Relation between the Solubility of Proteins in Aqueous Solutions and the Second Virial Coefficient of the Solution

C. Haas and J. Drenth*

Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG, Groningen, The Netherlands

W. William Wilson

Department of Chemistry, Mississippi State University, Mississippi State, Mississippi 39762

Received: October 12, 1998; In Final Form: January 26, 1999

In recent publications it was pointed out that there is a correlation between the observed values of the solubility of proteins in aqueous solutions and the second virial coefficient of the solution. In this paper we give a theoretical explanation of this relation. The derived theoretical expression describes the experimentally observed relation between solubility and virial coefficient quite accurately. It is concluded that a variation of the crystallization conditions has little effect on the anisotropy or the range of the interactions between the protein molecules. Analysis of the data for lysozyme indicates a strong anisotropy of the interactions between the molecules.

Introduction

Stimulated by the need to grow diffraction quality protein crystals for X-ray structure determination, there is a growing interest in the thermodynamics and the phase diagram of protein−water solutions, because the phase diagram is the basis for all phase separation processes. It is expected that a better understanding of the phase diagram and of the underlying interactions between the protein molecules will make it possible to choose optimal conditions for crystal growth.

George et al.[1,2] observed that for protein solutions which are suitable for crystal growth, the second virial coefficient B falls in a narrow range, the so-called crystallization slot. In recent publications[3,4] it was shown that there is a theoretical relation between the location of the liquid−liquid immiscibility region in the phase diagram and the value of the second virial coefficient B. It was pointed out that the range of B values of the crystallization slot corresponds closely to the conditions for the presence of a liquid−liquid immiscibility region in the phase diagram.[3,4] This was taken as an indication that a mechanism with nucleation via a liquid−liquid phase separation as a first step is a favorable mechanism for crystal growth.[4−6]

Thus, in addition to the solubility of the protein, the second virial coefficient of the solution is also an important parameter for crystal growth. However, the solubility and second virial coefficient are not independent parameters. In recent publications[2,7,8] it was shown that there is a correlation between the protein solubility in aqueous solution and the second virial coefficient. Intuitively, this is perhaps not unexpected, as both the solubility and the second virial coefficient are determined by the interaction between the protein molecules. However, the relation is by no means simple or trivial. This is because the solubility depends on the binding energy between protein molecules at a short distance from one another in the crystal for very specific orientations of the molecules with respect to each other. On the other hand, the second virial coefficient is a statistical average over all distances and orientations of two molecules in the liquid phase, with each configuration weighted by a Boltzmann factor. Nevertheless, the observed correlation suggests that there must be a simple theoretical relation between the solubility and the second virial coefficient. In this paper, we derive such a relation, using a simple model for the interaction between the protein molecules.

Theoretical Considerations

In a recent publication[6] we discussed the thermodynamic properties of protein−water solutions. It was shown that the characteristic features of the phase diagram, such as the (metastable) liquid−liquid immiscibility region, are described in a satisfactory way with the following expressions for the Gibbs free energy of the crystalline and the liquid phase:

$$G_c = (1/\Omega)\phi_c g_c \tag{1}$$

$$G_l = (1/\Omega)\left[(\phi^2/\phi_c)g_l + kT\phi \ln \phi - kT\phi \ln m - kT\left\{\frac{\phi - 6\phi^2 + 4\phi^3}{(1-\phi)^2}\right\}\right] \tag{2}$$

G_c is the Gibbs free energy per unit volume of a protein crystal. The protein volume fraction in the crystal is ϕ_c; in most protein crystals ϕ_c is about 0.5. M is the molecular weight and ρ is the density of the protein ($\rho \cong 1.36$ g cm^{-3}). Ω is the volume of a protein molecule and $m = M/18\rho$ is the number of water molecules that can be placed in the volume of one protein molecule. G_l is the Gibbs free energy per unit volume of the aqueous solution with protein volume fraction ϕ. The last three terms in the expression for G_l represent the entropy of mixing, deduced from numerical calculations for hard spheres.[9] The interaction between the protein molecules in the crystal and in the solution is expressed by the parameters g_c and g_l, respectively.

