from Boltzmann-averaging over positions of salt and water is beyond the scope of statistical mechanics. The advantage of using McMillan–Mayer theory is that the effective protein–protein interactions can be measured using a variety of techniques. Consequently, most theoretical work on protein–solution phase diagrams is based on a simplified form of the protein–protein interaction potential function which is fit to experimental results from light scattering or osmometry [2–8].

Statistical mechanics provides a method for calculating a generating function from which all other equilibrium properties can be derived using standard thermodynamic calculations. However, the thermodynamic framework for calculating properties from the generating function must be chosen such that the independent variables of the framework match those of the statistical–mechanical ensemble employed by the calculations. For McMillan–Mayer solution theory, the independent variables are the set of solvent chemical potentials, solute concentration, and temperature. However, most experimental properties are measured in the Gibbs framework where the independent variables are the set of molalities (other than that of the principal solvent), temperature, and pressure. Thus, the results of a theoretical model must be

interaction is determined from comparing results from the pseudo-solvent framework with those from a mixed-solvent framework. As an example, Vrij and Overbeek [13] have studied charged colloidal particles in salt solutions using light scattering. They showed that the measured molecular weights need to be corrected by factors of 1.10–1.45, if the aqueous salt solution is considered as a pseudo-one-component instead of a mixed solvent. From the difference, the authors determined the amount of negative adsorption of salt on the colloidal particles.

Here, we describe a similar system containing one protein solute and a binary solvent of salt and water. While we present details for this specific ternary system, the results are readily generalized for any number of solutes dissolved in a multi-component solvent. Casassa and Eisenberg [14] have presented a thermodynamic analysis for a solute in a multi-component solvent for the Gibbs framework. Here, we show that protein–salt preferential interactions are directly determinable from data analyzed within the McMillan–Mayer framework. We then use results for the protein–salt preferential interaction to convert the independent variables of the McMillan–Mayer framework to those of a modified Gibbs framework where the independent variables are temperature, pressure, salt molality, and protein concentration. This conversion allows us to generate protein phase diagrams at constant pressure and constant salt molality, in contrast to the McMillan–Mayer framework, where all phase diagrams are determined as functions of the less well-defined set of solvent chemical potentials.

The first section summarizes the essentials of McMillan–Mayer solution theory. The theory of preferential interactions is given in the following section within the McMillan–Mayer framework. In the final section, to illustrate our discussion, we calculate the liquid–liquid phase diagram for protein solutions in the modified Gibbs framework and compare it with that in the McMillan–Mayer framework.

2. Essentials of McMillan–Mayer solution theory

Consider the system shown in Fig. 1 where a semi-permeable membrane separates an inside solution containing water (1); protein (2); and salt (3) and an outside solution that contains only solvent molecules,

“Outside”	“Inside”
1 - water 3 - salt μ_1, μ_3, T $m_{3,o}$ p_o	1 - water 2 - protein 3 - salt μ_1, μ_3, c_2, T $\mu_2, m_{3,i}$ $p_i = p_o + \Pi$

Fig. 1. McMillan–Mayer solution theory describes the properties of the inside solution in terms of an outside solution that is at the same temperature and chemical potential of solvent components (salt and water) denoted by μ_1 and μ_3 . Here, p is pressure, T temperature, μ_3 salt molality, and Π is osmotic pressure. Subscripts o and i refer to outside and inside solutions, respectively. Quantities, $m_{3,i}$, Π , and μ_2 are determined from a model whose inputs are protein concentration, c_2 , and the properties of the outside solution, either (μ_1, μ_3, T) or $(p_o, m_{3,o}, T)$.

salt and water. The activities of the solvent molecules are the same for the inside and the outside solutions, because they are permeable to the membrane. The thermodynamic properties of the outside solution are determined by three intensive properties chosen to be either the set of temperature, T , and chemical potentials of the solvent components, μ_1 and μ_3 , or the set of temperature, outside-solution pressure, p_o , and outside-solution salt molality, $m_{3,o}$. In McMillan–Mayer solution theory, the properties of the inside solution are determined from the protein concentration, c_2 , and the condition of osmotic equilibrium, which relates the thermodynamics of the inside solution to that of the reference or outside solution. Thus the independent variables in the McMillan–Mayer framework are T , μ_1 , μ_3 , and c_2 or equivalently T , p_o , $m_{3,o}$, and c_2 .

Thermodynamic properties of the inside solution are determined from the generating function that is the energy whose independent variables are T , μ_1 , μ_3 , and c_2 , given by the McMillan–Mayer free energy, F . F is given by the following transform of the Helmholtz free energy, A

$$F = A + p_o V - \mu_1 N_1 - \mu_3 N_3 \quad (1)$$

where V is the volume of the solution and N_i is the mole number of component i . Plugging the definition of A into Eq. (1) relates F to the osmotic pressure, Π , and the solute chemical potential, μ_2 , by

$$F = -\Pi V + \mu_2 N_2 \quad (2)$$

All other thermodynamic properties are derived from the differential of F , obtained from the total differential of Eq. (1)

$$dF = -\Pi dV + V dp_o - S dT - N_1 d\mu_1 + \mu_2 dN_2 - N_3 d\mu_3 \quad (3)$$

where S is the entropy of the solution. Eq. (3) implies that there are six independent variables, although only 5 d.f. are necessary to specify an extensive variable for a single phase with three components. The extra variable results from the Legendre transform of Eq. (1) with respect to $p_o V$. This transform is introduced to obtain conjugate variables for F that are referenced with respect to the outside solution as shown below; dp_o can be eliminated from Eq. (3) by using the Gibbs–Duhem relation for the outside solution

$$V_o dp_o = S_o dT + N_{1,o} d\mu_1 + N_{3,o} d\mu_3 \quad (4)$$

where subscript o refers to the outside solution, introduced to distinguish properties of the outside solution from those of the inside solution, denoted by subscript i. Introducing Eq. (4) into Eq. (3) and relating the extensive McMillan–Mayer free energy to the intensive form, F/V , gives

$$d\left(\frac{F}{V}\right) = \mu_2 dc_2 - (c_{1,i} - c_{1,o})d\mu_1 - (c_{3,i} - c_{3,o})d\mu_3 - (s_i - s_o)dT \quad (5)$$

where c_j refers to molar concentration of component j and s refers to entropy per unit volume.

From a model for the McMillan–Mayer free energy, all other thermodynamic properties of interest can be derived from the appropriate manipulations of Eqs. (2) and (5). For example, the protein chemical potential is given by the partial differential of Eq. (5)

$$\mu_2 = \left(\frac{\partial F/V}{\partial c_2}\right)_{T, \mu_1, \mu_3} \quad (6)$$

The pressure of the inside solution is given by the sum of the outside-solution pressure and the osmotic pressure, where the osmotic pressure is calculated from Eqs. (2) and (6) as

$$\Pi = -\frac{F}{V} + \mu_2 c_2 \quad (7)$$

An additional thermodynamic quantity of interest is the concentration of salt in the inside solution. This can be calculated from the a partial differential of Eq. (5)

$$c_{3,i} - c_{3,o} = -\left(\frac{\partial F/V}{\partial \mu_3}\right)_{T, \mu_1, c_2} \quad (8)$$

Models for the McMillan–Mayer free energy F are based on statistical mechanics where there is a direct analogy between a system of gas molecules interacting through a vacuum and a system of solute particles interacting through a continuous solvent. The residual Hemholtz energy, A^{res} , for a pure gas in terms of the set of n -body interaction potentials and gas density is the same function as the excess McMillan–Mayer energy function for a solute in a solvent in terms of the set of n -body potentials of mean force and the solute density. This mapping is symbolized by

$$A^{\text{res}}(c, T) \rightarrow F^{\text{ex}}(c_2, T, \mu_1, \mu_3) \quad (9)$$

where the left-hand side of Eq. (9) refers to the pure gas and the right-hand side refers to the solute/mixed-solvent system. A residual property, denoted by superscript *res*, is the difference of the real-system property and that of an ideal gas at the same temperature and density. An excess property, denoted by *ex*, is the difference of the real-system property and the ideal property evaluated at the same temperature, solute concentration, and the set of solvent chemical potentials; an ideal solution is a solution, where the solute molecules do not interact as discussed below. Because the solute/solvent system has two more degrees of freedom than those of the pure gas, the set of solvent chemical potentials need to be included as independent variables in the calculation of F^{ex} . These variables are contained in the description of the set of n -body potentials of mean force, $\{W^{(n)}\}$. The gradients of these potentials give the forces between n -solute particles with fixed positions averaged over positions of all solvent molecules. Because the averaging process involves only the solvent molecules, the potential of mean force is determined by the state of the outside solution or the variables (T, μ_1, μ_3) or the variables $(T, p_o, m_{3,o})$.

