

Modeling Opposite Effects of an Additive on Liquid-Liquid Phase Separation and Crystal Solubility of Proteins

Onofrio Annunziata, Shamberia Thomas

Department of Chemistry and Biochemistry, Texas Christian University, Fort Worth, TX 76109, USA

S1. Wertheim Perturbation Theory

To describe LLPS, we also examined Wertheim perturbation theory of associating spheres (WPT)

The residual Helmholtz free energy for this model is¹⁻³

$$a_R(\beta, \phi) = n_b \phi \left(\ln X + \frac{1-X}{2} \right) \quad (\text{S1})$$

where n_b is the number of binding sites, which is assumed identical and independent. These sites are randomly distributed on particle surface. Binding between two particles is modeled using a square-well potential with attraction energy, ε_b . Binding occurs only when the distance between the surface sites of two different particles is less than the interaction range, δ (as a multiple of particle diameter, σ). To ensure that binding does not involve more than two proteins, we must have: $\delta < 1 - \sqrt{3}/2 \cong 0.134$. In Eq. S1, X is the fraction of free sites, with $a_R = 0$ when $X = 1$. It is obtained from the mass-action equation:^{2,4}

$$X = \frac{1}{1 + \omega X} \quad (\text{S2})$$

or

$$X = \frac{2}{1 + \sqrt{1 + 4\omega}} \quad (\text{S3})$$

where

$$\omega \equiv \frac{n_b \phi}{V_p} \int_{\sigma}^{(1+\delta)\sigma} (4\pi r^2) \cdot f \cdot g_{\text{HS}} dr = 24n_b \phi \int_1^{1+\delta} x^2 \cdot f \cdot g_{\text{HS}} dx \quad (\text{S4})$$

with $x \equiv r/\sigma$, $V_p = (\pi/6)\sigma^3$ and

$$f = (e^{\beta\varepsilon_b} - 1) \frac{1}{6\sigma^2 r} (\sigma\delta + \sigma - r)^2 (2\sigma\delta - \sigma + r) = (e^{\beta\varepsilon_b} - 1) \frac{(1 + \delta - x)^2}{6} \left(1 - \frac{1 - 2\delta}{x} \right) \quad (\text{S5})$$

The function, $f(x)$, is an angle average of the site-site Mayer function.³⁻⁶ In Eq. S4, $g_{\text{HS}}(x, \phi)$ is the pair distribution function of the hard-sphere reference fluid. If we insert Eq. S5 into Eq. S4, we

obtain

$$\omega = 4n_b\phi(e^{\beta\varepsilon_b} - 1)\int_1^{1+\delta}(1+\delta-x)^2\left(1+\frac{2\delta-1}{x}\right)g_{\text{HS}}(x,\phi)x^2dx \quad (\text{S6})$$

We can formally rewrite Eq. S6 in the following way:

$$\omega = n_b\phi(e^{\beta\varepsilon_b} - 1)g_{\text{HS}}(1,\phi)\Omega_b \quad (\text{S7})$$

where

$$g_{\text{HS}}(1,\phi) = \frac{1-\phi/2}{(1-\phi)^3} \quad (\text{S8})$$

is the Carnahan-Starling contact value of the pair distribution function and Ω_b is a phase volume (as a multiple of V_p) over which a bond can form. After introducing, the difference, $\Delta g_{\text{HS}} \equiv g_{\text{HS}}(x,\phi) - g_{\text{HS}}(1,\phi)$, it follows from Eq. S6 that the expression of Ω_b is

$$\Omega_b = 4\int_1^{1+\delta}(1+\delta-x)^2\left(1+\frac{2\delta-1}{x}\right)x^2dx + 4\int_1^{1+\delta}\frac{\Delta g_{\text{HS}}(x,\phi)}{g_{\text{HS}}(1,\phi)}(1+\delta-x)^2\left(1+\frac{2\delta-1}{x}\right)x^2dx \quad (\text{S9})$$

In general, we need an expression for $g_{\text{HS}}(x,\phi)$. For sufficiently short distances, we can write:⁷

$$g_{\text{HS}}(x,\phi) = g_{\text{HS}}(1,\phi) + g'_{\text{HS}}(\phi) \cdot (x-1) \quad (\text{S10})$$

where

$$g'_{\text{HS}} \equiv \left(\frac{\partial g_{\text{HS}}}{\partial x}\right)\Bigg|_{\phi, x=1} = -\frac{9}{2}\frac{1+\phi}{(1-\phi)^3}\phi \quad (\text{S11})$$

