The Protein−Water Phase Diagram and the Growth of Protein Crystals from Aqueous Solution

C. Haas and J. Drenth*

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

Received: November 21, 1997; In Final Form: March 3, 1998

The phase diagram of a protein−water system is described with a simple model with parameters for the interaction between protein molecules in the liquid and in the solid phase. The model reproduces essential features of the phase diagram, such as the (metastable) liquid−liquid immiscibility region and the triple point. The calculated phase diagram is used to analyze conditions for crystal growth. It is found that the condition for a crystallization mechanism with nucleation via a liquid−liquid phase separation as a first step corresponds closely to the favorable conditions for crystal growth reported by George and Wilson (*Acta Crystallogr.* **1994**, *D50*, 361), which suggests that nucleation via liquid−liquid phase separation is the prefered mechanism for growing high quality protein crystals.

Introduction

The growth of suitable protein crystals is an essential step in the structure determination of a protein by x-ray diffraction. The crystals are usually grown by a somewhat organized trial and error procedure. Generally, it is not known why a certain set of parameters of temperature, solvent composition, and supersaturation is the optimal one for crystal growth.[1−3] This uncertainty is at least partly due to the fact that a basic understanding of the crystal growing process of proteins from aqueous solution is lacking.

In this paper we want to contribute to a better understanding of the process of protein crystal growth. First we calculate the phase diagram of a protein−water system, which forms the basis for understanding phase separation processes. In the next sections we try to deduce from the phase diagram optimal conditions for the growth of protein crystals.

The Protein−Water Phase Diagram

We derive the phase diagram of a protein−water system on the basis of a simple model with only two parameters, which are characteristic for the interaction between the protein molecules.[4] For the Gibbs free energy per unit volume of an aqueous protein−water solution we use the expression

$$G_\lambda(\phi) = (1/\Omega)\left((\phi^2/\phi_c)g_\lambda + kT\phi \ln \phi - kT\phi \ln m - kT\left\{\frac{\phi - 6\phi^2 + 4\phi^3}{(1-\phi)^2}\right\}\right) \quad (1)$$

For the Gibbs free energy per unit volume of a protein crystal we use the expression

$$G_c = (1/\Omega)\phi_c g_c \quad (2)$$

In these equations, ϕ is the volume fraction of protein. Protein crystals contain a considerable amount of water, and the protein volume fraction ϕ_c in the crystal is ∼0.5.

The Gibbs free energies G_λ and G_c are defined with respect to the pure solvent and the state of the protein molecule in a dilute solution. The solvent mainly consists of water, but usually also contains other components, such as dissolved salts.

We define Ω as the volume of a protein molecule, and ω as the molar volume of water divided by Avogadro's number. Thus, the ratio $m = \Omega/\omega$ is equal to the number of water molecules that can be placed in the volume of one protein molecule. We neglect changes of Ω and ω as a result of changes of temperature or pressure.

The first term in eq 1, which is quadratic in the protein concentration, describes the interaction between protein molecules in the solution. The other terms represent the entropy of mixing of the protein molecules with the solvent molecules. We have used in eq 1 the expression for the entropy proposed by Carnahan and Starling.[5] This expression was deduced from numerical calculations of the entropy of mixing of hard spheres. The expression has been used in the literature to calculate the free energy of colloids[6] and proteins.[7] The parameter g_c describes the interaction between protein molecules in the crystal; its value depends also on the composition of the solvent, because the state of the protein in the dilute solution is taken as the reference state.

Using eqs 1 and 2, it is possible to calculate the phase diagram. We find that for negative values of g_λ there is a region of liquid−liquid phase separation. The compositions of the two coexisting liquid phases ϕ_α and ϕ_β can be calculated from the equations

$$G_\lambda(\phi_\beta) - G_\lambda(\phi_\alpha) = \phi_\beta(\partial G_\lambda/\partial\phi)_{\phi_\beta} - \phi_\alpha(\partial G_\lambda/\partial\phi)_{\phi_\alpha} \quad (3)$$

$$(\partial G_\lambda/\partial\phi)_{\phi_\beta} = (\partial G_\lambda/\partial\phi)_{\phi_\alpha} \quad (4)$$

The two−phase region between ϕ_α and ϕ_β is separated into two parts, an unstable and a metastable region. In the part between the spinodal compositions ϕ_α* and ϕ_β*, the solution is unstable with respect to demixing. In the regions $\phi_\alpha < \phi < \phi_\alpha$* and

* To whom correspondence should be addressed.

S1089-5647(98)00296-X CCC: $15.00
Published on Web 04/28/1998

TABLE 1: Liquid−Liquid Phase Separation as a Function of T/θ[a]

T/θ	ϕ_α	ϕ_β	ϕ_α*	ϕ_β*
1.00	0.1305	0.1305	0.1305	0.1305
0.95	0.067	0. 21	0. 091	0.18
0.90	0.046	0. 25	0. 078	0. 20
0.85	0.032	0. 28	0. 068	0. 22
0.80	0.022	0. 31	0. 060	0. 23
0.75	0.014	0. 33	0. 054	0. 25
0.70	0.0093	0. 36	0. 048	0. 27
0.65	0.0058	0. 38	0. 042	0. 29
0.60	0.0032	0.41	0.038	0.31
0.55	1.5×10^{-3}	0.43	0.033	0.32
0.50	6.3×10^{-4}	0.45	0.029	0.34
0.45	1.9×10^{-4}	0.48	0.026	0.36
0.40	4.6×10^{-5}	0.50	0.022	0.38
0.35	6.3×10^{-6}	0.53	0.020	0.40
0.30	5.0×10^{-7}	0.56	0.017	0.43

[a] The volume fractions of the coexisting liquid phases are ϕ_α and ϕ_β; the spinodal compositions are ϕ_α* and ϕ_β*; θ is a parameter that measures the strength of the interactions between the protein molecules in the solution.