Eq. (9) provides the route for calculating the excess McMillan–Mayer free energy by using the same statistical–mechanical theories that are used for calculating properties of pure gases. Generally, in theories for calculating the thermodynamics of pure gases, the n -body interaction potentials are approximated by the sum of the two-body interactions, referred to as pairwise additivity. The corresponding approximation in McMillan–Mayer theory is known as the superposition approximation. With this approximation, the excess McMillan–Mayer free energy is a function of temperature, protein concentration, and the two-body potential of mean force.

The total McMillan–Mayer free energy is given by the sum of the ideal and excess contributions

$$F = F^{\text{id}} + F^{\text{ex}} \quad (10)$$

where an ideal solution is denoted by superscript *id*. An ideal McMillan–Mayer solution is analogous to an ideal gas. The expression equivalent to the ideal-gas law is given by the Van't Hoff equation

$$\Pi^{\text{id}} = c_2 RT \quad (11)$$

and the ideal McMillan–Mayer solute chemical potential is

$$\mu_2 = RT \ln \frac{\lambda_2^3}{q_2} + RT \ln c_2 + RT \ln \gamma_2^\infty \quad (12a)$$

where λ_2 is the thermal wavelength of the protein, q_2 the internal partition function of the protein molecule, and γ_2^∞ is the protein infinite-dilution activity coefficient. The sum of the first two terms on the left side of Eq. (12a) are analogous to the chemical potential of an ideal gas if the internal partition function of the protein molecule in the solvent can be approximated by that in an ideal-gas state. With this approximation, it follows that the third term on the right side of Eq. (12a) is equal to the free energy of transferring a protein molecule from an ideal gas state into the solvent. This free energy of transfer is referred to as the solvation free energy, G_s

$$G_s = RT \ln \gamma_2^\infty \quad (12b)$$

and is only a function of the thermodynamic state of the outside solution. Generally, the protein infinite-dilution standard state, μ_2^θ , is defined by

$$\mu_2^\theta = RT \ln \frac{\lambda_2^3 \gamma_2^\infty}{q_2} \quad (13)$$

The McMillan–Mayer free energy of an ideal solution is obtained by substituting Eqs. (11)–(13) into Eq. (2) to give

$$\frac{F^{\text{id}}}{V} = c_2 \mu_2^\theta + c_2 RT [\ln c_2 - 1] \quad (14)$$

3. Preferential-interaction parameter

Experimental measurements are performed as a function of the independent variables in the Gibbs framework, given by temperature, pressure, and protein and salt concentration. Consequently, for a predictive model, calculated thermodynamic properties in the McMillan–Mayer framework must be converted to give properties in the Gibbs framework. To perform this conversion, the inside-solution properties need to be related to the outside-solution properties. For example, to relate the inside-solution pressure to the outside-solution pressure, we need to determine the osmotic pressure. Similarly, the difference of the inside-solution salt concentration and the outside-solution salt concentration needs to be calculated. This difference has been evaluated previously using variables in the Gibbs framework [14,15]. Below, we evaluate this difference in salt concentration using the McMillan–Mayer framework via Eq. (8) and an appropriate form for the McMillan–Mayer free energy. The details of the calculation are given in the Appendix A, where we also show that the result of the same calculation in the Gibbs framework is equivalent to that in the McMillan–Mayer framework.

This difference in salt concentration, $c_{3,i} - c_{3,o}$, can be expanded in a power series in protein concentration

$$c_{3,i} - c_{3,o} = \sum_{n>0} \Delta c_{3,n} c_2 \quad (15)$$

where the n th-order coefficients, $\Delta c_{3,n}$ are related to the salt deficiency or excess of the n th-order interaction between solvent layers. Thus, the first-order coefficient is related to the amount of salt perturbation due to adding protein molecules in the limit of infinite dilution of protein. In this limit, there is a difference in salt composition between the outside solution and the inside solution due to the excluded volume of the protein molecules and because the region around the protein molecule is at a different salt composition than that of the outside solution. At higher protein concentrations, the regions of solvent perturbation about the protein molecules overlap and higher order terms need to be included in Eq. (15). For instance, the second-order coefficient is related to the average amount of excess or deficiency of salt in the two-body interaction minus that of two non-overlapping solvent layers.

To determine the coefficients of Eq. (15), we use Eq. (8) along with the osmotic virial expansion for the McMillan–Mayer free energy, given by [1]

$$\frac{F}{VRT} = \frac{c_2 \mu_2^\theta}{RT} + c_2 [\ln c_2 - 1] + c_2^2 B_{22} + c_2^3 B_{222} + \dots \quad (16)$$

where B_{22} is the second virial coefficient, B_{222} is the third virial coefficient, etc. The n th virial coefficient is related to a volume integral of the n -body potential of mean force.

Substituting Eq. (16) into Eq. (8) gives the desired expansion in protein concentration for the difference in salt concentration. The first-order coefficient is related to the derivative of the standard-state potential with respect to salt chemical potential, whereas the n th-order coefficients are related to derivatives of the n -body pmf with respect to salt chemical potential. As shown in the Appendix A, the first-order coefficient can be approximated by (see Eq. (A.15))

$$\Delta c_{3,1} = - \left(\frac{\partial \mu_2^\theta}{\partial m_{3,o}} \right)_{T,p_o} \frac{1 - \bar{v}_{3,o} c_{3,o}}{(\partial \mu_3 / \partial m_{3,o})_{T,p_o}} - \bar{v}_2^\infty c_{3,o} \quad (17)$$

where the superscript, ∞ , denotes a property evaluated at infinite dilution of protein. The relation $(\partial \mu_3 / \partial m_{3,o})_{T,p_o}$ can be calculated from an activity-coefficient model for the salt in the outside solution. On the other hand, evaluating $(\partial \mu_2^\theta / \partial m_{3,o})_{T,p_o}$ requires a model to calculate the solvation free energy of proteins in concentrated salt solutions. Generally, the protein solvation free energy is calculated by summing the contributions to the free energy from atomic groups on the protein surface. The solvation free energies of the different atomic groups are determined from salting-out experiments of small model compounds. However, detailed investigations have not yet been performed to show if this extrapolation is reliable. As shown below, we use a model where the effects of the individual atomic groups are averaged-out.

The second-order coefficient is given by

$$\Delta c_{3,2} = \frac{(1 - \bar{v}_{3,o} c_{3,o}) RT}{(\partial \mu_3 / \partial m_{3,o})_{T,p_o}} \left(\frac{\partial B_{22}}{\partial m_{3,o}} \right)_{T,p_o} - c_{3,o} RT \left(\frac{\partial B_{22}}{\partial p_o} \right)_{T,m_{3,o}} \quad (18)$$

where we have derived Eq. (18) following the methods given in the Appendix A. Here, the second-order correction to the difference in salt concentrations is related to derivatives of the protein pair potential of mean force with respect to outside-solution properties.

To illustrate the previous paragraph, we calculate the preferential interaction between lysozyme and sodium chloride up to second-order in protein concentration using Eqs. (17) and (18). In concentrated salt solutions, the solvation free energy of protein molecules has been correlated with the following relation [16]:

$$\left(\frac{\partial \mu_2^\theta}{\partial m_{3,o}}\right)_{T,p_o} = \left(\frac{\partial G_s}{\partial m_{3,o}}\right)_{T,p_o} = -\Lambda_2 + \frac{d\sigma}{dm_3} S_2 \quad (19)$$

The first term on the right-hand side of Eq. (19) is the contribution from favorable electrostatic interactions between the charge on the protein surface and salt ions. For salt molalities greater than 0.5m, this term is proportional to the dipole moment of the protein, Λ_2 . The second term is related to the work to form a cavity in the salt solution. The incremental free energy is proportional to the molal surface-tension increment of the salt, $(d\sigma/dm_3)$, multiplied by the surface area of the protein molecule, S_2 . Because all salts have positive surface-tension increments, this effect always leads to the preferential exclusion of salt in the domain of the protein.