* To whom correspondence should be addressed.

10.1021/jp984035l CCC: $18.00
Published on Web 03/20/1999

Using standard thermodynamic relations one can show (see ref 3) that the second virial coefficient B is given by

$$B = (1/M\rho)[4 + (g_l/kT\phi_c)] \quad (3)$$

The solubility ϕ_s of the protein can be calculated from eqs 1 and 3.[4] For small values of the solubility ϕ_s one obtains

$$\phi_s = m \exp(g_c/kT) \quad (4)$$

In order to derive a relation between B and the solubility ϕ_s it is necessary to relate both quantities to an interaction potential between the protein molecules. In many publications in the literature the interaction between protein molecules is modeled by an isotropic potential which depends only on the distance between the molecules. However, even so-called globular proteins are not perfect spheres, but have folds and crevices on the surface of the molecule. Moreover, there is an inhomogeneous distribution of polar and hydrophobic groups at the surface. As a consequence, the interaction between two protein molecules is a complicated, highly anisotropic function of the orientation of the molecules with respect to one another:[10]

$$V = V(r,\theta,\varphi,\alpha,\beta,\gamma) \quad (5)$$

The vector between the centers of gravity of the two molecules is **r**; r is the length of **r** and the orientation of **r** with respect to a coordinate system fixed in the first of the two protein molecules is specified by the polar angles θ and φ. The orientation of the second protein molecule is represented by the Eulerian angles α, β, γ.

The parameter for the interaction between protein molecules in the crystal, g_c, is given by the sum of the interaction energies at equilibrium distances r_i for specified orientations of the two molecules as they occur in the crystal:

$$g_c = (1/2)\sum_{i=1}^{i=z} V(r_i,\theta_i,\varphi_i,\alpha_i,\beta_i,\gamma_i) \quad (6)$$

where z is the coordination number, i.e., the number of nearest-neighbor protein molecules in the crystal. Generally, the coordination number z will depend on the crystal structure. For a simple lattice with molecules at the corners of a primitive unit cell, $z = 6$.

For isotropic spherical molecules the relation between the second virial coefficient and the interaction potential is given by[11]

$$B = (1/2M\rho\Omega)\int_0^\infty r^2\,dr\int_0^\pi \sin\beta\,d\beta\int_0^{2\pi} \times d\varphi[1 - \exp(-V(r)/kT)] \quad (7)$$

For anisotropic interactions between nonspherical molecules this is easily generalized:

$$B = (1/2M\rho\Omega)\int_0^\infty r^2\,dr\int_0^\pi \sin\theta\,d\theta\int_0^{2\pi} \times d\varphi(1/8\pi^2)\int_0^{2\pi}d\alpha\int_0^\pi \sin\beta\,d\beta\int_0^{2\pi} \times d\gamma[1 - \exp(-V(r,\theta,\varphi,\alpha,\beta,\gamma)/kT)] \quad (8)$$

In order to derive a simple relation between B and ϕ_s, it is necessary to make further assumptions, and to introduce a specific model for the interaction potential. We will model the interactions with a square well potential. For an isotropic square well potential $V(r) = \infty$ for $r < 2a$, $V(r) = -\epsilon$ for $2a < r < 2a\lambda$, and $V(r) = 0$ for $r > 2a\lambda$. The radius of the spherical molecules is a, and $\Omega = 4\pi a^3/3$. For $r < 2a$ there is the hard sphere repulsion. The attractive potential well extends from $r = 2a$ to $r = 2a\lambda$. The parameter λ is a measure of the range of the interaction; the molecules interact if the distance between the molecules is smaller than $2a(\lambda - 1)$. Thus $2a(\lambda - 1)$ is the range of the interaction between the molecules.

For anisotropic interactions between nonspherical molecules the parameters λ and ϵ will depend on the mutual orientation of the two molecules. For the interaction parameter g_c we can write

$$g_c = -\frac{1}{2}\sum_{i=1}^{i=z}\epsilon_i = -(z/2)\epsilon_0 \quad (9)$$

where ϵ_i is the depth of the potential well for the interaction with near-neighbor molecule i in the crystal and ϵ_0 is the average value of ϵ_i for the z nearest neighbors.