Thus, the expression of Ω_b becomes

$$\begin{aligned} \Omega_b &= 4\int_1^{1+\delta}(1+\delta-x)^2\left(1+\frac{2\delta-1}{x}\right)x^2dx + \\ &-18\frac{1+\phi}{1-\phi/2}\phi\int_1^{1+\delta}(1+\delta-x)^2\left(1+\frac{2\delta-1}{x}\right)(x-1)x^2dx \end{aligned} \quad (\text{S12})$$

After solving the two integrals in Eq. S12, we obtain:

$$\Omega_b = \left(3 + \frac{4}{5}\delta\right)\delta^4 - \frac{9}{2} \frac{1+\phi}{1-\phi/2} \phi \left(\frac{4}{5} + \frac{1}{3}\delta\right)\delta^5 \quad (\text{S13})$$

The second term in Eq. S13 is may be neglected under the assumption that δ is sufficiently short that the Sticky-limit approximation is good.^{2, 3} Hence, we can introduce the sticky-limit approximation of Ω_b :

$$\Omega_b = \left(3 + \frac{4}{5}\delta\right)\delta^4 \quad (\text{S14})$$

To assess accuracy of WPT model, we consider preciously reported Monte Carlo simulation data on critical coordinates, available for $n_b = 5$ and $\delta = 0.119$.⁷ We specifically calculated $a_R(\beta, \phi)$ using Eq. S1 with Ω_b given by Eq. S14 (sticky limit) and Eq. S13 (general case), and apply Eq. 2a,b in the manuscript to determine $\phi^{(l)}(\beta)$, $\phi^{(m)}(\beta)$ and (ϕ_c, β_c^{-1}) . For completeness, we also calculate a_R using the Percus-Yevick expression of the contact value,⁸ $g_{\text{HS}}(1, \phi) = (1 + \phi/2)/(1 - \phi)^2$, in Eq. S7, which was previously^{3, 6} employed, with Ω_b given by Eq. S14. The comparison between the three calculated phase boundaries is shown in Fig. S1A. Here, we can see that the three different choices of g_{HS} produce curves that are fairly close to each other (deviations of $\approx 1\%$ in the values of β_c^{-1}). We can also appreciate that the Sticky-limit approximation is sufficiently accurate for $\delta \approx 0.1$ (and less). For convenience, we will use the Carnahan-Starling expression of $g_{\text{HS}}(1, \phi)$ as it is consistent with the chosen expression of a_{HS} . It is also convenient to introduce the phase volume, Ω_b , as a fixed parameter independent of ϕ . Thus, the sticky-limit approximation for Ω_b given by Eq. S14 will be used.

As we can also see in Fig. S1A, the value of β_c^{-1} from simulations is just $\approx 2\%$ higher than that calculated using Eq. 8, showing that WPT accurately describes LLPS temperature. However, the critical volume fraction from simulations, $\phi_c = 0.185$,⁷ is significantly larger than the WPT value of $\phi_c = 0.13$. Since the value of ϕ_c increases with number of sites, we examined the effect of n_b on ϕ_c using the WPT model. As shown in Fig. S1B, we can achieve $\phi_c \approx 0.18$ only if we increase n_b from 5 to 7. Thus, our analysis indicates that the value of n_b from WPT model is about two units larger than the actual number of sites. Since critical volume fractions of protein solutions are

typically about $\phi_c \approx 0.2$,⁹ it follows from Fig. 4B that the best WPT value to describe these mixtures is $n_b \approx 8$. It follows from simulation results that this WPT value of n_b is about two units larger than the actual number of sites, $n_b \approx 6$.