In the calculation of the preferential-interaction parameter, the second-order term is also included in the calculation using the first term on the right of Eq. (18). The derivative of B_{22} with respect to salt molality has been evaluated numerically from B_{22} data for lysozyme in solutions of sodium chloride [17]. All other properties are shown in Table 1.

In Fig. 2, the inside-solution salt molality, calculated with and without the second-order correction term, is plotted versus protein concentration for an outside solution salt molality of 1.0. For all protein concentrations, the salt is excluded from the inside solution. This is typical of most salt solutions at molalities greater than 0.5 where salting-out behavior is observed. Generally, the unfavorable surface tension-increment effect is greater than the favorable electrostatic interaction between the protein dipole and the ion atmosphere. This results in an increase in the infinite-dilution activity coefficient and a reduction in protein solubility. The second-order correction is not significant at protein concentrations less than 100 g/l; at higher protein concentrations the preferential exclusion of the salt decreases because the pmf is more attractive with the addition of salt. However, at protein concentrations greater than 100 g/l, it is likely that higher-order interactions are significant. Consequently, truncating the virial expansion at second-order may introduce significant error. The inset of Fig. 2 shows that the preferential-interaction parameter is proportional to the outside-solution salt molality for a protein molality of 0.005m (~ 65 g/l). This proportionality is a consequence of the law of mass action.

Table 1

Thermodynamic properties required to calculate preferential-interaction parameter^a

Property	Reference
Partial molar volumes	[25]
Activity coefficient and osmotic coefficient	[26]
Surface tension	[27]

^a All properties refer to aqueous solutions of sodium chloride calculated as a function of salt molality and temperature.

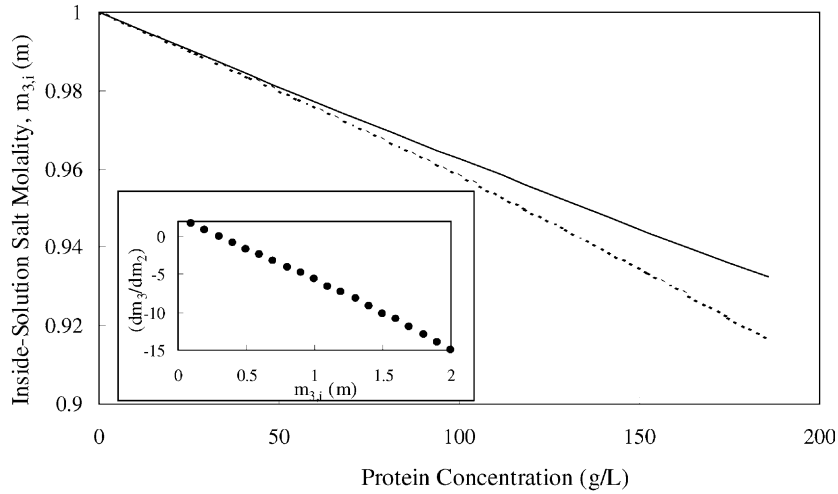


Fig. 2. Plot of the inside-solution salt molality vs. protein concentration for a hypothetical outside-solution salt molality of 1.0m. The dashed line refers to the calculation to first-order in protein concentration and the solid line refers to the calculation to second-order in protein concentration. In the inset, the preferential-interaction parameter is plotted vs. salt molality for a protein molality of 0.005m.

4. Phase-equilibrium calculations

4.1. McMillan–Mayer phase equilibrium

Consider the case of determining the liquid–liquid equilibrium for the system of protein dissolved in an aqueous electrolyte solution (the liquid–liquid equilibrium for solute in a solvent is analogous to the liquid–vapor equilibrium of a pure substance). The light liquid phase (l) and the dense liquid phase (d) have the same outside solution due to the equilibrium condition that each solvent species has the same chemical potential in all phases. The choices of the independent variables are made from the set of thermodynamic variables corresponding to the framework. The natural intensive thermodynamic variables in the McMillan–Mayer framework for the two-phase system are T , c_2^l , c_2^d , p_o , and $m_{3,o}$. According to the phase rule, there are three intensive degrees of freedom. Here, the properties of the outside solution, T , p_o , and $m_{3,o}$ are the independent variables; c_2^l and c_2^d are calculated from the two equilibrium conditions

$$p_o + \Pi^l = p_o + \Pi^d \tag{20}$$

and

$$\mu_2^l = \mu_2^d \tag{21}$$

Because both phases have the same outside solutions, the outside-solution pressure cancels in Eq. (20) that becomes

$$\Pi^l = \Pi^d \tag{22}$$

The osmotic pressure and protein chemical potential are determined from a model whose inputs are the two-body potential of mean force, the protein concentration, and temperature. The potential of mean

force is the same for both phases, because it is a function of the outside-solution variables, T , p_o , and $m_{3,o}$.

On a coexistence curve, one of the independent variables varies and the other independent variables are constant. In this illustration, we generate cloud-point (temperature) curves where coexistence temperature is plotted versus the light and dense-phase protein concentrations. In the McMillan–Mayer framework, the outside-solution pressure and the outside-solution salt molality are constant, whereas the pressure and the salt molality of the equilibrated phases vary along the cloud-point curve. The pressure is given by the sum of the outside-solution pressure and the osmotic pressure and the salt molality is determined from the preferential-interaction parameter that can be calculated from Eqs. (15) and (17)–(19).

When salt and water are considered as a pseudo-solvent, it is implicitly (and erroneously) assumed that the preferential-interaction parameter expressed in molality units is zero, i.e. the outside-solution salt molality is equal to the inside-solution salt molality. Because the preferential-interaction parameter analysis is not used in the determination of the cloud-point curve, this curve is independent of the treatment of the solvent as long as we choose the outside-solution salt molality of the mixed-solvent system equal to that of the pseudo-solvent. However, the pseudo-solvent condition gives the incorrect result that the salt molality in the light phase is equal to that in the dense phase.

4.2. Phase equilibrium in the Gibbs framework

Cloud-point temperature curves are generated by cooling protein solutions at constant pressure until the onset of phase separation as observed by a discontinuous change in the turbidity of the solution. Because only a differential amount of the dense phase is formed, the salt molality in the light phase is given by the salt molality of the protein solution before phase separation. Thus to match experimental data, it is desirable to generate coexistence curves at constant pressure and at constant salt molality of the light phase instead of generating curves at constant properties of the hypothetical outside solution. To perform calculations in this experimental framework, we first use McMillan–Mayer solution theory to determine the thermodynamic properties from the two-body potential of mean force. Because the potential of mean force is a function of the properties of the hypothetical outside solution, T , p_o , and $m_{3,o}$, we then need to relate the variables of the McMillan–Mayer framework to those of the modified Gibbs framework. The properties of the outside solution also determine the protein infinite-dilution standard state and the chemical potentials of the solvent components that are also required for calculation of phase equilibria in the Gibbs framework.

We need the outside pressure, p_o , and the outside-solution salt molality, $m_{3,o}$, as a function of T , p_i , c_2 , $m_{3,i}$. The preferential-interaction parameter relates the difference in salt concentration of the inside and of the outside solution to the McMillan–Mayer free energy

$$c_{3,i} - c_{3,o} = - \left(\frac{\partial F/V}{\partial \mu_3} \right)_{T, \mu_1, c_2} \sim | [c_2, T, W^{(2)}(T, p_o, m_{3,o}), \mu_2^\theta(T, p_o, m_{3,o})] \quad (23)$$

where $|$ denotes a function of those variables contained in the brackets and where the salt concentrations are given by

$$c_{3,o} = \frac{m_{3,o}}{V_{m,o}(T, p_o, m_{3,o})} \quad (24)$$

and

$$c_{3,i} = \frac{m_{3,i}}{V_{m,i}(T, p_i, c_2, m_{3,i})} \quad (25)$$

The outside-solution pressure can be determined from the osmotic pressure given the pressure of the inside solution

$$p_o = p_i - \Pi \quad (26)$$

where

$$\Pi \sim | [c_2, T, W^{(2)}(T, p_o, m_{3,o})] \quad (27)$$

Relations given by Eqs. (23) and (27) provide two implicit relations in $m_{3,o}$ and p_o which can be obtained in terms of the independent variables of the inside solution, T, p_i, c_2 , and $m_{3,i}$.