The second virial coefficient can be expressed in terms of the ϵ and λ parameters. For an isotropic square well potential one obtains from eq 7

$$B = (4/M\rho)[1 - (\lambda^3 - 1)\{\exp(\epsilon/kT) - 1\}] \quad (10)$$

For anisotropic interactions one has to take into account that the values of ϵ and λ depend on the orientation of the molecules. It has been pointed out[10] that due to the Boltzmann factor the configurations with a large, negative value of V will strongly dominate in the integration over all orientations. Because of this property the expression for B for an attractive potential can be approximated by the following equation:

$$B \cong (4/M\rho)[1 - p(\lambda^3 - 1)\{\exp(\epsilon_0/kT) - 1\}] \quad (11)$$

The factor p represents the anisotropy; it is a result of the fact that the strong attractive interactions with average strength ϵ_0 occur only for a restricted range of orientations of the molecules with respect to one another. This range is accounted for by the factor p. For isotropic interactions $p = 1$, but in the case of strong anisotropy p will be smaller.

The approximations used to derive eq 11 are valid for sufficiently attractive interaction potentials; only then does the Boltzmann factor make certain that preferred configurations with strong bonding dominate. For repulsive and for weakly attractive interactions, that is no longer the case.

From eqs 4 and 9 we obtain

$$\epsilon_0 = -(2g_c/z) = -(2kT/z)\ln(\phi_s/m) \quad (12)$$

Substituting this expression for ϵ_0 in eq 11 we find the desired relation between the solubility ϕ_s and the second virial coefficient B:

$$B = (4/M\rho)[1 - A\{(\phi_s/m)^{-(2/z)} - 1\}] \quad (13)$$

where

$$A = p(\lambda^3 - 1) \quad (14)$$

The only free parameter in eq 13 is A, which depends on the anisotropy p and the range $2a(\lambda - 1)$ of the interactions. For strong anisotropy (small p) and short-range interactions ($(\lambda - 1) \ll 1$) the parameter A is small. Figure 1 shows the relation between ϕ_s and B for a coordination nunber $z = 6$ and for several values of A. In Figure 2 the relation between ϕ_s and B is shown for two values of the coordination number: $z = 3$ and $z = 6$. The value of A was chosen in such a way that the curves for z

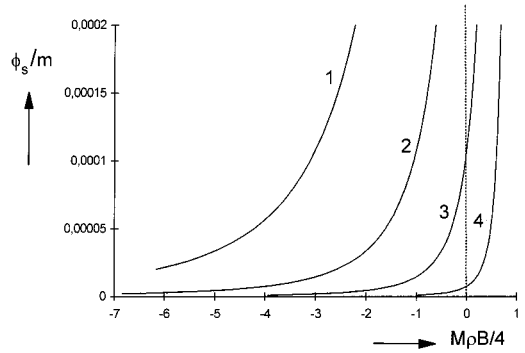

Figure 1. Solubility ϕ_s (volume fraction protein) as a function of $M\rho B/4$, where B is the second virial coefficient, calculated for a square well potential. The coordination number $z = 6$. The solubility is scaled with a factor $m = M/18\rho$ for the volume of the protein molecule. (1) $A = 0.2$, (2) $A = 0.1$, (3) $A = 0.05$, (4) $A = 0.02$.

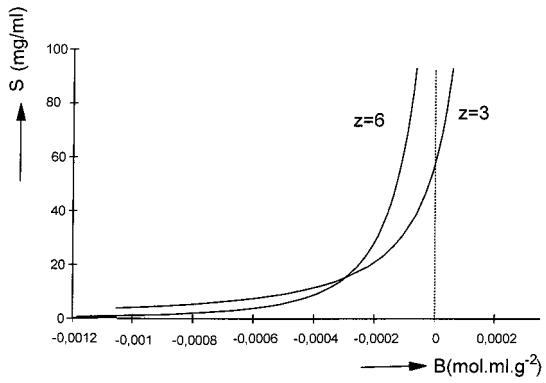

Figure 2. Relation between the solubility ϕ_S and the second virial coefficient B calculated for a square well potential. The two curves are calculated for the parameter values $z = 3$; $A = 0.001\,76$ and $z = 6$; $A = 0.067$. The calculation is for a protein with $M = 14\,000$ and $\rho = 1.36$ g cm^{-3}.