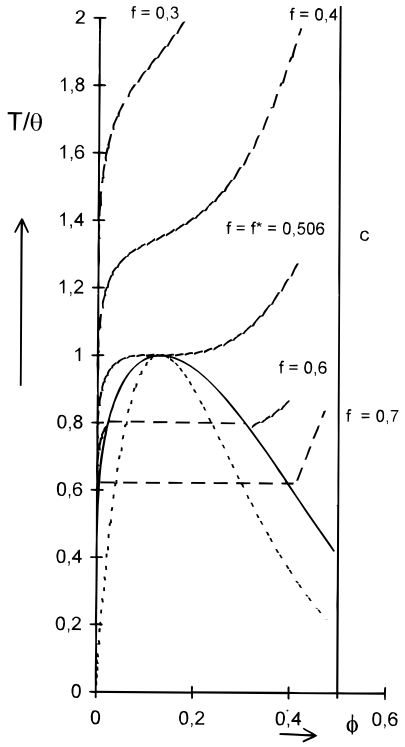

Figure 1. Calculated phase diagram. The figure shows the liquid−liquid two-phase region, with the volume fractions ϕ_α and ϕ_β of the coexisting liquid phases (thick line), and the spinodal compositions ϕ_α* and ϕ_β* (dotted line). The dashed lines are solubility curves for different values of $f = g_\lambda/g_c$ (calculated for $m = 1000$).

$\phi_\beta^* < \phi < \phi_\beta$, the solution is metastable with respect to liquid−liquid phase separation. The spinodal volume fractions ϕ_α* and ϕ_β* can be calculated from the condition

$$\partial^2 G_\lambda/\partial\phi^2 = 0 \tag{5}$$

Equations 3, 4, and 5 were used to calculate values of ϕ_α, ϕ_β, ϕ_α*, and ϕ_β*. These values depend on the temperature T and on the parameter g_λ for the interaction between the protein molecules in the solution. The calculated values of ϕ_α, ϕ_β, ϕ_α*, and ϕ_β* are given in Table 1, and plotted as a function of the temperature in Figure 1. It is found that liquid−liquid phase separation occurs only below a critical temperature T_0. The volume fractions ϕ_α, ϕ_β, ϕ_α*, and ϕ_β* at the critical temperature are equal to the critical concentration, $\phi_0 = 0.1305$.

For convenience we have introduced a parameter θ (with dimension of a temperature) defined by

$$\theta \equiv -\frac{g_\lambda}{10.601 \times k\phi_c} \tag{6}$$

The numerical factor 10.601 is chosen so that at the critical temperature, $T/\theta = 1$; k is the Bolzmann constant. Generally, g_λ and θ depend on the temperature. Therefore, some caution is necessary in using Figure 1, because the T/θ scale is not linear in T. Generally, the critical temperature T_0 can be calculated from the equation

$$T_0 = \theta(T_0) = -\frac{g_\lambda(T_0)}{10.601 \times k\phi_c} \tag{7}$$

The solubility of the solid in the liquid can be calculated from the condition

$$G_c - G_\lambda = (\phi_c - \phi)(\partial G_\lambda/\partial\phi) \tag{8}$$

The solubility ϕ_s depends on the parameters g_λ and g_c, and also on $m = \Omega/\omega$. If the solubilty is small ($\phi_s \ll 1$), it is given in good approximation by

$$\phi_s = m \exp(g_c/kT) \tag{9}$$

We calculated from eq 8 the solubility ϕ_s as a function of T/θ for different values of the ratio $f = g_\lambda/g_c$. The results are given in Figure 1 for a chosen value $m = 1000$. It is found that for values of f smaller than a critical value

$$f^* = \frac{5.30}{3.563 + \ln m} \qquad (\text{for } \phi_c = 0.5) \tag{10}$$

the solubility curve ϕ_s (T) lies above the liquid−liquid two-phase region. This result means that for $f < f^*$, the liquid−liquid two-phase region is metastable; that is, it lies completely in the region where the liquid phase is not stable with respect to a phase separation into a solid phase (ϕ_c) and a liquid phase (ϕ_s). For $f > f^*$, on the other hand, the phase diagram shows a triple point T_t , with equilibrium between a solid phase ϕ_c and two liquid phases with compositions ϕ_α and ϕ_β. The liquid−liquid two-phase region is metastable below T_t and stable above T_t. The triple point temperature T_t is determined by the values of f and θ. Calculated values of T_t/θ as a function of the ratio $f = g_\lambda/g_c$ are shown in Figure 2 for different values of m.

Conditions for Crystal Growth

The process of nucleation and growth of crystals is usually discussed in terms of the classical theory of crystal growth.[8,9] In this theory, several steps can be distinguished. In the first step small nuclei of the crystalline phase are formed by a thermally activated process. If the size of the crystals is below a certain threshold, they are not stable in a supersaturated solution, because of the large contribution of the surface energy; therefore these very small crystals, if formed, will dissolve again and disappear. Only if the crystalline nuclei have obtained a certain critical size, are they stable with respect to dissolution. The second step of the crystallization process is the further growth of the critical nuclei. Finally, the process of Ostwald

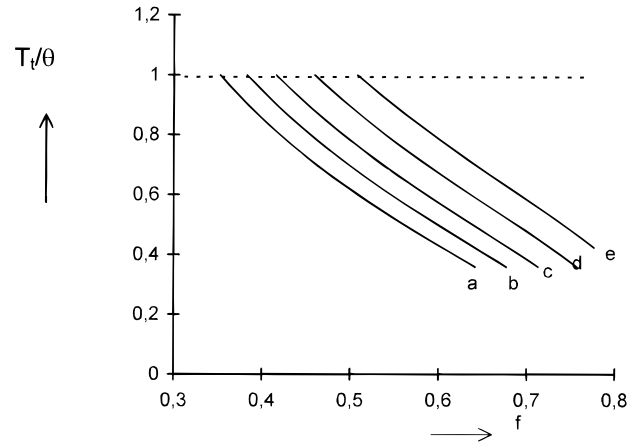

Figure 2. Triple point temperature T_t divided by the parameter θ as a function of the ratio $f = g_\lambda/g_c$, for (a) $m = 10^5$; (b) $m = 3 \times 10^4$; (c) $m = 10^4$; (d) $m = 3 \times 10^3$; (e) $m = 10^3$.