In this phase equilibrium there are 3 d.f. The temperature, pressure, and salt molality of the light phase, $m_{3,i}^l$, are a convenient set of independent variables. The salt molality in the dense phase, $m_{3,i}^d$ and the protein concentrations in the light and dense phases, c_2^l and c_2^d are then determined from the equilibrium relations

$$\mu_1^l = \mu_1^d \quad (28)$$

$$\mu_2^l = \mu_2^d \quad (29)$$

$$\mu_3^l = \mu_3^d \quad (30)$$

Because the chemical potentials of the solvent components in the light phase are equal to those in the dense phase, these phases have the same outside solution. Thus, we can replace these equilibrium criteria (Eqs. (28) and (30)) with the following:

$$m_{3,o}^l = m_{3,o}^d \quad (31)$$

and

$$p_o^l = p_o^d \quad (32)$$

Eqs. (31) and (32) provide a more direct method for equating the chemical potentials of the salt and of the water avoiding the additional calculation of the salt and water chemical potentials. Because both phases are also at the same pressure, Eq. (32) is also satisfied by

$$\Pi^l = \Pi^d \quad (33)$$

In the Gibbs framework, the salt molality of the light phase and the pressure remain constant along the cloud-point curve and the properties of the outside solution vary as a function of temperature.

For the pseudo-one-component solvent case in the Gibbs framework, because the salt molality of the light phase is equal to that in the dense phase, the phase-equilibrium criteria reduce to Eqs. (29) and (33), which need to be solved for the light and dense-phase protein concentrations. These conditions are identical to the McMillan–Mayer equilibrium conditions. The difference in the two systems of equations is that the outside-solution pressure varies along a Gibbs cloud-point curve, in contrast to the McMillan–Mayer cloud-point curve, where the outside-solution pressure is held constant. The

difference in the outside-solution pressure of the Gibbs framework and that of the McMillan–Mayer is given by the osmotic pressure of the solution that is on the order of 0.01 atm for typical protein solutions. Because it is unlikely that the potential of mean force changes significantly over this range, we expect that the cloud-point curve in the Gibbs framework is similar to that of McMillan–Mayer for the case of a single-component pseudo-solvent and consequently also similar to the McMillan–Mayer cloud-point curve for the case of a mixed-solvent at the same outside-solution salt molality as the pseudo-solvent.

If the protein–salt preferential interaction is strong, there will be a large difference in the salt molality of the dense phase and that of the light phase. In this case, the salt is treated as a separate component and we need the additional phase-equilibrium criterion of Eq. (31) to solve for the salt molality in the dense phase given that in the light phase. In the modified Gibbs framework, we can generate a phase diagram at constant pressure and salt molality of the light phase. Here the outside-solution salt molality and the outside-solution pressure vary along the cloud-point curve in contrast to the McMillan–Mayer curve where the outside-solution properties are held constant. As mentioned in the preceding paragraph, we do not expect the small changes in outside-solution pressure to alter the shape of the cloud-point curve. In this study, we determine the effect of varying the outside-solution salt molality along the Gibbs cloud-point curve versus holding this variable constant along that of the McMillan–Mayer framework.

To simplify the computation, we ignore the dependence of the potential of mean force and the protein infinite-dilution standard state on the outside-solution pressure. Because osmotic pressures of protein solutions are on the order of 0.01 atm, we expect that the potential of mean force and the protein infinite-dilution standard state can be approximated by their values evaluated at the pressure of the inside solution, set to 1.0 atm.

4.3. Phase-equilibrium results

To illustrate our discussion, we fit the liquid–liquid equilibrium data for lysozyme in sodium-chloride solutions using the McMillan–Mayer and the Gibbs frameworks. We use a simple statistical–mechanical model [18,19] to calculate the thermodynamic properties of the McMillan–Mayer phases. Here, the protein–protein interaction in the light and dense fluid phases is given by the same two-body potential of mean force. This two-body pmf is given by a hard-sphere term and a power-law attractive interaction:

$$W^{(2)}(r) = \begin{cases} \infty & \text{for } r \leq d_2 \\ -\varepsilon \left(\frac{r}{d_2}\right)^{-n} & \text{for } r > d_2 \end{cases} \quad (34)$$

where d_2 is the hard-sphere protein diameter, r center-to-center separation, n the power-law exponent and ε is the interaction strength at contact. The random-phase approximation is used to calculate the thermodynamic properties from the power-law potential. The RPA is a first-order perturbation theory for the McMillan–Mayer free energy (or the analogous Helmholtz energy for a pure gas). The reference system is a fluid of hard spheres and the attractive power-law potential gives the perturbation. The McMillan–Mayer free energy is given by

$$\frac{F}{V} = c_2 \mu_2^\theta + c_2 RT [\ln c_2 - 1] + RT \frac{4\eta - 3\eta^2}{(1 - \eta)^2} + \frac{1}{2} N_{\text{av}}^2 c_2^2 \int_{d_2}^{\infty} 4\pi r^2 W_{\text{pert}}^{(2)}(r) dr \quad (35)$$

where η is the hard-sphere packing fraction ($\pi d_2^3 \rho_2 / 6$), ρ_2 protein number density, R the gas constant, N_{av} Avogadro's number, and $W_{\text{pert}}^{(2)}$ refers to the attractive power-law contribution to $W^{(2)}$. The first two terms on the right side of Eq. (35) give the ideal contribution to the McMillan–Mayer free energy. The contribution of the hard-sphere reference system is calculated from the Carnahan–Starling equation of state [20], given by the third term on the right-hand side of Eq. (35). The last term includes the contribution of the attractive power-law potential to the McMillan–Mayer free energy.

The protein chemical potential is obtained from substituting Eq. (35) into Eq. (6). Similarly, the osmotic pressure is obtained from the protein chemical potential using Eq. (7). After substituting in the attractive power-law potential, we have the following expressions

$$\mu_2 = \mu_2^\theta + RT \ln c_2 + RT \frac{8\eta - 9\eta^2 + 3\eta^3}{(1 - \eta)^3} - 2\pi R c_2 N_{\text{av}} \frac{d_2^3}{n - 3} \frac{\varepsilon}{k_B} \quad (36)$$

and

$$\frac{\Pi}{c_2 RT} = \frac{1 + \eta + \eta^2 - \eta^3}{(1 - \eta)^3} - 2\pi c_2 N_{\text{av}} \frac{d_2^3}{n - 3} \frac{\varepsilon}{k_B T} \quad (37)$$

where we use Eq. (13) for μ_2^θ and k_B is Boltzmann's constant. The infinite-dilution activity coefficient is determined from the solvation free energy of the protein. Here the protein solvation free energy is given by the integrated form of Eq. (19), where the lower limit of integration gives the solvation free energy of the protein in salt-free water. This term cancels out in the phase-equilibrium calculations given here.

The interaction strength, ε is taken from Fornaseiro et al. [19] for fitting the phase-equilibrium data for lysozyme in solutions of sodium chloride where the salt molality ranges from 0.8 to 1.3m [21]

$$\frac{\varepsilon}{k_B} = 750 + 120m_{3,o} \quad (38)$$

with n equal to 6. The calculated cloud-point curves are shown in Fig. 3 along with the experimental data for lysozyme in solutions of 7% w/v ($\sim 1.3\text{m}$) NaCl [21]. The dashed line refers to the McMillan–Mayer cloud-point curve; the outside-solution salt molality is held constant at 1.0m on this curve. The solid line refers to the Gibbs cloud-point curve; the salt molality in the light phase is set to 1.0m on this curve.

The main purpose of this work is to contrast the cloud-point curves generated in the McMillan–Mayer framework and those generated in the Gibbs framework. To simplify the calculation, we have used a simplified form of the two-body potential of mean force and a statistical–mechanical model that neglects higher-body interactions. For this reason, the model does not describe the experimental data accurately; however, more sophisticated (and more realistic) statistical–mechanical models can be used to achieve better agreement with the data.

