Polymerization–Depolymerization of Tobacco Mosaic Virus Protein. V. Osmotic Pressure Theory for Hydrated Proteins*

Max A. Lauffer

ABSTRACT: The equation for the osmotic pressure of a "hydrated" protein was derived from the equilibrium conditions of Gibbs and the definition of hydration of Lauffer [Lauffer, M. A. (1964), *Biochemistry 3*, 731]. Highly accurate mathematical approximations were made in the derivation, but the only physical assumption was incompressibility of solvent. The meaning of the molecular weight obtained from osmotic pressure studies on "hydrated" protein depends only on the way in which concentration is defined. If it is defined in terms of weight of dried isoionic protein, the molecular weight obtained refers to anhydrous un-ionized protein. The second virial coefficient contains, in addition to the usual Donnan and interaction plus excluded volume terms, a term proportional to the degree of hydration. Under normal circumstances, this should contribute even less than the excluded volume effect to the total virial coefficient. No contribution involving salt–salt interaction appears explicitly in terms lower than the fourth virial coefficient.

The protein extracted from tobacco mosaic virus either by alkaline degradation or by treatment with concentrated acetic acid undergoes reversible endothermic polymerization into rodlike particles resembling the original virus. Detailed descriptions of many aspects of this reaction have been published (Lauffer *et al.*, 1958; Ansevin and Lauffer, 1959, 1963; Ansevin *et al.*, 1964; Stevens and Lauffer, 1965). One of its features is that water is released. It is, therefore, clear that a study of the mechanism of polymerization of tobacco mosaic virus (TMV) protein must involve an understanding of the meaning and behavior of "hydrated" protein.

Banerjee and Lauffer (1966) have carried out osmotic pressure studies on dilute solutions of tobacco mosaic virus protein in order to investigate the beginning stages of the polymerization process. To interpret these results, it is necessary to have an exact theory for the osmotic pressure of hydrated proteins. An exhaustive thermodynamic treatment of the theory of osmotic pressure as applied to protein solutions was published by Scatchard (1946). Simplification can be achieved without loss of rigor by applying the definition of protein hydration previously published by the author (Lauffer, 1964).

Theory

For the sake of simplicity, consider a system composed of water, 1, protein, 2, and potassium chloride, 3. If n_2' moles of uncharged "hydrated" protein is added to vessel I, which initially contains an aqueous potassium chloride solution of exactly the same composition as that in vessel II, eq 1–3 must be obeyed (Lauffer, 1964). The subscripts refer to "free" water, potassium chloride, potassium ion, and chloride ion, the superscripts ′ and ″ refer to vessels I and II, respectively, and N, a, and f mean mole fraction, activity, and activity coefficient, respectively.

$$\frac{N_1'}{N_1''} = \frac{N_3'}{N_3''} \tag{1}$$

$$\frac{a_1'}{a_1''} = \frac{a_3'}{a_3''} = \frac{\sqrt{a_+'a_-'}}{\sqrt{a_+''a_-''}} \tag{2}$$

$$\frac{f_1'}{f_1''} = \frac{f_3'}{f_3''} = \frac{\sqrt{f_+'f_-'}}{\sqrt{f_+''f_-''}} \tag{3}$$

Equations 1–3 are identical equalities, dependent only on definition.

Next, let charged instead of uncharged protein be added to vessel I. The situation is now slightly more complex. For reasons of convenience, define 1 mole of component 2 as $K_x P_r - x/2KCl$, in the manner proposed by Scatchard (1946). With this definition, adding n_2' moles of "hydrated" protein amounts to adding n_2' moles of anhydrous potassium protein salt, withdrawing $n_2'x/2$ moles of KCl, and adding sufficient water to satisfy eq 3. The fact that one ordinarily does not know how much water to add and how much KCl to take away is of no consequence, because ultimately solution I will be equilibrated across a semipermeable

* From the Department of Biophysics, University of Pittsburgh, Pittsburgh, Pennsylvania. *Received October 26, 1965; revised March 29, 1966.* Publications I–IV of this series are Ansevin and Lauffer (1963), Lauffer (1964), Ansevin *et al.* (1964), and Stevens and Lauffer (1965), respectively. This is publication No. 116 of the Department of Biophysics, University of Pittsburgh. Work was supported by a U. S. Public Health Service Grant (GM 10403-03).

membrane with a large volume of external solution and the required adjustments will then take place.