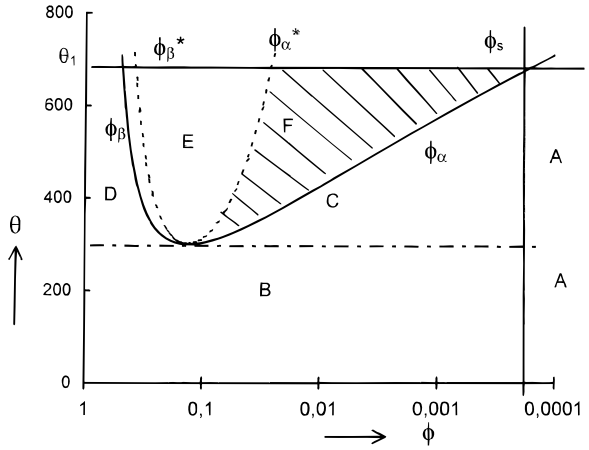

Figure 3. The composition at 300 K of the coexisting liquid phases ϕ_α and ϕ_β, the spinodal compositions ϕ_α* and ϕ_β*, and the solubility ϕ_s of the solid in the liquid are given as a function of the strength θ of the interaction between the protein molecules in the solution at 300 K. Crystal growth by the mechanism of nucleation of small liquid droplets of high protein concentration is possible only in the hatched region F. In region C, the first step of the nucleation process is the forming of local concentration fluctuations in the liquid phase with a high protein concentration. Note that ϕ increases to the left on a logarithmic scale.

ripening may set in, with large crystals growing at the expense of small crystals. We mention that under certain conditions in colloidal systems and also in aqueous protein solutions, amorphous precipitates may form instead of crystals.[10,11]

Recently, it was proposed that the crystallization of proteins from aqueous solutions may start by a phase separation into two liquid phases.[12,13] This mechanism can occur only if the phase diagram shows a metastable liquid−liquid two-phase region. The first step here is the formation of small regions (droplets) with a high protein concentration. These regions can be either metastable droplets that initiate a macroscopic liquid−liquid phase separation or flickering unstable concentration fluctuations. The next step in the nucleation process is the formation of small crystals in these droplets. It was suggested that this mechanism is possibly a suitable procedure for the growth of good protein crystals, adequate for the structure determination by x-ray diffraction. Numerical calculations by tenWolde and Frenkel[13] indicated that the process in which the formation of crystalline nuclei proceeds via the intermediate formation of liquid droplets with a high protein concentration has a lower activation energy and is expected to be more rapid than the direct formation of crystalline nuclei. This is in fact a quite general phenomenon; in many cases a certain transition proceeds more rapidly via a succession of two or more intermediate transitions, each with a low activation energy, than by a single-step transition with a higher activation energy. Thus, one expects that if there is a metastable liquid−liquid two-phase region in the phase diagram, then crystallization in that region will proceed via the formation of liquid droplets with a high protein concentration and not via the much slower process of the direct formation of crystalline nuclei.

In this section we analyze the conditions for crystal growth, using the phase diagram of the protein−water system. In the previous section it was shown that the phase diagram may show a stable or metastable liquid−liquid two-phase region and a triple point. The occurrence of these features depends on the parameters g_λ and g_c for the interaction between protein molecules in the liquid and the solid phase, respectively.

We first discuss a phase diagram without a triple point. In this case it is evident that above the critical temperature T_0, crystallization is only possible via the classical mechanism of the formation of crystalline nuclei. Below T_0, crystallization may also proceed via a mechanism involving as a first step a phase separation in the liquid (i.e., the formation of small droplets with a high protein concentration of about ϕ_β). The nucleation in the liquid phase for liquid−liquid phase separation has been discussed extensively in the literature.[9,14,15] In the region between the spinodal compositions $\phi_\alpha^* < \phi < \phi_\beta^*$, the phase separation proceeds very rapidly, without an activation energy. The rapid phase separation in this region leaves little time for establishing the proper order and the steric orientations of the protein molecules required for crystallization (i.e., desolubilization and self-association are faster than the rate at which molecules achieve the proper orientations). Many small crystals will form if a large number of separate self-association events occur, whereby a large net rate is maintained but the rate for any one nucleation event is low enough to maintain order during the assembly process.

The next case is that of a phase diagram with a triple point. In this case, one should distinguish between temperatures above and below T_t. For $T_t < T < T_0$, there is a stable liquid−liquid two-phase region. In this temperature range, solutions will separate into two phases with compositions ϕ_α and ϕ_β. Crystals can be obtained in this temperature range only from a solution with a very high protein concentration, $\phi > \phi_\beta$. For temperatures below T_t, there is a metastable two-phase region. Thus, for $T < T_t$, the second mechanism, with nucleation by liquid−liquid phase separation as a first step, is also possible. For $T < T_t$ and $\phi_\alpha^* < \phi < \phi_\beta^*$, there will be a rapid phase separation, presumably leading to amorphous precipitates or many small crystals.

