

## Molecular thermodynamics for some applications in biotechnology\*

John M. Prausnitz

*Chemical Engineering Department, University of California, Berkeley and  
Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA  
94720, USA*

*Abstract:* As biotechnology sweeps the world, it is appropriate to remember that the great virtue of thermodynamics is its broad range of applicability. As a result, there is a growing literature describing how chemical thermodynamics can be used to inform processes for old and new biochemical products for industry and medicine. A particular application of molecular thermodynamics concerns separation of aqueous proteins by selective precipitation. For this purpose, we need phase diagrams; for constructing such diagrams, we need to understand not only the qualitative nature of phase equilibria of aqueous proteins, but also the quantitative intermolecular forces between proteins in solution. Some examples are given to show how aqueous protein–protein forces can be calculated or measured to yield a potential of mean force and how that potential is then used along with a statistical-thermodynamic model to establish liquid–liquid and liquid–crystal equilibria. Such equilibria are useful not only for separation processes, but also for understanding diseases like Alzheimer’s, eye cataracts, and sickle-cell anemia that appear to be caused by protein agglomeration.

### INTRODUCTION

Historians of science have often pointed out that thermodynamics provides two strong, possibly unique, advantages. First, thermodynamics has a very wide range of application as it is used in a large variety of scientific disciplines (physics, chemistry, earth sciences, astronomy, engineering, life sciences). Second, because it is an integrative science concerned with bulk properties, it can provide useful answers to scientific problems with a minimum of experimental data. Thermodynamics can provide useful information on material properties based on a few experimental measurements without specification of molecular details. It is this integrative, coarse-grained feature of thermodynamics that makes it attractive for describing some of the overall features of biological systems, that is, for systems that, on a molecular scale, are so complex, that a detailed description is extremely difficult because it would require a large and expensive experimental effort.

This article considers thermodynamic properties of proteins in aqueous solutions that may also contain salts or other water-soluble solutes. Special attention is given to phase equilibria. Our discussion is simplistic in the sense that consideration is given only to gross, overall properties, toward obtaining a bird’s-eye view that gives the main features while neglecting numerous details. The goal-oriented viewpoint represented here uses molecular thermodynamics toward attaining some broad approximate results that can be reached relatively easily, rather than specific, highly accurate results that require a formidable experimental effort for a limited application. Molecular thermodynamics combines

---

\*Rossini lecture presented at the 17<sup>th</sup> IUPAC Conference on Chemical Thermodynamics (ICCT), Rostock, Germany, 28 July–2 August 2002. Other presentations are published in this issue, pp. 859–936.

classical and statistical thermodynamics coupled with models from physics and chemistry toward achieving practical ends. Phase behavior of aqueous proteins is of particular importance in medicine and for separation processes in biotechnology. There is strong evidence to suggest that protein aggregation is responsible for a variety of serious diseases including eye cataracts, sickle-cell anemia, and Alzheimer's disease; better understanding of protein phase behavior may assist in preventing or curing such illnesses. In the separation of biochemical products, the aqueous broth leaving a bioreactor inevitably is a mixture; a common primary separation step is to add a precipitation agent (typically ammonium sulfate) that selectively precipitates some of the proteins in the broth. Better understanding of the pertinent phase equilibria can help to optimize that and other separation processes [1].

To establish a quantitative method for describing phase equilibria in aqueous protein solutions, we require two kinds of information. First, we require a thermodynamic relation that gives the chemical potential of a protein as a function of its state, that is, temperature, concentration of protein and of other species present in the solution, and, very important, pH because the net electric charge of a protein (a polymer of amino acids) depends strongly on pH. The desired thermodynamic relation can be obtained from any one of a variety of theories through statistical thermodynamics. Second, to utilize a statistical-thermodynamic theory, we need a quantitative expression that describes the intermolecular forces between two dissolved protein particles; this quantitative expression depends on their separation, that is, on center-to-center distance  $r$ .

The force  $F$  between two globular proteins in solution is related to the potential of mean force  $W(r)$  by

$$F = \frac{-dW(r)}{dr} \quad (1)$$

The potential of mean force is similar to the more commonly known potential between two molecules in vacuum [2]. For example, according to Coulomb's law, the potential  $\Gamma$  for two charged particles  $i$  and  $j$ , separated by center-to-center distance  $r$ , is given by

$$\Gamma = \frac{e_i e_j}{r} \quad (2a)$$

where  $e_i$  is the charge on particle  $i$  and  $e_j$  is the charge on particle  $j$ . However, eq. 2a assumes that the two particles are *in vacuo*. If, instead, they are in a medium (i.e., a solvent), the corresponding potential of mean force is

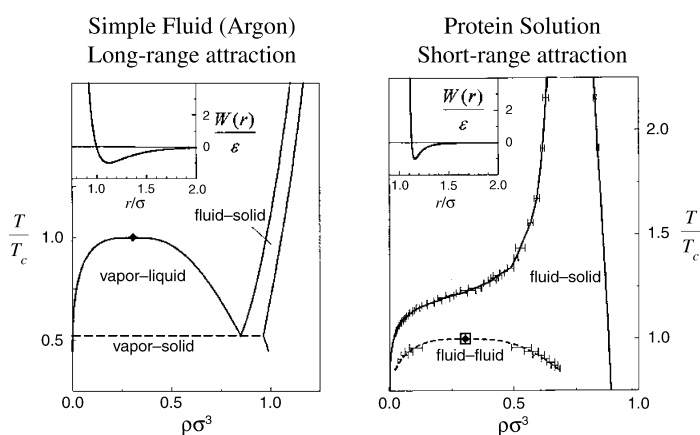
$$W(r) = \frac{e_i e_j}{rD} \quad (2b)$$

where  $D$  is the medium's dielectric constant. The adjective "mean force" indicates that, for any two particles  $i$  and  $j$  in a medium, the potential of mean force is the free energy of the  $ij$  pair, averaged with respect to all possible configurations of the solvent molecules. For symmetric spherical particles  $i$  and  $j$ , the potential  $\Gamma$  depends only on  $r$  and on the properties of  $i$  and  $j$ . By contrast,  $W(r)$  depends, in addition, on the properties of the medium.