When one adds uncharged colloid it does not make any difference whether one defines hydration as that amount of water which must be added along with anhydrous protein to satisfy eq 2 or 3, for the same amount satisfies both. However, as is shown below, it makes a difference when the colloid is charged. We have specified that eq 3 be satisfied in the definition of hydration. The following equations show the relationships between mole fractions, activities, and activity coefficients in vessels I and II following the operation described above.

$$\frac{N_+''N_-''}{(N_1'')^2} = \left(\frac{n_3''}{n_1''}\right)^2 = \left(\frac{n_3'}{n_1'}\right)^2 \tag{4}$$

$$\frac{N_+'N_-'}{(N_1')^2} = \left(\frac{n_3'}{n_1'}\right)^2 - \left(\frac{xn_2'}{2n_1'}\right)^2 \tag{5}$$

$$\frac{f_+'f_-'}{(f_1')^2}\left[\left(\frac{n_3'}{n_1'}\right)^2 - \left(\frac{xn_2'}{2n_1'}\right)^2\right] = \frac{a_+'a_-'}{(a_1')^2} \neq \frac{a_+''a_-''}{(a_1'')^2} = \frac{f_+''f_-''}{(f_1'')^2}\left(\frac{n_3''}{n_1''}\right) \tag{6}$$

The product of activities on the left side of the inequality in eq 6 can be made equal to the product of activities on the right side by changing slightly the concentration of potassium chloride in the outside vessel, which will result in a small change in activity coefficients. One produces solution III by decreasing the molality of potassium chloride of II until eq 7 is satisfied.

$$\frac{a_+'a_-'}{(a_1')^2} = \frac{a_+'''a_-'''}{(a_1''')^2} \tag{7}$$

The extent to which the molality of KCl must be decreased, Δm_3, can be evaluated as follows. First, consider the change in activity coefficients. In general, in a binary system, $\ln f_i$ can be expressed as a function of solute molality, m_i, by a power series. A convenient way to write this is in the form of eq 8 where $\beta_{ii} = \partial \ln f_i/\partial m_i$, β_{iii} is the second derivative, etc., and superscript ^{0} means the value at infinite dilution.

$$\ln f_i = \beta_{ii}^0 m_i + \frac{\beta_{iii}^0}{2} m_i^2 + \frac{\beta_{iiii}^0}{6} m_i^3 + \frac{\beta_{iiiii}^0}{24} m_i^4 + \text{etc.} \tag{8}$$

Ordinarily, a similar equation is written for f_i defined on the practical basis, $f_i \equiv a_i/m_i$. However, throughout the present development, f_i is defined on the rational basis, $f_i \equiv a_i/N_i$. The coefficients defined on the rational and on the practical basis differ slightly, especially at high molalities; for protein the difference is negligible. Nevertheless, the only requirement for writing eq 8 is that f_i, however defined, be a function of m_i, which it is. The various β values will have values slightly, but for protein not measurably, different from those encountered in the practical system. For multicomponent systems, additional terms must be added to express the effect of other components on f_i. From the Gibbs–Duhem equation one can obtain eq 9.

$$\frac{\partial}{\partial m_3} \ln\left(\frac{f_+f_-}{f_i^2}\right) = 2\left(1 + 2\frac{n_3}{n_1}\right)\frac{\partial \ln \sqrt{f_+f_-}}{\partial m_3} \equiv 2\alpha\beta_{33} \tag{9}$$

Strictly speaking, $\alpha \equiv (1 + 2n_3/n_1)$ is a variable depending on composition. For a 0.1 m solution of KCl, it has a value of 1.0036. As will be seen later, the salt concentrations on the two sides of the membrane differ very little at equilibrium. Ordinarily, the corresponding values of α will differ by <1 part in 10,000. Thus, no appreciable error is introduced by treating α as a constant for each experiment. Therefore

$$\ln\left[\frac{f_+'''f_-'''}{(f_1''')^2}\Big/\frac{f_+''f_-''}{(f_1'')^2}\right] = \int_{II}^{III} 2\alpha\beta_{33}dm_3 = -2\alpha\bar{\beta}_{33}\Delta m_3 \tag{10}$$