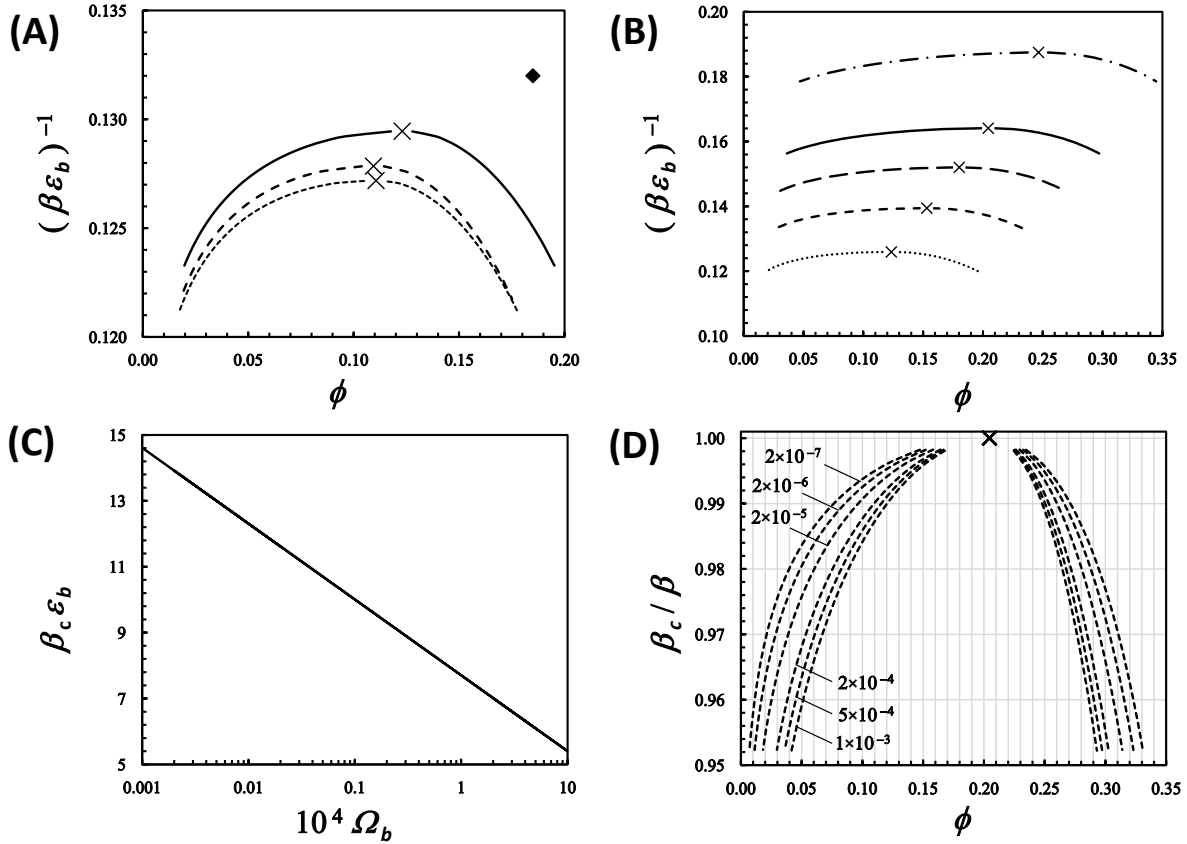


Figure S1. (A) Phase diagram showing normalized temperature, $(\beta\epsilon_b)^{-1}$, as a function of protein volume fraction, ϕ . LLPS boundaries extracted using $n_b = 5$, $\Omega_b = 6.3 \times 10^4$ ($\delta = 0.119$) and Carnahan-Starling (solid curve) and Percus-Yevick (dashed curve) values of $g_{\text{HS}}^{\text{con}}$, and integration of $g_{\text{HS}}(r)$ (dotted curve). Crosses indicate the corresponding critical points. Diamond indicates critical point extracted from Monte Carlo simulations. (B) Phase diagram showing LLPS boundaries extracted using $n_b = 5$ (dotted curve), 6 (dashed curve), 7 (long dashed curve), 8 (solid curve) and 10 (dash-dotted curve), and $\Omega_b = 5.0 \times 10^4$. (C) Plot describing dependence of $\beta_c \epsilon_b$ on Ω_b . (D) Phase diagram showing how the width of LLPS boundaries depend on Ω_b at $n_b = 8$. Numbers associated with curves are the corresponding values of Ω_b .

In our use of WPT, we should consider n_b as a fitting parameter that is just needed to reproduce experimental values of ϕ_c . It plays the same role as the range of interactions, λ , in the BHPT model. In contrast, the phase volume, Ω_b , does not affect ϕ_c . However, it does have an impact on both critical temperature, β_c^{-1} , and width of LLPS boundary. This is shown in Figs. S1C,D. It follows from Fig. S1C that decreasing Ω_b lowers LLPS temperature (e.g., β_c^{-1}) at constant attraction energy, ε_b . In Fig. S1D, we can appreciate that boundary width increases as Ω_b decreases. These are the characteristic trends associated with an increase in the anisotropic character of protein-protein interactions.^{2, 3, 9, 10} Indeed, the orientational degrees of freedom associated with protein-protein binding decrease as the phase volume of a site is reduced. The quantity, Ω_b in the WPT model plays the same role as α in the revised BHPT model.