While much is known about the statistical thermodynamics of spherical particles [3], much less is known about the potential of mean force for such particles that we use to represent globular proteins. To construct a phase diagram, we need a statistical-thermodynamic theory for fluids and a statistical-thermodynamic theory for solids. Here, phase diagram refers to a plot of temperature vs. density (protein concentration), and fluid refers to a fluid of protein particles in a solvent medium. For a reasonable first approximation, we can use the random-phase approximation (RPA) theory of fluids [4,5] and the (modern) van der Waals theory of solids [6]. But what shall we use for  $W(r)$ ? We can obtain some information from theory and from molecular simulations. But a more reliable source of information is one of several experimental methods as discussed in a later section.

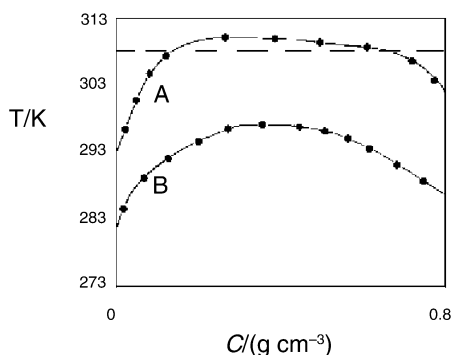
## PHASE DIAGRAMS FOR AQUEOUS PROTEINS

The potential of mean force determines not only the quantitative details of a phase diagram, but also its qualitative features, as illustrated in Fig. 1. When the range of attractive forces is reasonably large, of the order of particle diameter  $\sigma$ , we obtain the conventional phase diagram shown on the left side of Fig. 1 where, as density rises, we first have a transition from dilute fluid (gas) to dense fluid (liquid), and then from liquid to solid. However, if the attractive range is short (that is, a small fraction of  $\sigma$ ), then, as density rises, we go directly from dilute fluid to solid unless the temperature is below critical temperature  $T_c$ . In that event, as density rises, we first go through a metastable fluid–fluid transition and then, with further increase in density, to the solid. The essential difference is this: for long-range attraction, the freezing line lies to the right of the fluid–fluid transition. For short-range attraction, the freezing line lies initially to the left and then above the metastable liquid–liquid region. These calculated phase diagrams [4] were obtained by molecular simulations using the potentials of mean force indicated in the diagrams [6] where  $\varepsilon$  is the depth of the energy well (i.e., where  $W(r)/\varepsilon = -1$ ).



**Fig. 1** Phase transitions for argon and for protein solutions with reduced temperature plotted against reduced density  $\rho\sigma^3$ . (a) Long-range attraction is shown for the simple fluid argon. (b) Short-range attraction is shown for a protein solution. From ref. 4.

There is much evidence to suggest that the metastable liquid–liquid region of a protein solution is responsible for cataract formation [7]. As the eye ages, its protein concentration changes, and one (or more) proteins may achieve a concentration that exceeds the saturation concentration as indicated in Fig. 2. Following phase separation, the highly concentrated liquid phase does not settle, but remains in the eye as a fine dispersion whose optical properties seriously interfere with vision. The experimental data in Fig. 2 show that when the concentration of native  $\gamma$ -crystallin exceeds about  $100 \text{ mg cm}^{-3}$  at body temperature  $310 \text{ K}$  ( $37^\circ \text{C}$ ), a second liquid phase is formed whose concentration is about  $700 \text{ mg cm}^{-3}$ . It is this second liquid phase (which may eventually become crystalline) that is responsible for cataract symptoms. Precipitation of the second phase can be avoided by adding a small amount of glutathione that forms a soluble complex with  $\gamma$ -crystallin to keep it in dilute solution. These experimental results provide the first step for a possible method to prevent cataracts in the human eye, a disease that strikes many millions of elderly men and women throughout the world. Similar studies have been reported for sickle-cell anemia and for fibril formation that appears to be responsible for Alzheimer's disease [8–11].



**Fig. 2** Cataract formation: liquid–liquid separation of  $\gamma$ -crystallin in the eye shown as temperature plotted against  $\gamma$ -crystallin concentration. (A) Native  $\gamma$ -crystallin. (B)  $\gamma$ -crystallin with glutathione. The broken line denotes body temperature 310 K (37 °C).

### POTENTIAL OF MEAN FORCE FROM THEORY OR SIMULATION

Although there is a huge (colloid) literature concerning the theory of  $W(r)$  for charged particles, we can here discuss only a few representative results. We consider a globular protein of diameter  $\sigma$  with net electric charge  $q$ . We consider three contributions to  $W(r)$ :

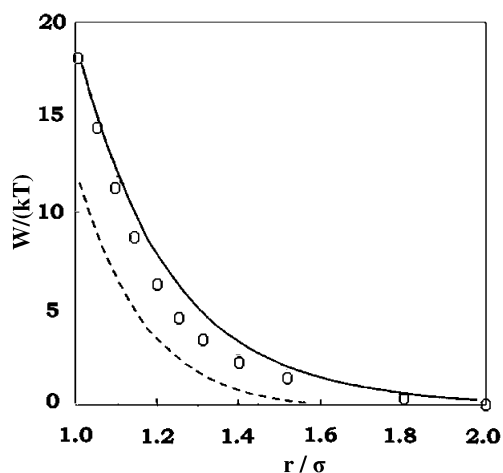
$$W(r) = W(r)_{\text{hs}} + W(r)_{\text{el}} + W(r)_{\text{disp}} \quad (3)$$

where hs stands for hard-sphere repulsion, el stands for electrostatic repulsion (or attraction), and disp stands for attractive dispersion. The first term is readily obtained using diameter  $\sigma$ . The second term is found from Debye–Hückel theory (or one of its modifications). Owing to fluctuating dipoles of the polarizable protein particles, the third term is calculated from Hamaker's theory based on the London dispersion potential. Equation 3 is the basis of the well-known, 54-year-old DLVO theory where the initials stand for Derjaguin, Landau, Verwey, and Overbeek [12]. Numerous variations of DLVO theory have been proposed, including that by Sogami and Ise (SI) [13].