In eq 10

$$-\Delta m_3 \equiv m_3''' - m_3'' = m_3''' - m_3'$$

$$\bar{\beta}_{33} \equiv \left(\int_{II}^{III} \beta_{33}dm_3\right)\Big/ \Delta m_3$$

From eq 3 and 10 one obtains eq 11.

$$\frac{f_+'''f_-'''}{(f_1''')^2}\Big/\frac{f_+'f_-'}{(f_1')^2} = \frac{f_+'''f_-'''}{(f_1''')^2}\Big/\frac{f_+''f_-''}{(f_1'')^2} = e^{-2\alpha\bar{\beta}_{33}\Delta m_3} \doteq 1 - 2\alpha\bar{\beta}_{33}\Delta m_3 \tag{11}$$

The approximation in eq 11 is very accurate when $\bar{\beta}_{33} \cdot \Delta m_3 < 0.001$.

Next, eq 6, 7, and 11 can be solved, since

$$m_i \equiv k\frac{n_i}{n_1} \cong 55.5\frac{n_i}{n_1}$$

to obtain

$$\frac{f_+'f_-'}{(f_1')^2}\left[(m_3')^2 - \left(\frac{x}{2}m_2'\right)^2\right] = \frac{f_+'f_-'}{(f_1')^2}(1 - 2\alpha\bar{\beta}_{33}\Delta m_3)(m_3''')^2$$

and then the square root is taken to give eq 12.

$$m_3' - \left[\frac{x^2(m_2')^2}{8m_3'} + \frac{x^4(m_2')^4}{128(m_3')^3} + \cdots\right] = (1 - \alpha\bar{\beta}_{33}\Delta m_3)m_3''' \tag{12}$$

Approximations are involved in both the left and right members of eq 12. However, the right member is highly accurate when $\alpha\bar{\beta}_{33}\Delta m_3 < 0.001$ and the series in brackets on the left, which converges rapidly, can be carried to as many terms as is necessary to achieve high accuracy. Let the converging series on the left of eq 12 be represented by $\phi_1(m_2')$, and apply the definition of Δm_3.

$$\Delta m_3 = \frac{\varphi_1(m_2')}{1 + \alpha\bar{\beta}_{33}m_3'''} \tag{13}$$

β_{33} can be calculated for KCl in water at 25° with the aid of the Debye–Hückel theory. Its values range from *ca.* 10 when m_3 is 0.001 to *ca.* 1 when m_3 is 0.1. From eq 12 and 13 it is apparent that when $m_3' > 4xm_2'$, $\Delta m_3 < 0.01m_3'$. For a protein of mol wt 50,000 and valence 25 at 1% concentration, $xm_2' = 0.005$. Thus, when $m_3' > 0.02$, $\Delta m_3 < 0.0002$, and $\alpha\bar{\beta}_{33}\Delta m_3 < 0.001$, since α is nearly one. It is apparent, therefore, that the condition for high precision in the approximations involved in deriving eq 13 can usually be met.

It follows from the criteria of Gibbs that, at equilibrium, every diffusible constituent must obey eq 14, where P_0 is $P' - P'''$ and $[\bar{V}_i]$ is an average partial molal volume equal to

$$\frac{1}{P_0}\int_{P'''}^{P'} \bar{V}_i dP$$

$$\ln \frac{a_i'''}{a_i'} = \frac{P_0[\bar{V}_i]}{RT} \equiv h_i \tag{14}$$

When this equation is applied to KCl and water, one obtains eq 15 as shown by Adair (1937) and by Lauffer (1964)

$$\frac{a_+'a_-'}{(a_1')^2} = \frac{a_+'''a_-'''}{(a_1''')^2} e^{2h_1 - h_3} \tag{15}$$