S2. Effect of HEPES on lysozyme LLPS boundary

We shall now employ the WTP model to describe the experimental LLPS boundaries for REF and HEPES systems. These are shown in the phase diagrams of Fig. S2 together with the corresponding crystal solubility curves. We shall set $n_b=8$ to accurately reproduce the critical volume fraction of $\phi_c \approx 0.20$. These parameters are assumed to be constant, independent of solvent composition. To make crystal cell model consistent with WPT model, we shall also assume that the attraction energy parameters, ε_b , for the fluid phase are the same as ε_s extracted from solubility data. We are then left with identifying the values of Ω_b that best fit experimental LLPS data. We determine that $\Omega_b = 5.2 \times 10^{-4}$ describes LLPS of the REF system. This value must be decreased to $\Omega_b = 3.2 \times 10^{-4}$ in order to accurately represent the lower LLPS temperatures of the HEPES system. Noteworthy, these values of Ω_b also describe the width of the LLPS boundaries fairly well. Finally, theoretical solubility curves are now obtained by calculating the protein chemical potential using WPT and BHPT models. These theoretical curves, which are also shown in Fig. S2, accurately describe solubility data justifying the initial use of $\mu \approx \mu_0 + \ln \phi$.

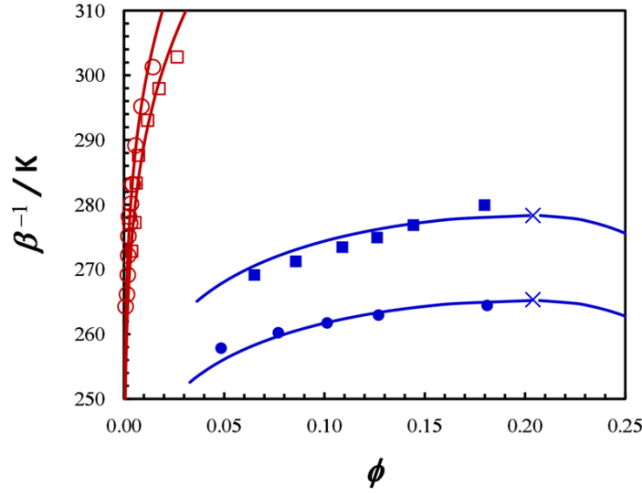


Figure S2. Phase diagram showing experimental LLPS data for the HEPES (solid circles) and REF (solid squares) systems. Theoretical LLPS boundaries are calculated from WPT model with $n_b=8$ in both cases. The values of $\varepsilon_b=1732$ K and $\Omega_b=3.23\times 10^{-4}$, and $\varepsilon_b=1687$ K and $\Omega_b=5.17\times 10^{-4}$ are used for the HEPES and REF systems, respectively. Solubility data for the HEPES (open circles) and REF (open squares) systems are also included. Theoretical solubility curves are obtained using $\Omega_s=2.6\times 10^{-6}$, and $n_s\varepsilon_s=10390$ K and 10120 K for the HEPES and REF systems, respectively.

S3. Second Virial Coefficient from WPT theory

The second-virial coefficient may be introduced as

$$B = \lim_{\phi \rightarrow 0} \frac{\hat{\pi} / \phi - 1}{\phi} = 4 + \lim_{\phi \rightarrow 0} \left(\frac{\partial a_R / \phi}{\partial \phi} \right)_{\beta} \quad (\text{S15})$$

where we have used $\hat{\pi} = \phi [\partial(a / \phi) / \partial \phi]_{\beta}$ from $\Pi = -(\partial A / \partial V)_{\beta}$ and “4” is the second-virial coefficient associated with hard spheres. It follows from Eqs. S1,S2 that:

$$\frac{a_R}{\phi} = n_b \left(\ln \frac{2}{1 + \sqrt{1 + 4\omega}} + \frac{1}{2} \frac{\sqrt{1 + 4\omega} - 1}{1 + \sqrt{1 + 4\omega}} \right) \quad (\text{S16})$$

where ω is given by Eq. S7 and Ω_b is a constant as described by Eq. S14. In the limit of $\phi \rightarrow 0$, we can rewrite Eq. S16 to first order with respect to ω :

$$\frac{a_R}{\phi} = n_b \left(\ln \frac{1}{1+\omega+\dots} + \frac{1}{2} \frac{\omega+\dots}{1+\omega+\dots} \right) = -\frac{n_b}{2} \omega + \dots \quad (\text{S17})$$