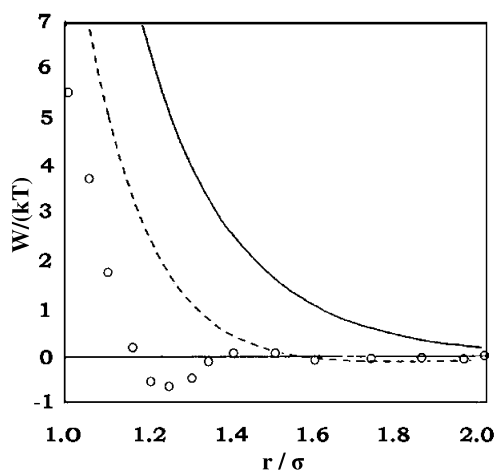
**Table 1** Net charge  $q$  as a function of pH for ovalbumin and for lysozyme.

pH	Ovalbumin	Lysozyme
3	28	13
4	12	10.5
5	-1	9
6	-8	8
7	-12	7.5
8	-14	7
9	-16	6
10	-20	3.5
11	-28	0

The net charge  $q$  on a protein depends on pH as illustrated in Table 1 that gives net charges for lysozyme and for ovalbumin. At any fixed pH, two identical protein particles carry the same net charge; both are either positive or negative. If we neglect the small contribution from dispersion attraction, we would expect the forces between these two protein particles to be repulsive, which is essentially equivalent to saying that  $W(r)$  should be positive. When the medium contains a monovalent cation, the DLVO theory is in qualitative agreement with molecular simulations as shown in Fig. 3. However, as shown in



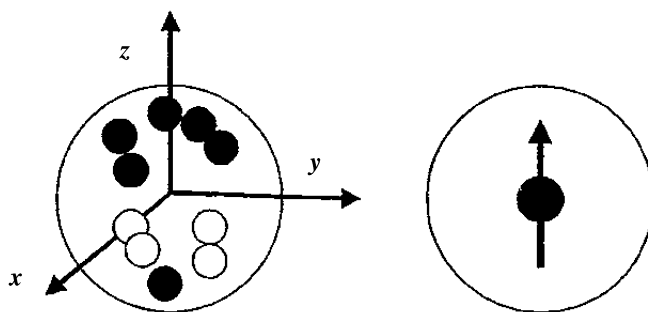
**Fig. 3** Potential of mean force  $W(r)$  plotted against reduced separation between a negatively charged protein at  $T = 298$  K.  $\circ$  = results from Monte Carlo simulation; ---- = theoretical model of SI; — = theoretical model of DLVO. For the protein, the charge is  $-20 \epsilon_0$  and diameter = 2.0 nm. For the salt with the monovalent cation (constituent valences +1, -2), ionic strength = 0.28 M with diameter 0.4 nm.



**Fig. 4** Potential of mean force plotted against reduced separation between a negatively charged protein at  $T = 298$  K.  $\circ$  = results from Monte Carlo simulation; ---- = theoretical model of SI; — = theoretical model of DLVO. For the protein, the charge is  $-20 \epsilon_0$  and diameter = 2.0 nm. For the salt with the divalent cation (constituent valences +2, -1), ionic strength = 0.31 M with diameter 0.4 nm.

Fig. 4, when the medium contains a divalent cation, molecular simulations show that the potential of mean force is negative for some values of  $r$ , indicating that essentially for those values of  $r$ , the two like-charged proteins experience attraction. The simple theories (DLVO and SI) are unable to show this result that follows from electrostatic binding of the divalent ions to the oppositely charged protein particles [14].

The DLVO potential of mean force can be improved by considering charge distribution. A protein particle has many electric charges situated at various locations;  $q$  is only the net charge. The charge distribution can be represented by a dipole moment,  $\mu$ , as indicated in Fig. 5.



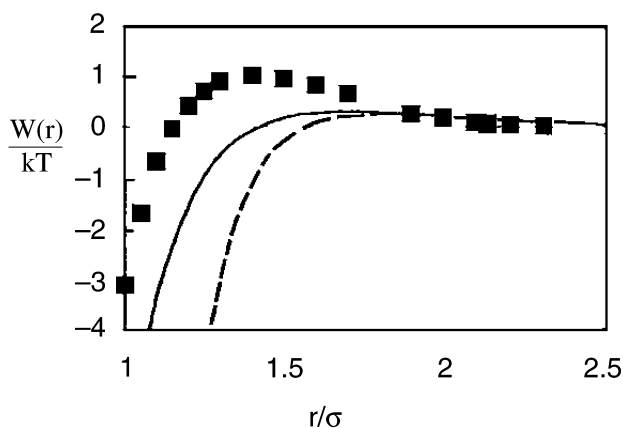
**Fig. 5** Schematic diagram to show the effect of charge distribution by representing the six positive charges (filled circles) and four negative charges (open circles) at specific positions (net charge +2) by the single equivalent dipole moment (net charge +2). The equivalent dipole moment enables analytical calculation of the potential of mean force  $W(r)$ .

To account for charge distribution through a dipole moment, the potential of mean force is

$$W(r) = W_{\text{hs}} + W_{qq} + W_{q\mu} + W_{\mu\mu} + W_{\text{disp}} \quad (4)$$

where again hs refers to hard-sphere repulsion due to finite size (diameter  $\sigma$ );  $qq$  refers to charge–charge interactions (Debye–Hückel);  $q\mu$  refers to charge–dipole interactions, and  $\mu\mu$  refers to dipole–dipole interactions.

Because contributions  $q\mu$  and  $\mu\mu$  are orientation-dependent, to obtain a potential of mean force  $W(r)$  for phase-diagram calculations, it is necessary to calculate orientation averages where each orientation is weighted by its Boltzmann factor. It has been common practice to angle-average  $W_{qq}$  and  $W_{q\mu}$  separately because the mathematics to do so is much simpler than that needed to average them together, as required by theory. The necessary analytic expressions for simultaneous averaging have been published only recently [15]. Figure 6 compares Monte Carlo simulations [16] with analytic results based on eq. 4, neglecting the last term. While simultaneous averaging gives better agreement with simulations than separate averaging, quantitative agreement is not good.