The salt molality in the light and dense phases at coexistence depends on the outside-solution salt molality which is different for the coexistence calculation in the Gibbs framework and for that in the McMillan–Mayer framework. This is demonstrated by Fig. 4 that shows the coexistence temperature versus inside and outside-solution salt molality. Because the salt is excluded in the domain of the protein, the outside solution is always at a higher salt molality than that of the light and dense phases. In the calculation performed in the Gibbs framework, the light-phase salt molality is constant. Thus the outside-solution salt molality increases with rising light-phase protein concentration (or rising temperature) because the amount of preferential exclusion of salt is proportional to protein concentration. This is in contrast to the McMillan–Mayer framework where the outside-solution salt molality is constant.

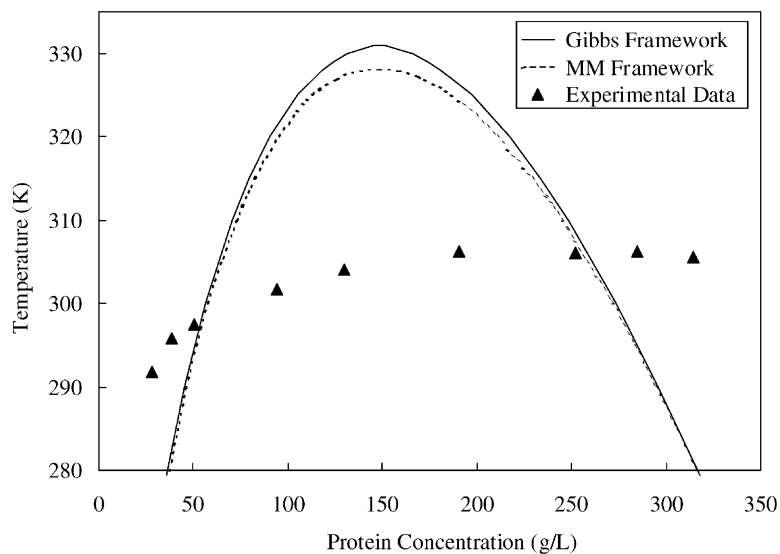


Fig. 3. Cloud-point temperature curves correlated using the Gibbs framework or the McMillan–Mayer framework with the random phase approximation and an attractive power-law potential. In the McMillan–Mayer cloud-point curve, the outside-solution salt molality is held constant at 1.0m, whereas the salt molality of the light phase is held constant at 1.0m in the Gibbs cloud-point curve. The experimental data for lysozyme in 7% w/v NaCl solution are from Muschol and Rosenberger [21].

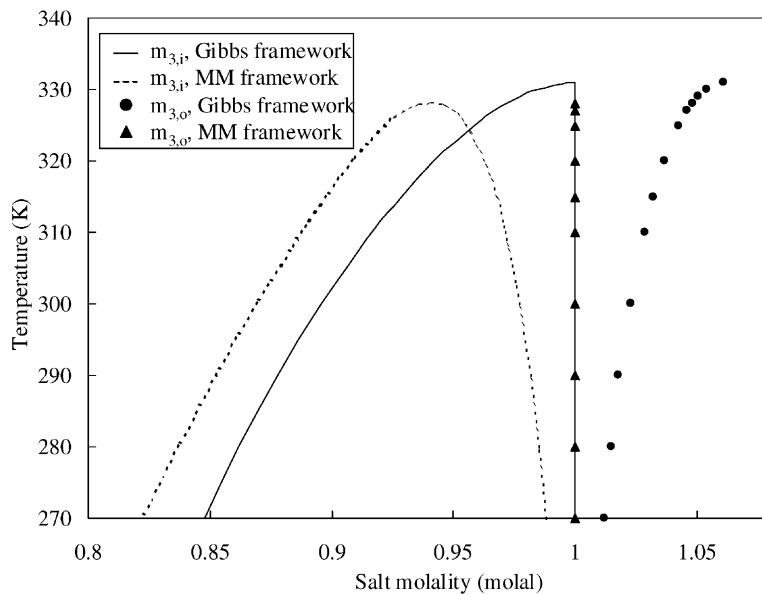


Fig. 4. Calculations of the coexistence temperature vs. inside- or outside-solution salt molality for the Gibbs framework and for the McMillan–Mayer framework.

At a given coexistence temperature, calculation of the protein concentration in the light phase and that in the dense phase is independent of the framework. The difference is that the cloud-point curves are evaluated at different properties of the outside solution. Because the effect of pressure is negligible, we are only concerned with the dependence of the phase diagram on the outside-solution salt molality. The length of the tie line joining the light and dense-phase protein concentrations increases with rising protein–protein attraction. Here, at a given temperature, the light and dense-phase protein concentrations of the Gibbs cloud-point curve are evaluated at a higher outside-solution salt molality than those for the McMillan–Mayer cloud-point curve. Because the protein–protein attraction is proportional to the outside-solution salt molality, the phase diagram is wider in the Gibbs framework relative to the McMillan–Mayer framework. The largest difference between the two cloud-point curves occurs near the critical point, where the outside-solution salt molality of the Gibbs cloud-point curve is a maximum.

Fig. 5 compares the cloud-point curves generated for a set of different salt molalities. As salt molality rises from 1.0 to 2.0m, the protein–protein attraction is enhanced and the solutions become less stable giving precipitation at higher temperatures. Here, the dashed lines refer to the McMillan–Mayer cloud-point curves evaluated at a constant outside-solution salt molality equal to the salt molality of the light phase on the Gibbs cloud-point curves denoted by the solid lines. Because of this equality, the difference in the outside-solution salt molalities of each set of curves is given by the preferential exclusion of the salt from the light-phase of the Gibbs cloud-point curve. The amount of preferential exclusion of salt from the light phase is proportional to the preferential-interaction parameter multiplied by the protein concentration of the light phase. As shown in the inset of Fig. 2, the preferential-interaction parameter

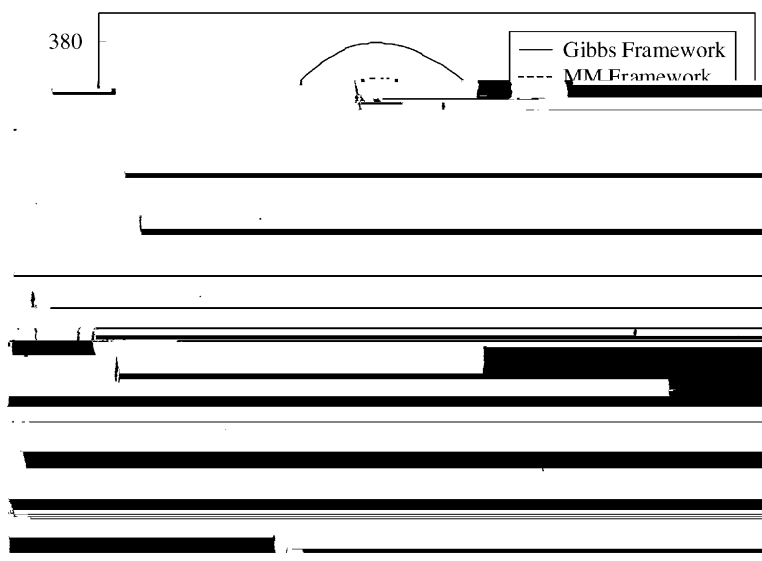


Fig. 5. Cloud-point curves from the Gibbs framework and from the McMillan–Mayer framework for three salt molalities. The light-phase salt molality is constant along the Gibbs cloud-point curve, whereas the outside-solution salt molality is constant along the McMillan–Mayer cloud-point curve. The McMillan–Mayer cloud-point curves are similar to the Gibbs cloud-point curves for a pseudo-one-component solvent.