When KCl is the electrolyte and when the protein solution has an osmotic pressure between 6×10^{-3} and 6×10^{-2} atm, the exponential in eq 15 differs from 1 by 1 part in something between 10^5 and 10^6. Therefore, eq 7 is a highly accurate statement of the condition for equilibrium. Since eq 14 must be obeyed by the solvent, one can write eq 16. Therefore

$$P_0[\bar{V}_1]/RT = \ln a_1'''/a_1' = \ln f_1'''/f_1' + \ln N_1'''/N_1' \tag{16}$$

$$\ln N_1'''/N_1' = \ln \frac{n_1'''(n_1' + 2n_3' + n_2')}{n_1'(n_1''' + 2n_3''')} = \ln \frac{1 + 2\frac{m_3'}{k} + \frac{m_2'}{k}}{1 + 2\frac{m_3'''}{k}}$$

When the definitions of α and of Δm_3 are introduced and the logarithm is expanded, eq 17 is obtained.

$$\ln N_1'''/N_1' \simeq \frac{m_2' + 2\Delta m_3}{k\alpha} \tag{17}$$

The approximation is very accurate. By taking the square root and then the logarithm of eq 11, one obtains eq 18, which is exact.

$$\ln f_1'''/f_1' = -\ln f_1'/f_1'' + \ln \sqrt{f_+'''f_-'''}/\sqrt{f_+''f_-''} + \alpha\bar{\beta}_{33}\Delta m_3 \tag{18}$$

Since

$$\ln \sqrt{f_+'''f_-'''}/\sqrt{f_+''f_-''} = \int_{II}^{III} \beta_{33} dm_3 = -\bar{\beta}_{33}\Delta m_3$$

where $\bar{\beta}_{33}$ is defined precisely as in eq 9, and since $\alpha - 1 = 2m_3''/k \simeq 2m_3'/k$, eq 19 follows.

$$\ln f_1'''/f_1' = -\ln f_1'/f_1'' + 2m_3'/k\bar{\beta}_{33}\Delta m_3 \tag{19}$$

It is a consequence of eq 3 that for "hydrated" component 2 $\partial \ln f_1/\partial n_2 = \partial \ln \sqrt{f_+f_-}/\partial n_2$ and, therefore, of the Gibbs–Duhem eq that

$$f\partial \ln {}_1/\partial n_2 = -\frac{n_2}{n_1 + 2n_3}\partial \ln f_2/\partial n_2$$

$$-\ln f_1'/f_1'' = -\int_0^{n_2'} \frac{\partial \ln f_1}{\partial n_2} dn_2 = \frac{1}{n_1' + 2n_3'}\int_0^{n_2'} n_2 \frac{\partial \ln f_2}{\partial n_2} dn_2 = \frac{1}{k\alpha}\int_0^{m_2'} m_2 \frac{\partial \ln f_2}{\partial m_2} dm_2 \tag{20}$$

By applying eq 8 and integrating, one obtains eq 21.

$$-\ln f_1'/f_1'' = \frac{1}{k\alpha}\left[\frac{\beta_{22}^0}{2}(m_2')^2 + \frac{\beta_{222}^0}{3}(m_2')^3 + \frac{\beta_{2222}^0}{8}(m_2')^4 + \dots\right] \equiv \frac{1}{k\alpha}\varphi_2(m_2') \tag{21}$$

Substitute eq 21 into 19 and then eq 13, 17, and 19 into 16 and let $\rho_1 k[\bar{V}_1] = 1000$. This involves the assumption that $[\bar{V}_1] = \bar{V}_1$, or that solvent is incompressible.

$$P_0\frac{1000\alpha}{\rho_1 RT} = m_2' + \varphi_2(m_2') + 2\frac{(1 + \alpha\bar{\beta}_{33}m_3')}{(1 + \alpha\bar{\beta}_{33}m_3''')}\varphi_1(m_2') \tag{22}$$

Since $m_3' = m_3''' + \Delta m_3$, eq 23 follows.

$$P_0\frac{1000\alpha}{\rho_1 RT} = m_2' + \varphi_2(m_2') + 2\left[1 + \frac{\alpha\bar{\beta}_{33}\Delta m_3}{1 + \alpha\bar{\beta}_{33}m_3'''}\right]\varphi_1(m_2') \quad (23)$$

It is rarely useful to carry terms in m_2' of higher powers than 3. From the definition of $\phi_1(m_2')$ and of Δm_3, it is apparent that the $\bar{\beta}_{33}$ term in eq 23 does not appear in terms in m_2' below the power of 4 and thus will be found in the final equation first in the fourth virial coefficient.