In Figure 3 we have given the characteristic compositions ϕ_s (solubility), ϕ_α and ϕ_β (limits of the liquid−liquid two-phase region), and ϕ_α^* and ϕ_β^* (spinodal compositions) at a fixed temperature, the temperature of crystallization, which is usually ∼300 K. We compare in Figure 3 the values of ϕ_α, ϕ_β, ϕ_α^*, and ϕ_β^* for different values of θ at the constant temperature of 300 K. Remember that θ is a measure of the strength of the interaction between the protein molecules in the solution.

Figure 3 is of course directly related to the phase diagram of Figure 1. Region A corresponds to undersaturated solutions, and is of no relevance here. Solutions in region B are supersaturated with respect to the crystalline phase, but because $\theta < 300$ K, there is no liquid−liquid two-phase region at 300 K. Thus, for $\theta < 300$ K, crystallization at 300 K proceeds via

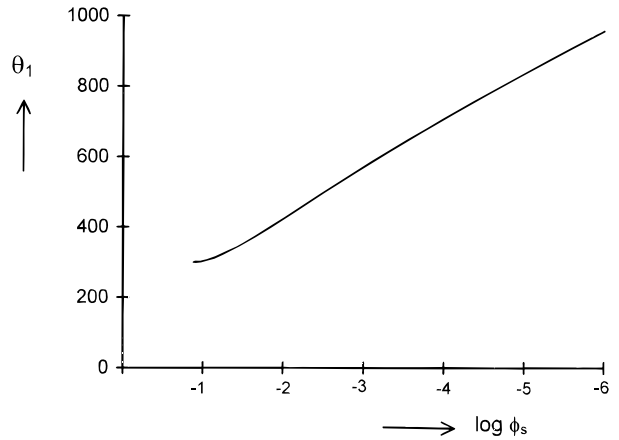

Figure 4. Critical value θ_1 as a function of the solubility ϕ_s.

the classical mechanism of the formation of small crystalline nuclei. In region C, with compositions $\phi_s < \phi < \phi_\alpha$, the solution is stable with respect to a liquid−liquid phase separation. However, because of the presence of a nearby metastable liquid−liquid two-phase region, concentration fluctuations are strongly enhanced. Therefore, in region C, the formation of small unstable regions of the liquid with a high protein concentration requires only a small activation energy, and the formation of these droplets is likely to be the first step of the nucleation process. For compositions in the hatched region *F*, with $\phi_\alpha < \phi < \phi_\alpha$*, the mechanism with nucleation via metastable liquid−liquid-phase separation is possible.

We find from Figure 3 that at a certain strength θ_1 of the interaction between the protein molecules, the ϕ_α line and the vertical line of constant ϕ_s cross; then, ϕ_s is equal to the lower limit of the metastable two-phase region in the liquid; that is, $\phi_s = \phi_\alpha$ for $\theta = \theta_1$. This situation corresponds to a triple point temperature T_t of 300 K. For larger values of θ ($\theta > \theta_1$), there is a stable liquid−liquid two-phase region and crystallization is only possible from a solution with a very large protein concentration. The value of θ_1 as a function of the solubility ϕ_s is given in Figure 4.

Remarks About Protein−Protein Interactions

The phase diagram of the protein−water system is determined by the parameters g_λ and g_c, which characterize the interactions between the protein molecules in the liquid and in the solid phase, respectively.

The value of g_c can be calculated from the experimentally determined solubility. It has been noted that for solutions that are suitable for the growth of protein crystals, the values of g_c show little variation; the average value being $-(58 \pm 3) \times 10^{-21}$ J (Table 2, and ref 4). The reason that g_c shows little variation is very simple. For the growth of crystals, a realistic value of the solubility is required. For values of g_c deviating considerably from the average value, the solubility is either much too large or much too small.

The parameter g_λ (or θ) for the protein−protein interaction in the solution is related to the second virial coefficient B by the relation (see *Appendix 1*):

$$B = (1/M\rho)[4 + (g_\lambda/kT\phi_c)] \qquad (11a)$$

or, using eq 6

$$B = (1/M\rho)[4 - 10.601(\Theta/T)] \qquad (11b)$$

TABLE 2: Molecular Weight *M* (in kD), Solubility ϕ_s (Volume Fraction Protein), and the Parameter g_c for the Interaction between Protein Molecules in the Crystalline Phase[a]

protein	*M* (kD)	*m*	ϕ_s	g_c (10^{-21} J)	ref
lysozyme	14	570	2.2×10^{-3}	−52	16
chymotrypsinogen	27	900	3.4×10^{-3}	−52	17, 18
thermolysine	34	1390	0.12×10^{-3}	−68	19, 20
concanavalin A	102	4160	0.74×10^{-3}	−65	21
concanavalin A	102	4160	34.8×10^{-3}	−49	21
pumpkin seed globulin	112	4580	9.6×10^{-3}	−55	22
STMV	1500	61 300	0.25×10^{-3}	−81	23, 24
porcine pancreatic α-amylase	54	2200	4.44×10^{-3}	−55	25
glucose isomerase	171	6990	0.074×10^{-3}	−77	26
horse serum albumin	67	3740	6.66×10^{-3}	−56	27
BPTI	6.5	266	5.18×10^{-3}	−46	28
procanavalin	42	1720	14.8×10^{-3}	−49	29
canavalin	21	860	14.8×10^{-3}	−46	29
collagenase of hypoderma lineatum	25	1020	0.44×10^{-3}	−62	30
ovalbumin	43	1760	7.4×10^{-3}	−52	31

[a] The solubility depends strongly on the composition of the solvent; details about the composition can be found in the references; the ratio of the molar volumes of protein and water is given by $m = M/18\rho$; for the protein density, a value of $\rho = 1.36$ g cm^{-3} is used.