It then follows from Eq. S7 that

$$\lim_{\phi \rightarrow 0} \left(\frac{\partial a_R / \phi}{\partial \phi} \right)_{\beta} = \lim_{\phi \rightarrow 0} \frac{1}{\phi} \left(\frac{\partial a_R}{\partial \phi} \right)_{\beta} = -\frac{n_b^2}{2} \cdot \Omega_b \cdot (e^{\beta \varepsilon_b} - 1) \quad (\text{S18})$$

where we have also used $g_{\text{HS}}(1,0) = 1$. Thus, Eq. S15 becomes:

$$B = 4 - 2n_b^2 \cdot \Omega_b \cdot (e^{\beta \varepsilon_b} - 1) \quad (\text{S19})$$

Finally, we can make predictions on the values of second virial coefficient at 298 K using $n_b = 8$ in Eq. S19. Specifically, we use $\varepsilon_b = 1687$ K and $\Omega_b = 5.17 \times 10^{-4}$ to determine that $B = -15.0$ for the REF system, and $\varepsilon_b = 1732$ K and $\Omega_b = 3.23 \times 10^{-4}$ to determine that $B = -9.8$ for the HEPES system at 298 K. These values of B are significantly more negative compared to those obtained from BHPT model. Moreover, their difference at 298 K is +5.2. This is significantly larger than that evaluated from DLS data ($+1.2 \pm 0.4$) and the revised BHPT model ($+1.3$).

In conclusion, the WPT model can be successfully used to describe the LLPS boundaries using the same protein-protein attraction energy parameter extracted from solubility data ($\varepsilon_b = \varepsilon_s$). Only the anisotropy parameter (Ω_b) is decreased in order to successfully explain the effect of HEPES on LLPS temperature. However, the magnitude of the calculated second virial coefficients is relatively large.

References

- (1) Wertheim, M. Fluids with highly directional attractive forces .2. Thermodynamic perturbation-theory and integral-equations. *Journal of Statistical Physics* **1984**, *35* (1-2), 35-47. DOI: 10.1007/BF01017363.
- (2) Sear, R. Phase behavior of a simple model of globular proteins. *Journal of Chemical Physics* **1999**, *111* (10), 4800-4806. DOI: 10.1063/1.479243.
- (3) Kastelic, M.; Kalyuzhnyi, Y. V.; Hribar-Lee, B.; Dill, K. A.; Vlady, V. Protein aggregation in salt solutions. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112* (21), 6766-6770. DOI: 10.1073/pnas.1507303112.
- (4) Chapman, W.; Jackson, G.; Gubbins, K. Phase-equilibria of associating fluids chain molecules with multiple bonding sites. *Molecular Physics* **1988**, *65* (5), 1057-1079. DOI: 10.1080/00268978800101601.

- (5) Wertheim, M. Fluids of dimerizing hard-spheres, and fluid mixtures of hard-spheres and dispheres. *Journal of Chemical Physics* **1986**, *85* (5), 2929-2936. DOI: 10.1063/1.451002.
- (6) Brudar, S.; Hribar-Lee, B. Effect of Buffer on Protein Stability in Aqueous Solutions: A Simple Protein Aggregation Model. *Journal of Physical Chemistry B* **2021**, *125* (10), 2504-2512. DOI: 10.1021/acs.jpcc.0c10339.
- (7) Bianchi, E.; Largo, J.; Tartaglia, P.; Zaccarelli, E.; Sciortino, F. Phase diagram of patchy colloids: Towards empty liquids. *Physical Review Letters* **2006**, *97* (16). DOI: 10.1103/PhysRevLett.97.168301.
- (8) Barker, J.; Henderson, D. What is liquid - Understanding states of matter. *Reviews of Modern Physics* **1976**, *48* (4), 587-671. DOI: 10.1103/RevModPhys.48.587.
- (9) Lomakin, A.; Asherie, N.; Benedek, G. B. Monte Carlo study of phase separation in aqueous protein solutions. *Journal of Chemical Physics* **1996**, *104* (4), 1646-1656. DOI: 10.1063/1.470751.
- (10) Lomakin, A.; Asherie, N.; Benedek, G. Aeolotopic interactions of globular proteins. *Proceedings of The National Academy of Sciences of the United States of America* **1999**, *96* (17), 9465-9468. DOI: 10.1073/pnas.96.17.9465.