$= 3$ and $z = 6$ give the same value $M\rho B/4 = -1.43$ for $\phi_s/m = 0.000\,02$. Figures 1 and 2 show that the relation between ϕ_s and B depends strongly on the parameter A which involves the anisotropy and the range of the interaction and depends to a lesser extent on the coordination number.

A square well potential is a highly simplified representation of the interaction between protein molecules. In order to find out what effect the shape of the potential function has on the relation between ϕ_s and B, calculations were carried out for a potential with a quite different form, the so-called Yukawa potential,[12] with $V(r) = \infty$ for $r < 2a$, and for $r > 2a$

$$V(r) = -\epsilon(2a/r)\exp\{-(r - 2a)/d\} \qquad (15)$$

This potential has a sharp minimum $V(r) = -\epsilon$ at $r = 2a$; the interaction decays exponentially with a range d. Using eqs 8 and 12, and taking account of the anisotropy of the interaction by a factor p in the same way as in the calculations for the square well potential, one obtains the following equation for the relation between ϕ_s and B:

$$B = (4/M\rho)[1 - p(\pi/2\Omega)\int_{2a}^{\infty} r^2 dr[\exp\{-(4a/rz)\ln(\phi_s/m)\exp[-(r - 2a)/d]\} - 1]] \qquad (16)$$

With this equation one can calculate by numerical integration ϕ_s as a function of B for given values of the parameters p, z, and d. Results of the calculations (see Figure 3) show that the relation between ϕ_s and B depends strongly on the anisotropy and the range of the interactions, and to a lesser extent on z,

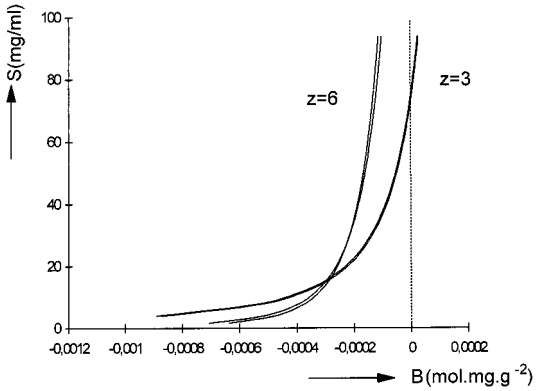

Figure 3. Relation between the solubility ϕ_S and the second virial coefficient B, calculated for a Yukawa potential. The calculation is for a protein with $M = 14\,000$ and $\rho = 1.36$ g cm^{-3}. The two curves for coordination number $z = 3$, calculated with $p = 1$; $d/a = 0.0072$ and $p = 0.1$; $d/a = 0.072$, nearly coincide. The same is true for the two curves for $z = 6$, calculated with $p = 1$; $d/a = 0.133$, and $p = 0.1$; $d/a = 1.33$.

just as was found for the square well potential. We also find that the curves are nearly the same for different values of p, provided that the range d is chosen so that the product pd is constant; this is equivalent to the result for the square well potential where the relation between ϕ_s and B depends only on $A = p(\lambda^3 - 1)$.

It is of interest to compare the results of the calculations for the square well potential (Figure 2) and the Yukawa potential (Figure 3). We find that the curves for ϕ_s as a function of B are nearly the same. We conclude that the relation between solubility and second virial coefficient depends on the range and the anisotropy of the interaction and on the coordination number, but hardly on the shape of the interaction potential.

Comparison with Experimental Data

Recently it was shown experimentally that there is a correlation between the solubility ϕ_s of lysozyme and the second virial coefficient B. Data were reported for a large variety of solvents, containing different salts in different concentrations and for different values of the pH and the temperature.[8] An overview of the results is shown in Figure 4; the solubility S in mg/mL is related to ϕ_s by $S = 10^3\rho\phi_s$. It is found that all data fall approximately on a single curve of ϕ_s versus B.