In these equations, *M* is the molecular weight and ρ is the density of the protein molecule. The second virial coefficient, and hence also g_λ and θ, can be obtained from experimental data on light scattering or osmotic pressure.

The parameters g_λ and g_c are Gibbs free energies that have an enthalpy and an entropy contribution ($g = h - Ts$). For the interaction between protein molecules in an aqueous solution, or in a crystal that also contains a considerable amount of water, the entropy contribution is quite large.[16,32] For example, it was found from measurements of the temperature dependence of the solubility of lysozyme that (for a specific aqueous solvent) $h_c = -167 \times 10^{21}$ J and $g_c = -53 \times 10^{-21}$ J;[33] the large difference between g_c and h_c is the entropy contribution $Ts_c = -114 \times 10^{-21}$ J. The values of h_c and s_c depend strongly on the composition of the solvent.[34] The temperature dependence of g_c is determined by the entropy s_c: $(\partial g_c/\partial T) = -s_c$. Because the entropy s_c is large, the temperature dependence of g_c will be large. It is expected that also g_λ depends strongly on the temperature.

Equation 1 contains a simple, phenomenological expression for the interaction between the protein molecules; only contributions proportional to ϕ^2 are considered, and higher order terms are neglected. Many authors[13,35−42] have computed the phase diagram of colloid or protein solutions using numerical calculations based on a simple effective interaction potential between the particles. These computations are accurate in the sense that all higher-order terms are taken into account, but they are approximate because a simple dependence of the interaction potential $U(r)$ on the distance *r* between the particles is assumed. The numerical phase diagram calculations show that a metastable liquid−liquid two-phase region occurs for a short-range interaction potential. The critical volume fraction of ∼0.2, obtained from the numerical calculations, is larger than our value of 0.1305, and the curve of *T* versus ϕ_α, ϕ_β is broader than the curve in Figure 1.

The pair potential $U(r)$ is related to the parameter g_λ by the relation[43]

$$g_\lambda = (kT\phi_c/2\Omega)\int_{2a}^{\infty}[1 - \exp(-U(r)/kT)]\,dV \quad (12)$$

The molecules are assumed to be spherically symmetric, with a radius a, and a distance r between the centers of the two molecules. For eq 12, it is further assumed that the interaction potential $U(r)$ has a hard core, with $U(r) = \infty$ for $r < 2a$. From eq 12 it follows directly that g_λ depends strongly on the range of the interactions between the protein molecules; for a short-range interaction, g_λ is small. Because g_c depends only on the strength of the interaction and not on its range, one expects that for a short-range interaction, the ratio $f = g_\lambda/g_c$ is small. Thus, the conditions for a metastable liquid−liquid two-phase region from numerical calculations (short-range interaction) and from our considerations ($f < f^*$) are equivalent (see *Appendix 2*).

In our opinion, the interaction between protein molecules cannot be represented by an isotropic, temperature-independent pair potenial. In the first place, even the so-called globular proteins do not have an exactly spherical shape. Moreover, the interactions are very anisotropic because on the outer surface of the protein molecule there is a highly inhomogeneous distribution of polar and nonpolar groups, charges, folds, and crevices. As a result, the interaction between two molecules is a sensitive function of the orientation of the two molecules with respect to each other. In the crystal, only orientations occur for which the interactions correspond to a maximum bonding, which is not so in the solution; the value of g_λ is determined by an average over all orientations of the molecules. Finally, we mention that the large entropy contribution cannot be accounted for by a temperature−independent pair potential.

Discussion

Only very few reliable data are available on the liquid−liquid phase separation in protein−water systems. It has been reported [44,45] that aqueous solutions of lysozyme separate into two liquid phases below a critical temperature T_0 of ∼285 K and a critical volume fraction ϕ_0 of 0.17. The value of T_0 depends somewhat on pH and ionic strength of the solution. The protein γ_{11}-cristalline shows a two-phase region with a critical temperature $T_0 = 268$ K and a critical volume fraction $\phi_0 = 0.17$.[46] In these two cases, the two-phase region is metastable; that is, the phase diagram does not show a triple point.

It has been suggested that a mechanism for protein crystal growth with phase separation in the liquid phase as a first step is a suitable method to grow protein crystals. The conditions for this mechanism can now be stated explicitly. First, it is necessary that the strength of protein−protein interactions in the solution is adjusted such that it lies between the limits $\theta =$ 300 K and θ_1; the latter value corresponds to the case where $\phi_s = \phi_\alpha$. Furthermore, the concentration of the protein solution should be chosen between the limits ϕ_s and ϕ_α*.

We compare these conditions for crystal growth via liquid−liquid-phase separation with the so-called "crystallization slot" suggested by George et al.[47,48] These authors calculated from light scattering measurements the second virial coefficient B of protein solutions that are suitable for crystal growth. The results are given in Table 3. The authors observed that the values of B for these solutions fall in a rather narrow range from -0.8×10^{-4} to -8.4×10^{-4} mol cm^3 g^{-2}. Because B is directly related to θ, a certain range of B values corresponds to a range for θ. We calculated the values of θ and g_λ from the B values reported by George et al. (Table 3).