The molality of protein, m_2', means moles of protein/1000 g of "free" solvent (water). However, this quantity cannot be determined operationally. When $\xi m_2'$ is the ratio of "bound"/"free" solvent (water), when M_2 is the anhydrous molecular weight of protein, and when c is the concentration expressed as grams of dry protein per gram of total solvent, an operationally definable quantity,

$$m_2' = \frac{1000c}{M_2}(1 + m_2') = \frac{1000c}{M_2} + \xi\left(\frac{1000c}{M_2}\right)^2 + \xi^2\left(\frac{1000c}{M_2}\right)^3 + \cdots$$

Furthermore, ξ can be expressed as $\xi_0 + \xi' c +$ etc. When these expressions for m_2' and ξ and the expressions for $\phi_1(m_2')$, $\phi_2(m_2')$, and Δm_3 are introduced into eq 23, eq 24 is obtained. Terms in higher powers of c can be retained whenever they are needed.

$$\frac{P_0}{c}\frac{\alpha}{\rho_1 RT} = \frac{1}{M_2} + \frac{1000}{M_2^2}\left(\xi_0 + \frac{\beta_{22}^0}{2} + \frac{x^2}{4m_3'}\right)c + \frac{(1000)^2}{M_2^3}\left[\xi_0^2 + \frac{M_2\xi'}{1000} + \frac{\beta_{222}^0}{3} + \xi_0\left(\frac{x^2}{2m_3'} + \beta_{22}^0\right)\right]c^2 + \cdots \quad (24)$$

From eq 24 it is evident that the second virial coefficient, B, equals

$$\frac{1000}{M_2^2}\left(\xi_0 + \frac{\beta_{22}^0}{2} + \frac{x^2}{4m_3'}\right)$$

Discussion

Equation 23 is physically complete in that no terms have been omitted. The only physical assumption made in deriving it is that water is incompressible, a negligible error. The equation is, however, inexact because numerous mathematical approximations were introduced. As long as $m_3' > 4xm_2'$, none of the approximations leads to significant error. The equation is, therefore, highly accurate, even for relatively concentrated protein solutions. Higher order terms can easily be introduced to go to even higher concentrations.

β_{22}^0 in eq 24 refers to hydrated protein. Involved in it are excluded volume and protein–protein interaction. The second virial coefficient has a term not exposed by other treatments, namely, ξ_0, or the hydration factor. The physical meaning of ξ_0 is the weight of solvent "bound" by 1 mole of component 2, divided by 1000. For spherical particles with no interaction, $\beta_{22}^0/2$ is equal to the molar volume of protein divided by (250 times the specific volume of the solvent). Nonspherical particles have larger calculated values for this excluded volume term. In the usual case, therefore, ξ_0 will be considerably smaller than $\beta_{22}/2$. Tanford (1961) has shown that the calculated value of the Donnan term, $x^2/4m_3'$, greatly exceeds the excluded volume term for most protein solutions.

It can be seen from eq 24 that one obtains the anhydrous molecular weight when one carries out osmotic pressure studies on hydrated protein. Because m_2' in eq 23 has the same value whether the protein is *considered* to be an ion, a salt, the component defined by Scatchard (1946), the isoionic particle, etc., and since, at infinite dilution, $m_2' = 1000c/M_2$, the meaning of M_2 is dependent solely on the operation involved in determining c; *e.g.*, if the operation can be reduced ultimately to weighing dried isoionic protein and water, regardless of intermediate operations such as measuring specific refractive increments and obtaining c refractrometrically, M_2 will be the anhydrous molecular weight of un-ionized protein. Furthermore, if any mathematical definitions other than those used here lead to eq 24 or even to one which agrees at infinite dilution, the same conclusion must be drawn. Tanford's (1961) discussion is not inconsistent with this, but does not emphasize the point that the meaning of M_2 is dependent solely on the operation used to measure c.