In order to compare the experimental data with the theoretical relation eq 13, it is necessary to know the value of the coordination number z, and to choose a value for A. In tetragonal lysozyme crystals all lysozyme molecules are equivalent, and each molecule has strong interactions (many hydrogen bonds and many van der Waals contacts) with four neighboring molecules, so that $z = 4$.[13] We found that a good fit between the experimental data and eq 13 is obtained for $A = 0.01$. The curve of the solubility S versus the second virial coefficient B, calculated from eq 13 with $z = 4$ and $A = 0.01$, is shown in Figure 4. We conclude that it is possible to describe the relation between B and ϕ_s quite well with eq 13, with only one adjustable parameter A.

The fact that experimental data for a particular protein in solvents containing different salts with different concentrations and with different values of the pH and the temperature all fall on the same curve indicates that the parameter A must be approximately the same in all cases. This is a rather surprising result. The anisotropy p and the range of the interaction (or at least the product $A = p(\lambda^3 - 1)$) is found to be nearly the same for all solvents and also independent of the temperature and the pH.

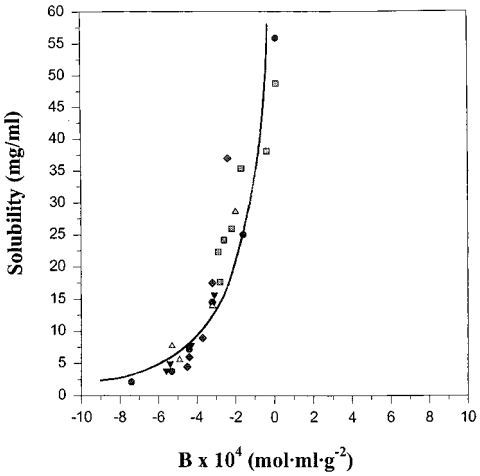

Figure 4. Composite plot of solubility S in mg/mL ($S = 10^3\rho\phi_c$) versus the second virial coefficient B for lysozyme in (○) 0.1 M HAc/NaAc, pH = 4.2, T = 25 °C, various [NaCl]. %(w/v) (ref a); (□) 0.1 M HAc/NaAc, 2.0% (w/v) NaCl, T = 25 °C, various pH (ref a); (△) 0.1 M HAc/NaAc, 2.0% (w/v) NaCl, pH = 4.2, various T (°C) (ref a); (▽) 0.05 M HAc/NaAc, pH = 4.5, T = 18 °C, various [NH_4Cl], (M) (ref b); (◇) 0.02 M HEPES, pH = 7.8, T = 23 °C, various [$MgBr_2$], (M) (ref c). The experimental data are from ref 5. The solubility data are from the following references: (a) Cacioppo, E.; Pusey, M. *J. Crystal Growth* **1991**, *114*, 286; (b) Ries-Kautt, M.; Ducruix, A. *J. Biol. Chem.* **1989**, *264*, 745; (c) Broide, M.; Tominc, T.; Saxowsky, M. *Phys. Rev.* **1996**, *E53*, 6325. The drawn line is calculated from the theoretical relation eq 13 with M = 14 000, ρ = 1.36 g cm^{-3} for z = 4 and A = 0.010.

From the value of A we can calculate the anisotropy factor p for a given value of the range of the interactions $2a(\lambda - 1)$. For lysozyme, with $A = 0.01$ and a = 16 Å, we obtain for a range of interaction $2a(\lambda - 1)$ of 1, 2, and 3 Å the values p = 0.1, 0.05, and 0.03, respectively. This shows that for all realistic values of the range of interactions the anisotropy is quite large. A consequence is that simple models with isotropic interactions, which were introduced to describe colloid systems, are not adequate to describe the properties of protein−water systems. The presence of a metastable liquid−liquid immiscibility region in the phase diagram of colloid systems is considered to be caused by the short range of the interaction between the colloid particles.[3,14,15] The same conjecture has been made for protein−water phase diagrams.[3,4,16−18] However, we like to point out that a factor of equal importance as the range is the anisotropy of the interactions between the protein molecules. In fact, it is not the range of the interactions, but rather the product of the range and the anisotropy factor p, which determines whether the phase diagram will show a metastable region of liquid−liquid immiscibility.