TABLE 3: Values of the Second Virial Coefficient B[a]

protein	M (kD)	B (10^{-4} mol cm^3 g^{-2})	θ (298 K)	g_λ (10^{-21} J)
BSA	65	−2.1	638	−47
canavalin	141	−0.8	547	−40
concanavalin A	102	−2.5	1095	−80
concanavalin A	102	−1.9	859	−63
α-chymotrypsin	25	−8.4	921	−67
α-lactalbumin	14	−7.3	506	−37
β-lactoglobulin A	36	−2.4	446	−33
β-lactoglobulin A	36	−6.2	972	−71
β-lactoglobulin B	36	−2.8	501	−37
β-lactoglobulin B	36	−6.2	972	−71
lysozyme	14	−2.8	264	−19
ovostatin	720	−7.1	19 800	−1448
ovalbumin	43	−6.1	1123	−82
pepsin	36	−7.8	1194	−87
pepsin	36	−2.8	501	−37
pepsin	36	−0.8	224	−16
ribonuclease A	14	−4.1	334	−24
STMV	1500	−1.8	10 508	−768
thaumatin	22	−3.0	367	−27

[a] Obtained from light scattering measurements, as reported by George et al.,[47,48] the molecular weight M in kD, and the parameters g_λ and θ for the interaction between the protein molecules in the solution, as calculated from B (assuming $\rho = 1.36$ g cm^{-3} and $\phi_c = 0.5$); the values of B depend on the composition of the solvent; for some of the proteins in the table, several values of B are reported, corresponding to different compositions of the solvent; details are found in the references given in ref 48.

In our considerations about the phase diagram, we found that nucleation at 300 K via liquid−liquid phase separation is only possible for θ values between $\theta = 300$ K and θ_1. The value of θ_1 depends on the solubility ϕ_s. Unfortunately, accurate data for the solubility is not available for the proteins in Table 3, so that the values of θ_1 cannot be calculated. However, for a realistic value of the solubility of 6×10^{-4} at 298 K, we find $\theta_1 = 600$ K (see Figure 4). The range from $\theta = 300$ K to $\theta =$ 600 K corresponds for a molecular weight of $M = 14\,000$ to a range of B values from -3.5×10^{-4} to -9.0×10^{-4} mol cm^3 g^{-2}, and for $M = 14\,0000$ to a range from -0.35×10^{-4} to -0.9×10^{-4} mol cm^3 g^{-2} (for $\rho = 1.36$ g cm^{-3}). This range of B values corresponds closely to the range of B values reported by George et al.[47,48] for solutions used for protein crystal growth. We conclude that the considerations of the previous sections provide a natural and nearly quantitative explanation of the observation by George et al. that the second virial coefficient B for solutions suitable for crystal growth fall in a narrow range, the so-called "crystallization slot". This agreement can be understood only if the prevailing mechanism for protein crystal growth is one with nucleation by liquid−liquid phase separation.

The values of θ (or g_λ) deduced from the light scattering experiments[47,48] for the large molecules ovostatin ($M = 720$ kD) and STMV ($M = 1500$ kD) deviate strongly from the results for the smaller molecules (see Table 3). From our considerations we would expect for crystallization via the mechanism of liquid−liquid phase separation θ values smaller than ∼1000 K and B values in the range from -0.1 to -0.2×10^{-4} mol cm^3 g^{-2} for ovostatin and from -0.03 to -0.1×10^{-4} mol cm^3 g^{-2} for STMV. The observed values are not within these ranges. This result could indicate that for these solutions of ovostatin and STMV, the crystallization process does not proceed via liquid−liquid phase separation as a first step, but rather via the classical mechanism with the direct formation of crystalline nuclei.

Generally, it is not possible to calculate from the value of B at 300 K the critical temperature T_0 because that B and θ depend

on T. For lysozyme, θ at 300 K is equal to 270 K, which happens to be close to 300 K. Because $\theta(T_0) = T_0$, the critical temperature for lysozyme must be close to 270 K. We can compare this with values of T_0 obtained from measurements of the temperature where the liquid−liquid phase separation begins or disappears (the cloud temperature). For lysozyme, a cloud temperature of $T_0 = 285 \pm 20$ K is reported, with a small variation as a function of pH and ionic strength of the solution.[44,45] This value of T_0 is in very good agreement with the value calculated with our equations from the experimental data for B.

Conclusion

The phase diagram of a protein−water system can be described in a satisfactory way with a simple model with only two parameters, which characterize the interaction between protein molecules in the liquid and in the solid phase. The calculated phase diagram reproduces features such as a metastable liquid−liquid two-phase region and a triple point.

The conditions for the growth of protein crystals were analyzed on the basis of the calculated phase diagram. A mechanism of crystal growth with the nucleation by a metastable liquid−liquid phase separation as a first step, that is, the formation of small droplets with a high protein concentration, is possible only in a fairly narrow range of values for the interaction between protein molecules in the solution. This range was found to correspond closely with the range of values for the second virial coefficient for solutions suitable for the growth of protein crystals. This correspondence is evidence that the prevailing mechanism for the growth of protein crystals is one with liquid−liquid phase separation as a first step (at least in most cases). The conditions for the protein−protein interaction strength and the required supersaturation for this mechanism can be stated explicitly.

An important aspect of protein crystallization is the transient nature of the process, and the associated long induction time before the crystallization sets in.[2,8,49−52] In the classical nucleation theory,[2,8,9] a stationary distribution of concentration fluctuations in the supersaturated solution is assumed, and this leads to a constant nucleation rate (i.e., a linear increase of the number of nuclei with time). However, in the experiment one starts with an unsaturated solution. By changing the conditions, the solution is brought to supersaturation. At the moment supersaturation is reached, the distribution of the fluctuations still corresponds to the distribution characteristic of the original solution. It takes a certain time to transform this distribution to the one corresponding to the supersaturated solution, and this is the induction time. Recent experiments[52] show that cycling through the liquid−liquid two-phase region strongly enhances the rate of formation of crystals, but is detrimental for obtaining large crystals of good quality. The induction time of the nucleation process is strongly reduced by the cycling procedure. In our opinion, a large number of small liquid nuclei are preformed during the cycling procedure, and this will as observed, strongly reduce the induction time. The fact that many small crystals are formed is a consequence of the large number of nuclei. Some experimental studies of the parameters, that influence the induction time, such as the degree of supersaturation and the composition of the solvent, have been reported,[23,53−55] but a quantitative understanding of the induction time in protein crystallization is lacking.