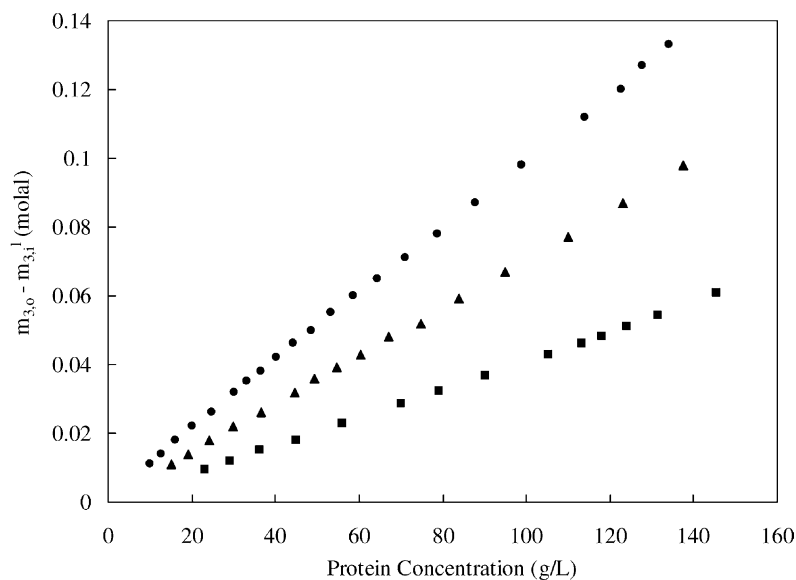


Fig. 6. Difference in the outside-solution salt molality and that of the light phase for the Gibbs phase diagrams in Fig. 5, where the symbols correspond to (■) 1.0m; (▲) 1.5m; and (●) 2.0m, respectively.

is also proportional to the salt molality of the solution at a given protein concentration. Consequently, as salt molality rises from 1.0 to 2.0m, the difference in the corresponding cloud-point curves is also magnified, as shown in Fig. 6, where the outside-solution salt molality minus light-phase salt molality is plotted versus light-phase protein concentration for the Gibbs phase diagram.

5. Conclusions

In the calculation of cloud-point curves, it is usually assumed that the aqueous salt solution can be considered to be a pseudo-one-component solvent; in that case, the salt molality of the light phase is identical to the salt molality of the hypothetical outside solution that determines the potential of mean force. We have shown that the error introduced by this approximation is determined by the magnitude of the preferential-interaction parameter. The preferential-interaction parameter is largest for conditions favorable for the salting-out of proteins: concentrated salt solutions and solutions of kosmotropic salts and for solutions of concentrated protein. As a result, the error introduced by the pseudo-solvent approximation increases as the critical point is approached from lower temperatures and also increases with rising salt molality. A similar correction factor is important in the determination of molecular weights of macromolecules from light-scattering data.

List of symbols

A	Helmholtz energy
B_{22}	protein osmotic second virial coefficient

B_{222}	protein osmotic third virial coefficient
c	molar concentration
d	hard-sphere diameter
F	McMillan–Mayer free energy
G_s	protein solvation free energy
k_B	Boltzmann's constant
m	molality
n	power-law exponent
N_{av}	Avogadro's number
p	pressure
q	intra-molecular partition function
r	center-to-center separation
R	gas constant
s	entropy per unit volume
S	entropy
T	temperature
U	internal energy
\bar{v}	partial molar volume
V	volume
V_m	molal volume
$W^{(n)}$	n -body potential of mean force

Greek symbols

β	dimensionless inverse temperature
ε	energy parameter
γ_i	activity coefficient of component i
γ_2^∞	protein infinite-dilution activity coefficient
η	protein packing fraction
λ	thermal wavelength
Δ	dipole moment
μ	chemical potential
Π	osmotic pressure
ρ	number density
σ	surface tension

Subscripts

1	water
2	protein
3	salt
i	inside solution property
o	outside solution property

Superscripts

d	dense-phase property
---	----------------------

- 1 light-phase property
 ∞ infinite dilution of protein

Acknowledgements

For financial support, the authors are grateful to the Office for Basic Energy Sciences of the US Department of Energy and to the National Science Foundation (Grant #CTS-9530793). Robin Curtis thanks Leo Lue for clarifying the statistical mechanics of McMillan–Mayer solution theory.

Appendix A. The preferential-interaction parameter in McMillan–Mayer theory

In this appendix, we provide the derivation of the preferential-interaction parameter in the limit of protein infinite dilution, $(\partial c_3/\partial c_2)_{T,\mu_1,\mu_3}^\infty$ which is approximately equal to $\Delta c_{3,1}$ of Eq. (15). Here, the superscript, ∞ , denotes a property evaluated at infinite dilution of protein in the inside solution. In this limit, there are two effects that contribute to the difference between the outside solution salt concentration and that of the inside solution. (i) The protein molecules on the inside solution have a finite volume so that the volume available to the solvent is less than the total volume of the inside solution. Consequently the salt concentration on the inside solution is reduced with respect to the outside solution. (ii) The salt molality in the domain of the protein is different from the salt molality far away from the protein, or equivalently, the outside-solution salt molality. This latter effect is related to the protein–salt preferential interaction. If the protein–salt interaction is favorable then there is an excess of salt over water in the domain of the protein with respect to the outside solution. If the protein–salt interaction is unfavorable, the salt molality in the domain of the protein is less than that of the outside solution. Thus, in contrast to $(\partial c_3/\partial c_2)_{T,\mu_1,\mu_3}^\infty$, the preferential-interaction parameter in molality units, $(\partial m_3/\partial m_2)_{T,\mu_1,\mu_3}^\infty$, is a more direct measure of the protein–salt interaction because this form of the parameter does not depend on the protein excluded volume. Here, because we use McMillan–Mayer solution theory where volume is the natural extensive variable, we first determine $(\partial c_3/\partial c_2)_{T,\mu_1,\mu_3}^\infty$. This result is then compared with the relation for $(\partial m_3/\partial m_2)_{T,\mu_1,\mu_3}^\infty$, which is derived in the literature using the variables in the Gibbs framework [14].

The starting point is Eq. (A.1), which relates the salt concentration to the McMillan–Mayer free energy

$$c_{3,i} - c_{3,o} = - \left(\frac{\partial F/V}{\partial \mu_3} \right)_{T,\mu_1,c_2} \quad (\text{A.1})$$

where the ideal McMillan–Mayer free energy (Eq. (14)) is used in evaluating Eq. (A.1), because we are only concerned with the infinite-dilution value of the preferential-interaction parameter. Deviations from ideal behavior are associated with protein–protein interactions that occur at finite protein concentrations. Substituting Eq. (14) into Eq. (A.1) gives

$$\left(\frac{\partial F/V}{\partial \mu_3} \right)_{T,\mu_1,c_2}^\infty = c_2 \left(\frac{\partial \mu_2^\theta}{\partial \mu_3} \right)_{T,\mu_1} \quad (\text{A.2})$$

The partial derivative on the right side of Eq. (A.2) involves only properties of the outside solution, so that the dependence on protein concentration can be dropped from the notation. Furthermore, the following

change of variables is used because functions of the properties of the outside solution are usually known in terms of $(T, p_o, m_{3,o})$ instead of (T, μ_1, μ_3)

$$\left(\frac{\partial \mu_2^\theta}{\partial \mu_3}\right)_{T, \mu_1} = \left(\frac{\partial \mu_2^\theta}{\partial p_o}\right)_{T, m_{3,o}} \left(\frac{\partial p_o}{\partial \mu_3}\right)_{T, \mu_1} + \left(\frac{\partial \mu_2^\theta}{\partial m_{3,o}}\right)_{T, p_o} \left(\frac{\partial m_{3,o}}{\partial \mu_3}\right)_{T, \mu_1} \quad (\text{A.3})$$

The main goal of the derivation is to relate the preferential-interaction parameter to thermodynamic properties of the outside solution and to $(\partial \mu_2^\theta / \partial m_{3,o})_{T, p_o}$, because this term can be evaluated from a model for the solvation free energy of the protein molecule (see Eq. (19)). Thus we seek to evaluate three of the four partial differentials on the right side of Eq. (A.3), $(\partial m_{3,o} / \partial \mu_3)_{T, \mu_1}$, $(\partial p_o / \partial \mu_3)_{T, \mu_1}$, and $(\partial \mu_2^\theta / \partial p_o)_{T, m_{3,o}}$. The first two of these partial derivatives involve only variables corresponding to properties of the outside solution and consequently these two derivatives can be evaluated from experimental data of aqueous salt solutions. However, first we must convert the variables involved in the differentiation to the independent variables of the Gibbs framework $(T, p_o, m_{3,o})$.