**Fig. 6** Potential of mean force  $W(r)/kT$  plotted against reduced separation between two proteins with a charge and a dipole at  $T = 298$  K. Solid squares = results from simulation; ----- and ——— = electrostatic theory with the terms angle-averaged separately and together, respectively. For the protein, the net charge is  $+8 \epsilon_0$ , dipole moment = 490 Debye and diameter = 2 nm. For the monovalent salt, ionic strength = 0.12 M with the salt-ion diameter = 0.4 nm.

## EXPERIMENTAL METHODS

Because theoretical models and simulation calculations are based on numerous simplifying assumptions, they can provide only qualitative or, at best, semi-quantitative results for the potential of mean force. Such results are nevertheless useful because they can serve as a guide for trends. But for quantitative results, we must make experimental measurements. Regrettably, even these can give us only rough results because, to interpret experimental data toward obtaining a potential of mean force, we must also make a variety of simplifying assumptions.

A well-known procedure for investigating intermolecular forces in gases is to measure low-density volumetric (*PVT*) properties that give the second virial coefficient  $B$ . Per mole of spherical molecules,  $B$  is related to the two-body potential  $\Gamma(r)$  by

$$B = -2\pi N_A \int_0^\infty \left[ \exp\left(\frac{-\Gamma(r)}{kT}\right) - 1 \right] r^2 dr \quad (5a)$$

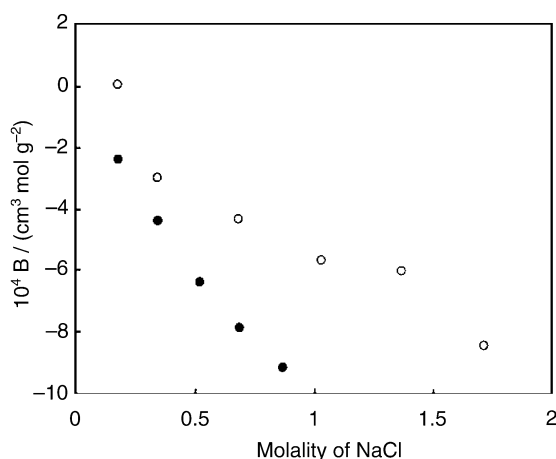
where  $k$  is Boltzmann's constant and  $N_A$  is Avogadro's number.

Similarly, for globular proteins dissolved in a fluid medium, the osmotic second virial coefficient  $B$  is related to the two-body potential of mean force  $W(r)$  by essentially the same relation:

$$B = -2\pi N_A \int_0^\infty \left[ \exp\left(\frac{-W(r)}{kT}\right) - 1 \right] r^2 dr \quad (5b)$$

The osmotic second virial coefficient is readily measured by osmometry [17] or, more accurately, by low-angle laser-light scattering (LALLS) [18].

Figure 7 shows osmotic second virial coefficient  $B$  for the aqueous globular protein hen-egg lysozyme (molar mass 14 000) at pH 4.5 and  $T = 298.15$  K as a function of sodium chloride concentration [19]. Results are shown for native lysozyme and for a mutant where a phenylalanine residue (F) has replaced the aspartic-acid residue (D) at position 101. Because  $B$  is negative, the protein-protein forces are predominantly attractive, becoming more so as the molality of salt rises. Attractive forces between two mutant protein particles are stronger than those between two native protein particles because the F residue is more lipophilic than the D residue; thus, hydrophobic attraction is stronger for the mutant.



**Fig. 7** Osmotic second virial coefficient  $B$  for native lysozyme and for a lysozyme mutant plotted against molality of NaCl for  $T = 298$  K and pH = 4.5: open circles denote native lysozyme, closed circles denote D101F lysozyme. In the mutant, phenyl alanine F has replaced aspartic acid D in the 101 position of lysozyme.

Because  $W(r)$  appears under an integral sign in eq. 5b, experimental data for  $B$  cannot give a unique potential of mean force. For a fixed temperature and given value of  $B$  at that temperature, numerous forms of  $W(r)$  can satisfy eq. 5b. A common procedure is to assume that  $W(r)$  is given by the DLVO theory plus an additional attractive contribution represented by a simple potential function (e.g., the Yukawa potential or the sticky-sphere potential or the square-well potential). For data reduction, we require two independent parameters: the protein diameter  $\sigma$  that is often available from X-ray crystallography, and the protein's pH-dependent net charge  $q$  obtained from acid-base titration data. For native lysozyme,  $\sigma = 3.44$  nm and at pH = 4.5,  $q = +9.8$ . Hamaker's constant  $H$  is typically set equal to  $5 kT$ . For the Yukawa or square-well potential, the additional contribution concerning  $W(r)$  requires two parameters: a range (domain) where the additional attractive force is effective and, most important, a characteristic attractive energy  $\varepsilon$  or  $\varepsilon$  (specific). Figure 7 shows that, for the mutant,  $\varepsilon$  is appreciably larger than that for native lysozyme.

Another relatively simple experimental method to obtain information concerning  $W(r)$  is provided by cloud-point measurements. At a temperature somewhat in excess of ambient, a homogeneous aqueous solution contains known concentrations of protein and salt in a glass container kept at a known temperature. A strong light is placed at the rear of the container, and a scope is placed in front. As the solution is slowly cooled, the temperature is closely monitored. When the scope reports the first trace of a finely dispersed precipitate, we record the cloud-point temperature. This temperature, coupled with some assumptions and a statistical thermodynamic equation of state, can give us a potential of mean force. Toward that end, the essential steps are as follows:

A mathematical form is assumed for  $W(r)$ . Grigsby [20] assumed that  $W(r)$  contains the DLVO terms and a specific square-well potential with specified width and unknown depth  $\varepsilon$ . This unknown depth, a measure of specific protein-protein attractions, is obtained from simultaneous solution of two equations of phase equilibrium:

$$\mu'_{\text{Pr}} = \mu''_{\text{Pr}} \quad (6)$$

$$P' = P'' \quad (7)$$

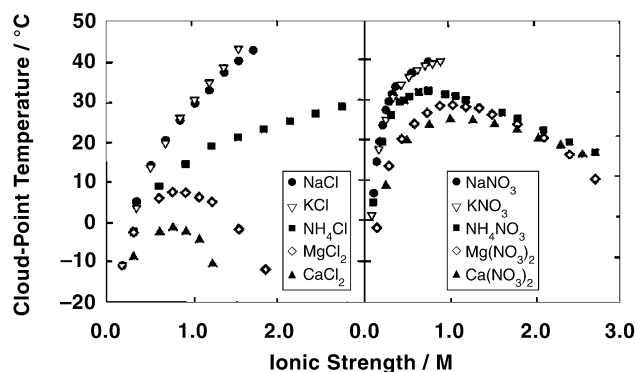
where  $\mu_{\text{Pr}}$  is the chemical potential of the protein and  $P$  is osmotic pressure; prime (') and double prime (") refer to the equilibrated phases at the cloud-point temperature. From the statistical-thermodynamic model, we have expressions for  $\mu_{\text{Pr}}$  and  $P$  as functions of number density (protein concentration), temperature, and  $W(r)$  [20]. From independent data, we know protein diameter  $\sigma$ , net charge  $q$  corresponding to the experimental pH, and the dielectric constant of water. There are two unknowns: square-well depth  $\varepsilon$  and the number density of the precipitated phase.

Figure 8 shows some experimental cloud-point temperatures for lysozyme in a variety of aqueous salt solutions [20]. For most cases, as salt concentration increases, the cloud-point temperature rises, indicating that increasing salt concentration raises protein-protein attraction. At the pH indicated, lysozyme has a positive net charge. Consistent with the theoretical results shown in Figs. 3 and 4, the cloud-point data show that the effect of a divalent cation is qualitatively different from that of a univalent cation.

While relatively simple, the three experimental methods indicated above have some disadvantages. Osmometry and LALLS and cloud-point measurements require a relatively large amount of protein (in the region of 100 or 200 mg). Osmometry and LALLS are often not useful for small proteins or protein segments because for such cases the membrane is not sufficiently semi-permeable and because the particle size is too small relative to the commonly used wavelength of light. These limitations are overcome by a different experimental method: fluorescence-anisotropy [21,22]. This method measures the dimerization constant for the equilibrium







**Fig. 8** Cloud-point temperatures plotted against molar ionic strength for lysozyme at pH = 7. Effect of cations in chloride or nitrate salts.

where Pr stands for the protein or protein segment. As shown by Chandler [23], the association equilibrium constant  $K$  is related to  $W(r)$  by

$$K = \frac{1}{2} \int_{\sigma}^{r_c} \left[ \exp\left(\frac{-W(r)}{kT}\right) \right] 4\pi r^2 dr \quad (9)$$

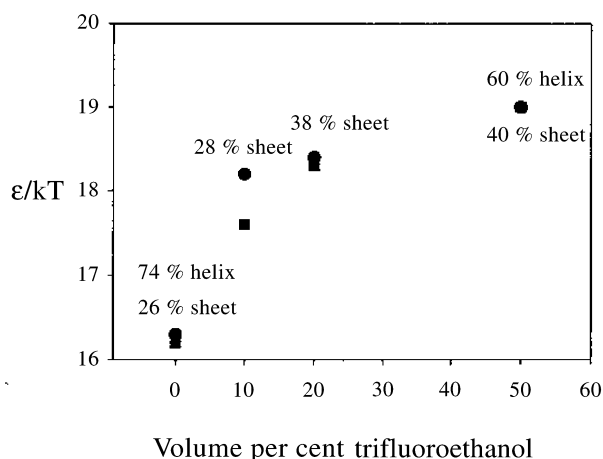
where  $\sigma$  is the diameter of the protein and  $r_c$  is the range (or domain) of the attractive forces that lead to dimerization.

In the fluorescence-anisotropy method, a fluorescent label is attached to a very small amount of Pr and mixed with a larger amount of nonlabeled Pr, all in liquid solution. The solution is subjected to monochromatic polarized light that optically excites the label. The label emits radiation with two components: one parallel ( $I_{\parallel}$ ) and one perpendicular ( $I_{\perp}$ ) to the incident light. The intensities  $I$  of these two components are measured to give anisotropy  $A$ :

$$A = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad (10)$$

If the labeled Pr rotates rapidly,  $A$  is small, but if the labeled Pr is dimerized with a nonlabeled Pr, the speed of rotation is reduced and  $A$  becomes large. By making measurements of  $A$  as a function of total concentration, dimerization constant  $K$  can be obtained providing information about  $W(r)$  through eq. 9.

To illustrate, Fig. 9 shows some results obtained by fluorescence-anisotropy [24] for a protein segment from hen-egg lysozyme. The segment [T4LYS(11-36)] has a molar mass of 3000, and its diameter is 0.188 nm. In data reduction,  $W(r)$  was assumed to be the three DLVO terms plus a specific square-well potential whose depth is  $\varepsilon$  (specific). Figure 9 shows the effect of protein secondary structure; when a denaturant, trifluoroethanol, is added to the aqueous solution, the secondary structure changes from  $\alpha$ -helix to  $\beta$ -sheet as determined by circular-dichroism measurements. As shown in Fig. 9,  $\varepsilon/kT$  is large compared to unity. Therefore, for this protein segment in 0.16 M sodium chloride solution, the DLVO electrostatic and Hamaker-dispersion contributions to  $W(r)$  are much smaller than that from the specific square-well potential. Attractive forces for the  $\beta$ -sheet structure are appreciably stronger than those for the  $\alpha$ -helix structure.



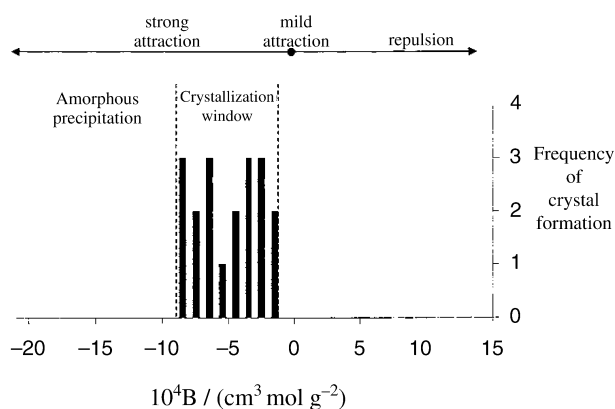
**Fig. 9** Peptide interactions from fluorescence anisotropy are shown as a plot of the depth of the square-well potential against volume per cent trifluoroethanol at pH = 4, 0.16 M NaCl. The solid circles denote  $T = 298$  K, and solid squares  $T = 310$  K.