Conclusion

Using a simple model potential for the interaction between protein molecules it was possible to derive a relation between the solubility of the protein and the second virial coefficient of the solution. This relation depends strongly on the anisotropy and on the range of the interaction between the molecules, and to a lesser extent on the coordination number of the molecules in the crystal. Calculations for two different interaction potentials, a square well and a Yukawa potential, gave nearly the same result. This indicates that the shape of the potential has hardly any effect on the relation between the solubility and the second virial coefficient.

The obtained theoretical relation is in excellent agreement with experimental data for lysozyme. From this agreement it is concluded that a variation of the crystallization conditions, i.e., a variation of the type or concentration of the salt in the solvent, of the pH or of the temperature of the solution has little effect on the anisotropy or on the range of the interactions between the protein molecules.

Analysis of the data for lysozyme shows that the interactions between the protein molecules are strongly anisotropic. The presence of a metastable liquid−liquid immiscibility region in the lysozyme−water phase diagram is not only due to the short range of the interactions, but also to the large anisotropy of these interactions.

Acknowledgment. We thank Dr. Gert Vriend for supplying PDB data in a suitable form. This work was supported by the Commission of the European Community under contract B104-CT98-0086. This work was supported in part by NASA grants NAG8-965 and NAG8-1158.

References and Notes

(1) George, A.; Wilson, W. W. *Acta Crystallogr.* **1994**, *D50*, 361.
(2) George, A.; Chiang, Y.; Guo, B.; Arabshaki, A.; Cai, Z.; Wilson, W. W. *Methods Enzymol.* **1997**, *276*, 100.
(3) Lekkerkerker, H. N. W. *Physica* **1997**, *A244*, 227.
(4) Haas, C.; Drenth, J. *J. Phys. Chem.* **1998**, *B102*, 4225.
(5) Wolde, P. R. ten; Frenkel, D. *Science* **1997**, *277*, 1975.
(6) Haas, C.; Drenth, J. *J. Crystal Growth* **1999**, *196*, 388.
(7) Gripon, C.; Legrand, L.; Rosenman, I.; Vidal, O.; Robert, M. C.; Boué, F. *J. Crystal Growth* **1997**, *178*, 575.
(8) Guo, B.; Kao, S.; McDonald, H.; Wilson, W. William; Asanov, A.; Combs, L. L. *J. Crystal Growth* **1999**, *196*, 424.
(9) Carnahan, N. F.; Starling, K. F. *J. Chem. Phys.* **1969**, *51*, 635.
(10) Neal, B. L.; Asthagiri, D.; Velev, O. D.; Kaler, E. W.; Lenhoff, A. M. *J. Crystal Growth* **1999**, *196*, 377.
(11) Landau, L. D.; Lifshitz, E. M. *Statistical Phys.* **1980**, *5* (1).
(12) Cummings, P. T.; Monson, P. A. *J. Chem. Phys.* **1985**, *82*, 4303.
(13) PDB file 193L, Protein Data Bank, Biology Department, Brookhaven National Laboratory, Upton, NY.
(14) Meijer, E. J.; Frenkel, D. *Phys. Rev. Lett.* **1991**, *67*, 1110.
(15) Hagen, M. H. J.; Frenkel, D. *J. Chem. Phys.* **1994**, *101*, 4093.
(16) Rosenbaum, D.; Zamora, P. C.; Zukoski, C. F. *Phys. Rev. Lett.* **1996,** *76*, 150.
(17) Lomakin, A.; Asherie, N.; Benedek, G. B. *J. Chem. Phys.* **1996,** *104*, 1646.
(18) Malfois, M.; Bonneté, F.; Belloni, L.; Tardieu, A. *J. Chem. Phys.* **1996**, *105*, 3290.