We realize that the results of our study do not give the protein crystallographer a simple recipe for choosing successful crystallization conditions. For this recipe, further theoretical and experimental studies of the kinetics of the phase separation processes, and in particular of the induction time, are necessary.

Appendix 1

The osmotic pressure is given by the expression

$$\Pi = \frac{1}{\omega}[\mu_1(\phi = 0) - \mu_1(\phi)] \tag{13}$$

where μ_1 is the thermodynamic potential of the solvent. The second virial coefficient B is related to the osmotic pressure by

$$\Pi = RTc\left[\frac{1}{M} + Bc\right] \tag{14}$$

where c is the concentration ($c = \phi\rho$).

The thermodynamic potential can be calculated from the Gibbs free energy. Consider a volume V containing N_1 molecules of the solvent and N_2 protein molecules. The total Gibbs free energy is $G_{tot}(\phi) = VG_\lambda(\phi)$, where G_λ is the Gibbs free energy per unit volume, $V = N_1\omega + N_2\Omega$, and $\phi = N_2\Omega/(N_1\omega + N_2\Omega)$. The thermodynamic potential of the solvent is defined as

$$\mu_1 = (\partial G_{tot}/\partial N_1)_{N_2 = \text{constant}} \tag{15}$$

Using the expressions for G_{tot}, V, and ϕ, we find

$$\mu_1 = \omega\left(G_\lambda - \phi\frac{\partial G_\lambda}{\partial \phi}\right) \tag{16}$$

Because $G_\lambda(\phi = 0) = 0$, we find that the osmotic pressure is given by

$$\Pi = -G_\lambda + \phi\frac{\partial G_\lambda}{\partial \phi} \tag{17}$$

From eqs 1 and 14, we calculate for the second virial coefficient

$$B = (1/M\rho)[4 + g_\lambda/kT\phi_c] \tag{18}$$

Appendix 2

We will show that for a short-range interaction between the protein molecules, the ratio $f = g_\lambda/g_c$ is small.

Consider as an example a simple square well potential,[35] with $U(r) = \infty$ for $r < 2a$, $U(r) = -\epsilon$ for $2a < r < 2a\lambda$, and $U(r) = 0$ for $r > 2a\lambda$. Substitution in eq 12 gives

$$g_\lambda = -4kT\phi_c[\exp(\epsilon/kT) - 1](\lambda^3 - 1) \tag{19}$$

For a short-range interaction $\lambda \rightarrow 1$ and $g_\lambda \rightarrow 0$.

In the crystalline phase, the distance between the molecules will correspond to the maximum of the attractive interaction. Thus, the interaction between a pair of neighboring molecules in the crystal will be $-\epsilon$. If each protein molecule in the crystal has z neighbors, then the free energy of the crystal will be

$$G_c = (\phi_c/\Omega)g_c = -\frac{1}{2}z(\phi_c/\Omega)\epsilon \tag{20}$$

so that

$$g_c = -\frac{1}{2}z\epsilon \tag{21}$$

For the ratio $f = g_\lambda/g_c$, we find

$$f = (8\phi_c/z)[\exp(\epsilon/kT) - 1](\lambda^3 - 1) \qquad (22)$$

This proves that f is small for a short-range interaction ($\lambda \rightarrow 1$).

References and Notes

(1) McPherson, A. *Preparation and Analysis of Protein Crystals*; John Wiley & Sons: New York, 1982.

(2) Chernov, A. A.; Komatsu, H. In *Science and Technology of Crystal Growth*; van der Eerde, J. P., Bruinsma, O. L. S., Eds.; Kluwer Academic: Norwell, MA, 1995; p 329.

(3) Ducruix, A.; Giegé, R. *Crystallization of Nucleic Acids and Proteins; A Practical Approach*; IRL: Oxford, 1992.

(4) Haas, C.; Drenth, J. *J. Cryst. Growth* **1994**, *154*, 126.

(5) Carnahan, N. F.; Starling, K. F. *J. Chem. Phys.* **1969**, *51*, 635.

(6) Jansen, J. W.; de Kruif, C. G.; Vrij, A. *Chem. Phys. Lett.* **1984**, *107*, 456.

(7) Taratuta, V. G.; Holockbach, A.; Thurstein, G. M.; Blankschtein, D. B.; Benedek, G. B. *J. Phys. Chem.* **1990**, *94*, 2140.

(8) Chernov, A. A. *Modern Crystallography*, vol. III; Springer Series in Solid State Sciences, Berlin, 1984, *p 36*.

(9) Kelton, K. F. *Solid State Physics* **1991**, *45*, 75.

(10) Elgersma, A. V.; Ataka, M.; Katsura, T. *J. Cryst. Growth* **1992**, *122*, 31.

(11) Kam, Z.; Shore, H. B.; Feher, G. *J. Mol. Biol.* **1978**, *123*, 539.

(12) Giegé, B.; Drenth, J.; Ducruix, A.; McPherson, A.; Saenger, W. *Prog. Cryst. Growth Characterization* **1995**, *30*, 251.

(13) tenWolde, P. R.; Frenkel, D. *Science* **1997**, *277*, 1975.

(14) Cahn, J. W.; Hilliard; J. E. *J. Chem. Phys.* **1959**, *31*, 688.