### POTENTIAL OF MEAN FORCE AND CRYSTALLIZATION

Proteins do not crystallize readily from solution unless the solution conditions are “just right”. Because  $B$ , the osmotic second virial coefficient, can be measured readily for different solution conditions, it would be helpful for protein crystallographers to establish a possible correlation between  $B$  and the likelihood of protein-crystal formation.

Crystallographic data for proteins are in much demand by pharmaceutical chemists because they provide the best way for determining a protein’s three-dimensional structure. Often, the most difficult part of protein crystallography is to make the protein crystal; because protein particles are very “fussy” about forming or not forming crystals, producing such crystals is as much art as science, often requiring extensive trial and error. Toward finding optimum crystallization conditions (temperature, pH, nature and concentration of added solutes), it would be useful to obtain guidance from  $B$  data for a variety of such conditions. Because  $B$  is closely related to  $W(r)$  and because  $W(r)$  has a major influence on crystal formation, it is reasonable to hope that experimental osmotic second virial coefficients might correlate with successful or unsuccessful conditions for crystallization. Indeed, such a correlation has been established by George and Wilson [25] as shown in Fig. 10. George and Wilson found a crystallization window, that is, a region of  $B$  where crystallization is favored. Outside that window, crystallization did not occur.

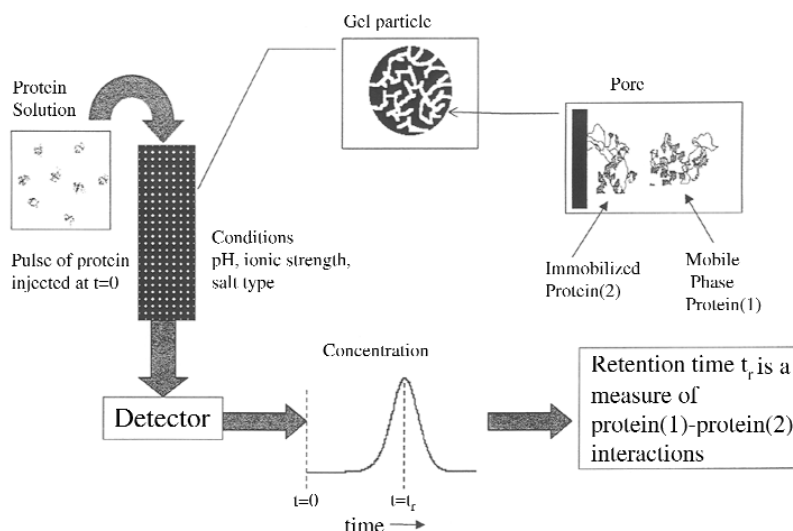
To form a crystal, dissolved protein particles must be in the correct orientation. If the forces of attraction are too strong, protein particles do not have enough time to “get together” in just the right way; they agglomerate to form an amorphous precipitate. On the other hand, if there are no forces of attraction ( $B > 0$ ), there is no precipitate, either amorphous or crystalline. For crystal formation, the forces of attraction should be weak; in other words,  $B$  should be negative but not excessively so. Figure 10 shows that, when  $B$  is in the region  $-1$  to  $-8$  cm<sup>3</sup> mol<sup>-2</sup>, crystallization is favored. When  $B$  is outside that region, crystallization is not likely. Thus, experimental  $B$  data that can be obtained relatively quickly serve as useful screening criteria for favorable crystallization conditions.



**Fig. 10** Protein crystallization window in which the frequency of crystal formation is plotted against the osmotic second virial coefficient. See ref. 25.

## TWO-PROTEIN SYSTEMS

For a system containing only one type of protein, the potential of mean force describes the intermolecular forces for two identical protein particles dissolved in a solvent that may also contain other solutes such as salts. This potential of mean force, coupled with a single-component equation of state (provided by a statistical-thermodynamic model), is sufficient to calculate an approximate phase diagram. However, if the solvent (that may include salts) contains two types of proteins 1 and 2, we need a two-component equation of state and three potentials of mean force:  $W_{11}(r)$ ,  $W_{22}(r)$ , and  $W_{12}(r)$ . How can we obtain  $W_{12}(r)$  experimentally? Osmometry, LALLS, and cloud-point measurements are not suitable for a direct measurement of  $W_{12}(r)$  because these measurements give the combined effect of all three potentials of mean force. Fluorescence-anisotropy can be used for direct measurement of  $W_{12}(r)$ , but a simpler method is provided by quantitative affinity-chromatography [26] as shown in Fig. 11. Protein 2 is immobilized on the surface of a solid support that fills the chromatographic column. A protein-free solution at the desired pH and at the desired salt concentration flows over the solid support. At time



**Fig. 11** Schematic flow chart for quantitative affinity chromatography to determine  $W_{12}(r)$ .

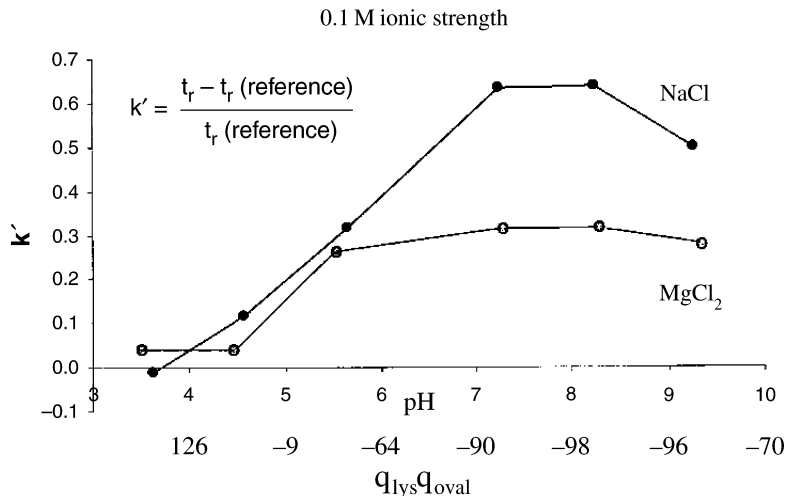
$t = 0$ , a small quantity of solution of protein 1 is injected immediately upstream of the chromatographic column. A detector immediately downstream from the column registers the arrival of protein 1 giving retention time  $t_r$ . In a separate experiment, the same procedure is used, except that now there is no immobilized protein 2 on the column; that experiment gives reference retention time  $t_{r0}$ . The key experimental quantity is the reduced relative retention time (retention factor)  $k'$ :