In contrast with eq 35 of Scatchard (1946) or eq 14–25 of Tanford (1961), our expression for B contains no explicit β_{23} or β_{33} terms. However, ξ is a function of solvent activity (Lauffer, 1964) and therefore of β_{33}. It is also a function of β_{21} and β_{23} when they are defined in terms of anhydrous protein, as do Scatchard and Tanford.

It is necessary to consider the effect of substituting a buffer for KCl in the solution. The simplest case would be a buffer composed of a uni-univalent salt, MA and the largely undissociated acid, HA. First, consider the effect of replacing KCl by MA. Component 2 will now be defined as $M_zP_r - x/2MA$. The amount of water that must be added along with $M_zP_r - x/2MA$ to satisfy eq 3 might be different from that for $K_xP_r - x/2KCl$. Equations 23 and 24 will still be valid, but the amount of "hydration" might be different.

When the requirements specified by Lauffer (1964) are fulfilled and HA is added as component 4, the effect will be to substitute $(\alpha\bar{\beta}_{33} + \bar{\beta}_{43}n_4/n_1)$ for $\alpha\bar{\beta}_{33}$ in eq 13. This will change the value of Δm_3. However, this change will always be small and can easily be made negligible. A substance with a salting out constant of 0.2, a representative value for nonionized solutes of low molecular

weight, would have a value of β_{43} of *ca.* 0.5, a magnitude half the minimum previously mentioned for β_{33}. When m_4 is 0.1, n_4/n_1 is $^1/_{550}$. Since α is *ca.* 1, $n_4/n_1\beta_{43}$ would be $<0.001\alpha\beta_{33}$. Dilution of a buffer increases β_{33} and decreases n_4/n_1. Thus the relative contribution of n_4/n_1 β_{43} can always be made as small as desired by reducing buffer concentration. Therefore, a buffer can replace KCl without affecting seriously the accuracy of the final equations.

An interesting consequence of eq 23 is the conclusion that the interaction term, $\phi_2(m_2')$, when negative, cannot exceed m_2' in absolute magnitude. From eq 20 and 21 it is seen that

$$\varphi_2(m_2') = \int_0^{m_2'} m_2 \frac{\partial \ln f_2}{\partial m_2} \mathrm{d}m_2$$

For a process which involves only the addition of "hydrated" component 2 to a solution, $(\partial \ln f_2/\partial m_2)\mathrm{d}m_2 = \mathrm{d} \ln f_2$. Therefore,

$$m_2' + \varphi_2(m_2') = \int_0^{m_2'} \mathrm{d}m_2 + \int_0^{m_2'} m_2 \mathrm{d} \ln f_2 = \int_0^{m_2'} m_2 \mathrm{d} \ln (f_2 m_2)$$

Since activity and mole fractions must be zero or positive, activity coefficients must be zero or positive. When f_2 is between 0 and 1, corresponding to negative values of $\phi_2(m_2')$, m_2 is always $\geqslant f_2 m_2$. Therefore,

$$\int_0^{m_2'} m_2 \mathrm{d} \ln f_2 m_2 \geqslant \int_0^{m_2} f_2 m_2 \mathrm{d} \ln (f_2 m_2) = f_2' m_2' \geqslant 0$$

It has thus been proved that $\phi_2(m_2')$, when negative, cannot exceed m_2' in absolute magnitude.

References

Adair, G. S. (1937), *Trans. Faraday Soc. 33*, 1106.

Ansevin, A. T., and Lauffer, M. A. (1959), *Nature 183*, 1601.

Ansevin, A. T., and Lauffer, M. A. (1963), *Biophys. J. 3*, 239.

Ansevin, A. T., Stevens, C. L., and Lauffer, M. A. (1964), *Biochemistry 3*, 1512.

Banerjee, K., and Lauffer, M. A. (1966), *Biochemistry 5*, 1957.

Lauffer, M. A. (1964), *Biochemistry 3*, 731.

Lauffer, M. A., Ansevin, A. T., Cartwright, T. E., and Brinton, C. C. (1958), *Nature 181*, 1338.

Scatchard, G. (1946), *J. Am. Chem. Soc. 68*, 2315.

Stevens, C. L., and Lauffer, M. A. (1965), *Biochemistry 4*, 31.

Tanford, C. (1961), Physical Chemistry of Macromolecules, New York, N. Y., Wiley, p 229.