(15) Binder, K.; Billotto, C.; Mirold, P. *Z. Physik* **1978**, *B30*, 183.

(16) Arakawa, T.; Timasheff, S. N. *Methods Enzymol.* **1985**, *114*, 49.

(17) Ries-Kautt, M. M.; Ducruix, A. F. *J. Biol. Chem.* **1985**, *264*, 745.

(18) Hamilton, J. A.; Koutsky, J. A.; Walton, A. G. *Nature* **1964**, *204*, 1085.

(19) Sasaki, T. G.; Ooshima, H.; Kato, J.; Harano, Y.; Hirikawa, N. *J. Cryst. Growth* **1993**, *130*, 357.

(20) Sasaki, G.; Ooshima, H.; Kato, J. *J. Cryst. Growth* **1994**, *135*, 199.

(21) Mikol, V.; Giegé, R. *J. Cryst. Growth* **1989**, *97*, 324.

(22) Malkin, A. J.; McPherson, A. *J. Cryst. Growth* **1993**, *133*, 29.

(23) Malkin, A. J.; Chueng, J.; McPherson, A. *J. Cryst. Growth* **1993**, *126*, 544.

(24) Larson, S. B.; Koszelak, S.; Day, J.; Greenwood, A.; Dodds, J. A.; McPherson, A. *J. Mol. Biol.* **1993**, *231*, 375.

(25) Boistelle, R.; Astier, J. P.; Marchis-Mourin, G.; Desseaux, V.; Haser, R. *J. Cryst. Growth* **1992**, *123*, 109.

(26) Chayen, N.; Atkins, J.; Campbell-Smith, S.; Blow, D. M. *J. Cryst. Growth* **1992**, *90*, 112.

(27) Rosenberger, F.; Howard, S. B.; Sowers, J. W.; Nyce, T. A. *J. Cryst. Growth* **1993**, *129*, 1.

(28) Lafont, S.; Veesler, S.; Astier, J. P.; Boistelle, R. *J. Cryst. Growth* **1994**, *143*, 249.

(29) DeMattei, R. C.; Feigelson, R. S. *J. Cryst. Growth* **1991**, *110*, 34.

(30) Carbonnaux, C.; Ries-Kraut, M.; Ducruix, A. *Protein Science* **1995**, *4*, 2123.

(31) Jugde, R. A.; Johns, M. R.; White, E. T. *J. Chem. Eng. Data* **1996**, *41*, 422.

(32) Blokzijl, W.; Engberts, J. B. F. N. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1545.

(33) Gripon, C.; Legrand, L.; Rosenman, T.; Vidal, O.; Robert, M. C.; Bouë, F. *J. Cryst. Growth* **1997**, *177*, 238.

(34) Cacioppo, E.; Pusey, M. L. *J. Cryst. Growth* **1991**, *114*, 286.

(35) Asheric, N.; Lomakin, A.; Benedek, G. B. *Phys. Rev. Lett.* **1996**, *76*, 4832.

(36) Rosenbaum, D.; Zamora, P. C.; Zukoski, C. F. *Phys. Rev. Lett.* **1996**, *76*, 150.

(37) Ilet, S. M.; Orrock, A.; Poon, W. C. K.; Pusley, P. N. *Phys. Rev.* **1995**, *E51*, 1344.

(38) Tejero, C. F.; Daanoun, A.; Lekkerkerker, H. N. W.; Baus, M. *Phys. Rev. Lett.* **1994**, *73*, 752.

(39) Tejero, C. F.; Daanoun, A.; Lekkerkerker, H. N. W.; Baus, M. Phys. Rev. **1995**, *E51*, 558.

(40) Hagen, M. H. J.; Frenkel, D. *J. Chem. Phys.* **1994**, *101*, 4093.

(41) Hagen, M. H. J.; Meijer, E. J.; Mooij, C. A. M.; Frenkel, D.; Lekkerkerker, H. N. W. *Nature* **1993**, *365*, 425.

(42) Poon, W. C. K. *Phys. Rev.* **1997**, *E55*, 3762.

(43) Landau, L. D.; Lifshitz, E. M. *Statistical Physics*, vol. 5, part I; Pergamon: Oxford, 1980.

(44) Broide, M.; Tominc, T. M.; Saxowsky, M. D. *Phys. Rev.* **1996**, *E56*, 6325.

(45) Tanaka, S.; Yamamoto, M.; Ito, K.; Hayakawa, R.; Ataka, M. *Phys. Rev.* **1997**, *E56*, R67.

(46) Broide, M. L.; Berland, C. R.; Pande, J.; Ogun Olutayo, O.; Benedek, G. B. *Proc. Acad. Sci. U.S.A.* **1991**, *88*, 5660.

(47) George, A.; Wilson, W. W. *Acta Crystallogr.* **1994**, *D50*, 361.

(48) George, A.; Chiang, Y.; Guo, B.; Arabshahi, A.; Cai, Z.; Wilson, W. W. *Methods Enzymol.* **1997**, *276*, 100.

(49) Binder, K.; Stauffer, D. *Adv. Phys.* **1976**, *25*, 343.

(50) Kashchiev, D. *Surf. Sci.* **1969**, *18*, 389.

(51) Feher, G.; Kam, Z. *Methods Enzymol.* **1985**,*114*, 77.

(52) Muschol, M.; Rosenberger, F. *J. Chem. Phys.* **1997**, *107*, 1953.

(53) Ataka, M.; Tanaka, S. *Biopolymers* **1986**, *25*, 337.

(54) Malkin, A.; McPherson, A. *J. Cryst. Growth* **1993**, *128*, 1232.

(55) Drenth, J.; Haas, C. *Acta Crystallogr.* **1998**, in press.