$$k' = (t_r - t_{r0})/t_{r0} \quad (11)$$

From  $k'$ , and some simplifying assumptions, we can calculate the osmotic second virial cross-coefficient  $B_{12}$  that, as in eq. 5b, is related to  $W_{12}(r)$  by

$$B_{12} = -2\pi N_A \int_0^\infty \left[ \exp\left(\frac{-W_{12}(r)}{kT}\right) - 1 \right] r^2 dr \quad (5c)$$

Figure 12 shows some results for  $k'$  obtained by Teske [26] for an aqueous two-protein system containing ovalbumin (1) and lysozyme (2). Results at  $T = 298.15$  K are shown as a function of pH for two cases: first, for an aqueous solvent containing 0.1 M ionic strength sodium chloride and second, for an aqueous solvent containing 0.1 M ionic strength magnesium chloride. Again, we see the difference between solutions that contain a univalent or a divalent cation. The horizontal axis also shows the product of the net charges,  $q$  (lysozyme) and  $q$  (ovalbumin) because that product gives a rough measure of electrostatic forces (see Table 1). If the product is positive, the two proteins carry either both positive or both negative charges indicating repulsion. Indeed, at the left side,  $k'$  is close to zero because in the absence of attraction, the retention time  $t_r$  of the flowing protein (ovalbumin) is nearly the same as  $t_{r0}$  where the column does not contain protein 2. As the product  $q$  (lysozyme)  $\cdot$   $q$  (ovalbumin) becomes increasingly negative, attraction increases and  $k'$  rises. As the product becomes less negative, we expect  $k'$  to decline.



**Fig. 12** Plot of the reduced relative retention time factor  $k'$  against pH for an aqueous two-protein system containing ovalbumin and lysozyme at 0.1 M ionic strength NaCl or  $\text{MgCl}_2$  at  $T = 298$  K. The product of the net charges  $q$  (lysozyme) and  $q$  (ovalbumin) is also shown along the x-axis.

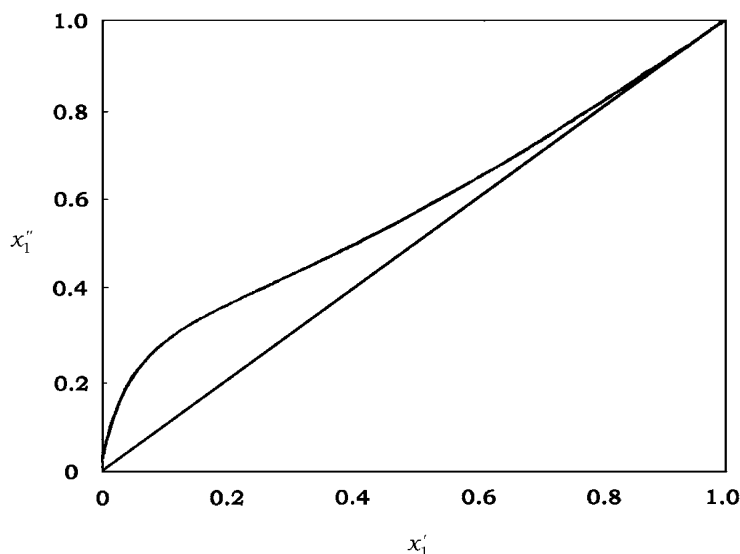
A liquid–liquid phase diagram for the aqueous two-protein system ovalbumin (1)–lysozyme (2) can be constructed when potentials of mean force  $W_{11}(r)$ ,  $W_{22}(r)$ , and  $W_{12}(r)$  are known by using a two-component form of the RPA theory. This gives chemical potentials  $\mu_1$  and  $\mu_2$  as functions of temperature and both protein number densities (concentrations) in the aqueous system. Each of the three po-

tentials of mean force is set equal to the sum of the appropriate DLVO terms plus an additional square-well potential whose well depth is  $\varepsilon_{ij}$  ( $i = 1, 2; j = 1, 2$ ). As a rough approximation, we may expect  $\varepsilon_{12} = \varepsilon_{21}$  to be the geometric mean of  $\varepsilon_{11}$  and  $\varepsilon_{22}$ , but this approximation requires a correction that is conventionally written

$$\varepsilon_{12} = \sqrt{\varepsilon_{11}\varepsilon_{22}}(1 - k_{12}) \quad (12)$$

where  $k_{12}$  is a binary parameter obtained from  $W_{12}(r)$ ,  $W_{11}(r)$ , and  $W_{22}(r)$ .

Figure 13 [27] shows the (solvent-free) mole fractions  $x'$  and  $x''$  of the two fluid phases for the aqueous {ovalbumin (1) + lysozyme (2)} system at  $T = 298.15$  K and  $\text{pH} = 7$  and a high ionic strength as often used in industrial separations [28], viz. 9 M ammonium sulfate. Figure 12 is analogous to the well-known  $y$ - $x$  diagrams used for vapor-liquid equilibria required for design of distillation columns, where  $y$  is the mole fraction in the vapor phase. Figure 12 provides the phase equilibria for design of an extraction process to separate an aqueous mixture of ovalbumin and lysozyme. Although Fig. 12 shows chromatographic results for sodium chloride and magnesium chloride rather than ammonium sulfate, and although the salt concentrations in Fig. 11 are much lower than those of Fig. 13, Fig. 12 suggests that easier separation may be achieved at  $\text{pH} = 4$  rather than  $\text{pH} = 7$ . This occurs because at the lower  $\text{pH}$ , lysozyme and ovalbumin show much lower mutual attraction.