To evaluate $(\partial m_{3,o} / \partial \mu_3)_{T, \mu_1}$, we use the change of variables

$$\left(\frac{\partial \mu_3}{\partial m_{3,o}}\right)_{T, \mu_1} = \left(\frac{\partial \mu_3}{\partial m_{3,o}}\right)_{T, p_o} + \bar{v}_{3,o} \left(\frac{\partial p_o}{\partial m_{3,o}}\right)_{T, \mu_1} = \left(\frac{\partial \mu_3}{\partial m_{3,o}}\right)_{T, p_o} + \bar{v}_{3,o} c_{3,o} \left(\frac{\partial \mu_3}{\partial m_{3,o}}\right)_{T, \mu_1} \quad (\text{A.4})$$

where quantities with overbars refer to partial molar properties. Here we have used the Maxwell relation, $(\partial \mu_3 / \partial p_o)_{T, m_{3,o}} = \bar{v}_{3,o}$, derived from the differential of the Gibbs energy for the outside solution and the relation $(\partial p_o / \partial m_{3,o})_{T, \mu_1} = c_{3,o} (\partial \mu_3 / \partial m_{3,o})_{T, \mu_1}$, derived from the Gibbs–Duhem equation of the outside solution (Eq. (4)). Collecting the identical terms of Eq. (A.4) to the left side yields

$$\left(\frac{\partial \mu_3}{\partial m_{3,o}}\right)_{T, \mu_1} = \frac{(\partial \mu_3 / \partial m_{3,o})_{T, p_o}}{1 - \bar{v}_{3,o} c_{3,o}} \quad (\text{A.5})$$

All the terms on the right side of Eq. (A.5) refer to properties of the outside solution in terms of variables of the Gibbs framework. The relation $(\partial \mu_3 / \partial m_{3,o})_{T, p_o}$ can be evaluated from an activity-coefficient model for aqueous salt solutions such as the Pitzer correlations [26] and $\bar{v}_{3,o}$ can be determined from densimetric measurements [25].

The relation $(\partial \mu_2^\theta / \partial p_o)_{T, m_{3,o}}$ needs to be evaluated, because we do not have a suitable model for the protein solvation free energy as a function of pressure. Here, the starting point is the Maxwell relation derived from the differential of the Gibbs energy applied to the inside solution

$$\left(\frac{\partial \mu_2}{\partial p}\right)_{T, m_2, m_3} = \bar{v}_2 = \left(\frac{\partial \mu_2^\theta}{\partial p}\right)_{T, m_2, m_3} \quad (\text{A.6})$$

where Eq. (12a) is used for the chemical potential of the protein. In addition, we assume that the solution is incompressible, thus $(\partial c_2 / \partial p)_{T, m_2, m_3} = 0$. To simplify the notation, quantities without a subscript o refer to properties of the inside solution. The right side of Eq. (A.6) is related to the desired quantity, $(\partial \mu_2^\theta / \partial p_o)_{T, m_{3,o}}$, by

$$\left(\frac{\partial \mu_2^\theta}{\partial p}\right)_{T, m_2, m_3} = \left(\frac{\partial \mu_2^\theta}{\partial p_o}\right)_{T, m_{3,o}} \left(\frac{\partial p_o}{\partial p}\right)_{T, m_2, m_3} + \left(\frac{\partial \mu_2^\theta}{\partial m_{3,o}}\right)_{T, p_o} \left(\frac{\partial m_{3,o}}{\partial p}\right)_{T, m_2, m_3} \quad (\text{A.7})$$

The derivative of the outside-solution pressure with respect to the inside-solution pressure, p , can be determined by setting $p_o = p - \Pi^{\text{id}}$, to give

$$\left(\frac{\partial p_o}{\partial p}\right)_{T,m_2,m_3} = 1 - RT\left(\frac{\partial c_2}{\partial p}\right)_{T,m_2,m_3} = 1 \quad (\text{A.8})$$

where the ideal osmotic pressure given by the Van't Hoff equation (Eq. (11)) has been used in deriving Eq. (A.8). The relation $(\partial m_{3,o}/\partial p)_{T,m_2,m_3}$ can be evaluated from the differential of $m_{3,o}$ in terms of the outside solution variables T , μ_1 , and μ_3 . Taking the derivative of this differential with respect to pressure gives

$$\left(\frac{\partial m_{3,o}}{\partial p}\right)_{T,m_2,m_3} = \bar{v}_1\left(\frac{\partial m_{3,o}}{\partial \mu_1}\right)_{T,\mu_3} + \bar{v}_3\left(\frac{\partial m_{3,o}}{\partial \mu_3}\right)_{T,\mu_1} \quad (\text{A.9})$$

The following Maxwell relations corresponding to the inside solution, $(\partial \mu_1/\partial p)_{T,m_2,m_3} = \bar{v}_1$ and $(\partial \mu_3/\partial p)_{T,m_2,m_3} = \bar{v}_3$, have been substituted into Eq. (A.9). The relation $(\partial \mu_1/\partial m_{3,o})_{T,\mu_3}$ is determined using the same methodology used in deriving Eqs. (A.4) and (A.5). This gives

$$\left(\frac{\partial \mu_1}{\partial m_{3,o}}\right)_{T,\mu_3} = -m_{3,o}\left(\frac{\partial \mu_3}{\partial m_{3,o}}\right)_{T,p_o} + \frac{\bar{v}_{1,o}}{V_{m,o}}\left(\frac{\partial \mu_1}{\partial m_{3,o}}\right)_{T,\mu_3} \quad (\text{A.10})$$

where the Gibbs–Duhem derived relation, $(\partial \mu_1/\partial m_{3,o})_{T,p_o} = -m_{3,o}(\partial \mu_3/\partial m_{3,o})_{T,p_o}$, is substituted into Eq. (A.10). Combining the identical terms in Eq. (A.10) yields

$$\left(\frac{\partial \mu_1}{\partial m_{3,o}}\right)_{T,\mu_3} = -\frac{V_{m,o}(\partial \mu_3/\partial m_{3,o})_{T,p_o}}{\bar{v}_{3,o}} \quad (\text{A.11})$$

where $\bar{v}_{1,o}$ has been removed from Eq. (A.11) using the relationship between the molal volume and its partial molar quantities, $V_{m,o} = \bar{v}_{1,o} + m_{3,o}\bar{v}_{3,o}$. Eqs. (A.5) and (A.11) are substituted into Eq. (A.9) to obtain

$$\left(\frac{\partial m_{3,o}}{\partial p}\right)_{T,m_2,m_3} = \frac{\bar{v}_3\bar{v}_{1,o} - \bar{v}_{3,o}\bar{v}_1}{V_{m,o}(\partial \mu_3/\partial m_{3,o})_{T,p_o}} \approx 0 \quad (\text{A.12})$$

In the limit of protein infinite dilution, the partial molar volumes of the solvent components in the inside solution are identical to those of the outside solution and the second equality of Eq. (A.12) becomes exact. Combining Eqs. (A.6)–(A.8) and (A.12) gives

$$\left(\frac{\partial \mu_2^\theta}{\partial p_o}\right)$$

The right side of Eq. (A.14) is only a function of the properties of the outside solution, consequently Eq. (A.14) can be differentiated with respect to protein concentration holding all properties of outside solution constant, to obtain

$$\left(\frac{\partial c_3}{\partial c_2}\right)_{T,\mu_1,\mu_3}^{\infty} = -\left(\frac{\partial \mu_2^{\theta}}{\partial m_{3,o}}\right)_{T,p_o} \frac{1 - \bar{v}_{3,o}c_{3,o}}{(\partial \mu_3/\partial m_{3,o})_{T,p_o}} - \bar{v}_2^{\infty} c_{3,o} \quad (\text{A.15})$$

This result could have also been obtained from the Maxwell relation $(\partial c_3/\partial c_2)_{T,\mu_1,\mu_3} = (\partial \mu_2/\partial \mu_3)_{T,\mu_1,c_2}$ (see Eq. (A.1)).