**Fig. 13** Plot of solvent-free mole fraction  $x_1''$  in the light aqueous phase vs.  $x_1'$  for the heavy aqueous phase for the two-fluid phases of aqueous ovalbumin (1) and lysozyme (2) at  $T = 298$  K,  $\text{pH} = 7$  and ionic strength 9 M ammonium sulfate. The value of  $k_{12} = -0.02$  in eq. 12.

## CONCLUSION

The discussion presented here shows that well-known methods of molecular thermodynamics can be extended to provide useful results for aqueous systems that contain one or more proteins. The key to obtaining such results lies in the potential of mean force. While the potentials of mean force shown here are only rough approximations of the “true” potential of mean force, they enable us to understand the main features of phase behavior in protein systems, where detailed molecular information would require a large experimental effort. The models and methods introduced here can surely be much refined, and there is good reason to believe that such refinements will be forthcoming as more fundamental knowl-

edge concerning proteins becomes available. However, even now, when our fundamental knowledge is far from complete, we can begin to understand some of the essential features of protein phase behavior. The discussion presented here once again shows the wisdom of the preface of the classic book on chemical thermodynamics by Lewis and Randall published about 80 years ago:

“The fascination of any growing science lies in the work of the pioneers at the very borderland of the unknown. But to reach that frontier one must pass over well-traveled roads. One of the safest and surest is the broad highway of thermodynamics.”

## ACKNOWLEDGMENTS

The author is grateful for financial support provided by the National Science Foundation and the Office of Basic Sciences of the U.S. Department of Energy. For extensive contributions to the results shown here, the author is grateful to numerous coworkers including Prof. Harvey Blanch and Dr. Dusan Bratko and to former graduate students Robin Curtis, Camille Anderson, Chris Teske, Ying-Chou Shih, John Grigsby, Alberto Striolo, Daniel Kuehner, Chris Coen, Charles Haynes, A. Louise Creagh, and Jianzhong Wu.

## REFERENCES

1. F. Rothstein. *Protein Purification Process Engineering*, Dekker, New York (1994).
2. T. L. Hill. *An Introduction to Statistical Thermodynamics*, Chap. 19, Addison Wesley, Reading, MA (1960).
3. D. A. McQuarrie. *Statistical Mechanics*, University Science Books, Sausalito, CA (2000).
4. P. R. ten Wolde and D. Frenkel. *Science* **277**, 1975–1978 (1977).
5. D. E. Kuehner, H. W. Blanch, J. M. Prausnitz. *Fluid Phase Equilib.* **116**, 140 (1996).
6. A. Daanoun, C. F. Tejero, M. Baus. *Phys. Rev. E* **50**, 2913 (1994).
7. A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. King, G. B. Benedek. *Proc. Nat. Acad. Sci. USA* **98**, 6116–6120 (2001).
8. M. D. Serrano, O. Galkin, S. T. You, B. R. Thomas, R. L. Nagel, R. E. Hirsch, P. G. Vekilov. *J. Crystal Growth* **232**, 368 (2001).
9. O. Galkin, K. Chen, R. L. Nagel, R. E. Hirsch, P. G. Vekilov. *Proc. Nat. Acad. Sci. USA* **99**, 8479 (2002).
10. D. R. Booth, et al. *Nature* **385**, 787 (1997).
11. R. M. Murphy and M. R. J. Pallitto. *Struct. Biol.* **130**, 109 (2000).
12. E. J. W. Verwey and J. T. G. Overbeek. *Theory of the Stability of Lyophobic Colloids*, Elsevier, New York (1948).
13. I. Sogami and N. Ise. *J. Chem. Phys.* **81**, 6320 (1984).
14. J. Z. Wu, D. Bratko, J. M. Prausnitz. *Proc. Nat. Acad. Sci. USA* **95**, 15169 (1998).
15. D. Bratko, A. Striolo, J. Z. Wu, H. W. Blanch, J. M. Prausnitz. *J. Phys. Chem. B* **106**, 2714–2720 (2002).
16. A. Striolo, D. Bratko, H. Blanch, J. Prausnitz. *J. Chem. Phys.* **16**, 7733 (2002).
17. M. P. Tombs and A. R. Peacocke. *The Osmotic Pressure of Biological Macromolecules*, Clarendon, Oxford (1974).
18. M. J. Grimson. *J. Chem. Soc. Faraday Trans.* **79**, 817 (1983).
19. R. A. Curtis, C. Steinbrecher, M. Heinemann, H. W. Blanch, J. M. Prausnitz. *Biophys. J.* **98**, 249 (2002).
20. J. J. Grigsby, H. W. Blanch, J. M. Prausnitz. *Biophys. Chem.* **91**, 231 (2001).
21. J. R. Lundblad, M. Laurance, R. H. Goodman. *Mol. Endocrinol.* **10**, 607 (1996).
22. W. J. Checovich, R. E. Bolger, T. Burke. *Nature* **373**, 254 (1995).

23. D. Chandler. *Introduction to Modern Statistical Mechanics*, Oxford University Press, New York (1987).
24. C. A. Anderson, J. F. M. Niesen, H. W. Blanch, J. M. Prausnitz. *Biophys. Chem.* **84**, 177–188 (2000) and C. A. Anderson. Dissertation, University of California, Berkeley (2001).
25. A. George, et al. *Macromol. Crystallogr. A*, 100–110 (1997); A. George and W. W. Wilson. *Meth. Enzymol.* 276 (1997).
26. P. M. Tessier, A. M. Lenhoff, S. I. Sandler. *Biophys. J.* **82**, 1620 (2002) and C. Teske, Ph.D. dissertation, University of California, Berkeley (2003).
27. N. von Solms, C. Anderson, J. M. Prausnitz. *AIChE J.* **48**, 1292 (2002).
28. D. J. Bell, M. Hoare, P. Dunnill. “The formation of protein precipitates and their centrifugal recovery”, in *Advances in Biochemical Engineering/Biotechnology*, A. Frechter (Ed.), **26**, Downstream Processing, Springer, Berlin (1983).