In order to verify the above derivation with that given in the literature, we need to convert the preferential-interaction parameter to molality units. Here, it is straightforward to use the chain rule of calculus two times in succession. Application of the chain rule the first time yields

$$\begin{aligned} \left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3} &= \left(\frac{\partial m_3}{\partial c_2}\right)_{T,\mu_1,\mu_3} \left(\frac{\partial m_2/V_m}{\partial m_2}\right)_{T,\mu_1,\mu_3} \\ &= \frac{1}{V_m} \left(\frac{\partial m_3}{\partial c_2}\right)_{T,\mu_1,\mu_3} \left[1 - c_2\bar{v}_2 - c_2\bar{v}_3 \left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3}\right] \end{aligned} \quad (\text{A.16})$$

where we have used the relation

$$\left(\frac{\partial V_m}{\partial m_2}\right)_{T,\mu_1,\mu_3} = \bar{v}_2 + \bar{v}_3 \left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3} \quad (\text{A.17})$$

Eq. (A.17) is derived by writing the differential for the molal volume as a function of T , P , μ_2 , and μ_3 and taking the derivative with respect to protein molality holding properties of outside solution constant. The dependence of the volume on pressure is zero because the solution is assumed to be incompressible. Application of the chain rule again yields

$$\begin{aligned} \left(\frac{\partial c_2}{\partial m_3}\right)_{T,\mu_1,\mu_3} &= \left(\frac{\partial c_2}{\partial c_3}\right)_{T,\mu_1,\mu_3} \left(\frac{\partial m_3/V_m}{\partial m_3}\right)_{T,\mu_1,\mu_3} \\ &= \frac{1}{V_m} \left(\frac{\partial c_2}{\partial c_3}\right)_{T,\mu_1,\mu_3} \left[1 - c_{3,i}\bar{v}_3 - c_{3,i}\bar{v}_2 \left(\frac{\partial m_2}{\partial m_3}\right)_{T,\mu_1,\mu_3}\right] \end{aligned} \quad (\text{A.18})$$

where $(\partial V_m/\partial m_3)_{T,\mu_1,\mu_3}$ has been replaced in Eq. (A.18) by an analogous expression to Eq. (A.17). Combining Eqs. (A.16)–(A.18) and taking the limit of infinite dilution of protein gives

$$\left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3}^{\infty} = \frac{c_{3,o}\bar{v}_2^{\infty} + (\partial c_3/\partial c_2)_{T,\mu_1,\mu_3}^{\infty}}{1 - c_{3,o}\bar{v}_{3,o}} \quad (\text{A.19})$$

where the solvent properties of the inside solution are approximated by those of the outside solution. Substituting Eq. (A.15) into Eq. (A.19) yields the desired result

$$\left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3}^{\infty} = -\frac{(\partial \mu_2^{\theta}/\partial m_{3,o})_{T,p_o}}{(\partial \mu_3/\partial m_{3,o})_{T,p_o}} \quad (\text{A.20})$$

Thus, if addition of salt raises the solvation free energy of the protein, the preferential-interaction parameter is negative indicating that salt is excluded from the domain of the protein. If the solvation free energy of

the protein decreases with addition of salt, there is excess salt in the domain of the protein with respect to the bulk solution and the preferential-interaction parameter is positive.

The infinite-dilution value of the preferential-interaction parameter has also been derived starting with variables in the Gibbs framework [14] as outlined below. Consider the differential of the salt chemical potential as a function of the Gibbs variables

$$d\mu_3 = \left(\frac{\partial\mu_3}{\partial T}\right)_{p,m_2,m_3} dT + \left(\frac{\partial\mu_3}{\partial p}\right)_{T,m_2,m_3} dp + \left(\frac{\partial\mu_3}{\partial m_2}\right)_{T,p,m_3} dm_2 + \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,p,m_2} dm_3 \quad (\text{A.21})$$

The preferential-interaction parameter is determined at constant properties of the outside solution. Consequently, we take the derivative of Eq. (A.21) with respect to protein molality, while holding T , μ_1 , and μ_3 constant, to give

$$0 = \bar{v}_3 \left(\frac{\partial p}{\partial m_2}\right)_{T,\mu_1,\mu_3} + \left(\frac{\partial\mu_2}{\partial m_3}\right)_{T,p,m_2} + \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,p,m_2} \left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3} \quad (\text{A.22})$$

where the Maxwell relations, $(\partial\mu_3/\partial p)_{T,m_2,m_3} = \bar{v}_3$, and $(\partial\mu_3/\partial m_2)_{T,p,m_3} = (\partial\mu_2/\partial m_3)_{T,p,m_2}$, have been substituted into Eq. (A.22). Once again, we are deriving the infinite-dilution value of the preferential-interaction parameter. In this limit, we have

$$\left(\frac{\partial p}{\partial m_2}\right)_{T,\mu_1,\mu_3}^\infty = \left(\frac{\partial\Pi}{\partial c_2}\right)_{T,\mu_1,\mu_3}^\infty \left(\frac{\partial c_2}{\partial m_2}\right)_{T,\mu_1,\mu_3}^\infty = \frac{RT}{V_{m,o}} \quad (\text{A.23})$$

where changes in solution pressure at constant properties of the outside solution are given by changes in the osmotic pressure. Substituting Eq. (A.23) into Eq. (A.22) and taking the limit of low protein concentration gives

$$\left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3}^\infty = -\frac{(\partial\mu_2/\partial m_3)_{T,p,m_2}^\infty}{(\partial\mu_3/\partial m_{3,o})_{T,p_o}} - \frac{\bar{v}_{3,o}RT}{V_{m,o}(\partial\mu_3/\partial m_{3,o})_{T,p_o}} \quad (\text{A.24})$$

Because $(\partial\mu_2/\partial m_3)_{T,p,m_2}^\infty$ is evaluated at infinite dilution of protein, the derivative can be evaluated from the ideal protein chemical potential (Eqs. (12) and (13)), yielding

$$\left(\frac{\partial\mu_2}{\partial m_3}\right)_{T,p,m_2}^\infty = \left(\frac{\partial\mu_2^\theta}{\partial m_{3,o}}\right)_{T,p_o} - \frac{RT}{V_{m,o}}\bar{v}_{3,o} \quad (\text{A.25})$$

The second term on the left side of Eq. (A.25) appears because the derivative is evaluated at constant protein molality (moles protein per kg water) instead of in terms of protein concentration (moles protein per liter). In addition, we have used the equality, $(\partial\mu_2^\theta/\partial m_3)_{T,p,m_2}^\infty = (\partial\mu_2^\theta/\partial m_{3,o})_{T,p_o}$, which follows from a similar derivation to the one above where we showed that $(\partial\mu_2^\theta/\partial p)_{T,m_2,m_3}^\infty = (\partial\mu_2^\theta/\partial p_o)_{T,m_{3,o}}$. Substituting Eq. (A.25) into Eq. (A.24) gives Eq. (A.20) which is the same result as obtained using McMillan–Mayer solution theory.

Eq. (A.20) forms the basis for the salting-out studies of Arakawa and Timasheff [22–24]. In the absence of protein–protein interactions (i.e. in sparingly soluble protein solutions), the preferential-interaction parameter is related to the dependence of the protein chemical potential on salt concentration. Arakawa and Timasheff [22–24] determined the infinite-dilution protein activity coefficient from measurements of the protein–salt preferential-interaction parameter. From the infinite-dilution activity coefficient, the authors correlated protein crystal solubility with the assumption that the protein crystal is a pure phase.

Although the protein crystal contains a significant amount of solvent, the authors nevertheless obtained semi-qualitative agreement between protein solubility and the measured preferential-interaction parameters, stressing the importance of the relation between the protein solvation free energy and protein solubility.

References

- [1] W.G. McMillan, J.E. Mayer, *J. Chem. Phys.* 13 (1945) 276–305.
- [2] N. Asherie, A. Lomakin, G.B. Benedek, *Phys. Rev. Lett.* 77 (1996) 4832–4835.
- [3] C. Haas, J. Drenth, W.W. Wilson, *J. Phys. Chem. B* 103 (1999) 2808–2811.
- [4] D. Rosenbaum, P.C. Zamora, C.F. Zukoski, *Phys. Rev. Lett.* 76 (1996) 150–153.
- [5] Y. Chiew, D. Kuehner, H. Blanch, J. Prausnitz, *Am. Inst. Chem. Eng. J.* 41 (1995) 2150–2159.
- [6] M. Malfois, F. Bonnete, L. Belloni, A. Tardieu, *J. Chem. Phys.* 105 (1996) 3290–3300.
- [7] R. Piazza, *J. Cryst. Growth* 196 (1999) 415–423.
- [8] W.C.K. Poon, *Phys. Rev. E* 55 (1997) 3762–3764.