






RESEARCH ARTICLE | APRIL 24 2023

## An alternative approach to the osmotic second virial coefficient of protein solutions and its application to liquid–liquid phase separation ✓

Furio Surfarò ; Ralph Maier ; Kai-Florian Pastryk ; Fajun Zhang ; Frank Schreiber ; Roland Roth 

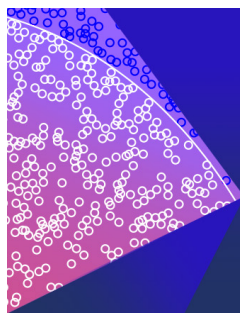


*J. Chem. Phys.* 158, 164902 (2023)

<https://doi.org/10.1063/5.0143696>



CrossMark



## The Journal of Chemical Physics

Special Topic: Monte Carlo methods,  
70 years after Metropolis *et al.* (1953)

**Submit Today**

# An alternative approach to the osmotic second virial coefficient of protein solutions and its application to liquid–liquid phase separation

Cite as: J. Chem. Phys. 158, 164902 (2023); doi: 10.1063/5.0143696

Submitted: 25 January 2023 • Accepted: 5 April 2023 •

Published Online: 24 April 2023



Furio Surfaro,<sup>1,a)</sup> Ralph Maier,<sup>1</sup> Kai-Florian Pastryk,<sup>1</sup> Fajun Zhang,<sup>1</sup> Frank Schreiber,<sup>1</sup> and Roland Roth<sup>2,b)</sup>

## AFFILIATIONS

<sup>1</sup>Institute of Applied Physics, University of Tübingen, 72076 Tübingen, Germany

<sup>2</sup>Institute of Theoretical Physics, University of Tübingen, 72076 Tübingen, Germany

<sup>a)</sup>Author to whom correspondence should be addressed: [furio.surfaro@uni-tuebingen.de](mailto:furio.surfaro@uni-tuebingen.de)

<sup>b)</sup>Electronic mail: [roland.roth@uni-tuebingen.de](mailto:roland.roth@uni-tuebingen.de)

## ABSTRACT

The osmotic second virial coefficient  $B_2$  is an important parameter to describe the interactions and phase behavior of protein solutions, including colloidal systems and macromolecular solutions. Another key parameter to describe the driving force of the nucleation of a new phase is the supersaturation, which is used in the classical nucleation theory framework and is connected with the favorable contribution in the Gibbs free energy in the bulk solution. In this article, we establish a connection between  $B_2$  calculated from small angle x-ray scattering (SAXS) data and the values of  $B_2$  obtained from supersaturation measurements using thermodynamics considerations. The values of the second virial coefficient calculated employing this method agree with those determined via SAXS in the region near the liquid–liquid phase separation border for human serum albumin and bovine serum albumin. The general relations adopted are shown to be useful for the estimation of the second virial coefficient  $B_2$  for globular proteins, in the proximity of the binodal biphasic coexistent region.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0143696>

## I. INTRODUCTION

Liquid–liquid phase separation (LLPS) in protein solutions and protein crystallization, as well as in different macromolecular and colloidal systems, are interesting and important phenomena that have far-reaching consequences in physics, chemistry, biology, and medicine.<sup>1–6</sup> LLPS may serve as a mechanism for cellular organization<sup>5</sup> or as a precursor on the pathway toward crystallization.<sup>7,8</sup> Despite the relatively simple observation of LLPS and protein crystals on the macroscopic scale, the rich phenomenology of protein solutions and their microscopic behavior are not fully understood yet. Given the complex nature of protein molecules and interactions on the microscopic scale, usually the key parameters used to describe protein solutions and their behavior are calculated with respect to macroscopic properties that are more easily accessible. Thus, thermodynamic relations are applied to provide

insights about the driving force of the nucleation process of a new phase.

An important parameter, often used to describe the tendency of a protein solution to phase separate, is the second virial coefficient  $B_2$  that measures the overall strength of the effective protein–protein interaction  $U(\mathbf{r}, \Omega)$ , which depends on the relative position  $\mathbf{r}$  and orientation  $\Omega$  between the proteins, averaged over all possible distances and orientations via

$$B_2(T) = -\frac{1}{2} \int [e^{-U(\mathbf{r}, \Omega)/kT} - 1] d\Omega d^3r. \quad (1)$$

Since the effective interaction potential between proteins is typically not known, a direct calculation of the second virial coefficient with Eq. (1) is impossible. Still,  $B_2$  is accessible either by thermodynamic considerations or experimentally. If  $B_2$  is positive, the effective inter-

action is overall repulsive, while it is attractive for negative values of  $B_2$ . For LLPS, a sufficiently strong attraction is required, and it has been observed for several colloidal systems  $B_2/B_{HS} \approx -1.5^9$  close to a critical point and more negative in the LLPS region. For protein solutions that display LLPS, even though the effective interaction is much more complicated than in the colloidal system, similar observations have been made close to the LLPS region.<sup>10,11</sup> Furthermore, it seems that frequently  $B_2$  falls into a narrow range when crystallization occurs, called crystallization slot.<sup>12</sup>

In order to obtain information about  $B_2$ , different experimental techniques are used, such as equilibrium sedimentation, dynamic light scattering (DLS), and small angle x-ray scattering (SAXS). However, these different approaches do not always lead to the same values, and the estimated results cannot easily be compared directly.<sup>13,14</sup> Additionally, the estimation of the  $B_2$  value requires different approximations due to the complex nature of the interactions.

Equation (1) is applied for the calculation of  $B_2$ , assuming different forms of simplified interaction potentials  $U(r)$ , that typically only depend on the relative distance  $r$  in order to improve the quality of the fit, i.e., assuming isotropy.<sup>15–18</sup>

In this work, we establish a thermodynamic relation in order to connect the osmotic second virial coefficient  $B_2$  with the supersaturation. Our method, explained in Sec. II, can be applied if the system, such as a protein solution, shows a meta-stable LLPS, and relies solely on easily accessible macroscopic observables. We explain different experimental procedures that have successfully been used for the determination of  $B_2$  so far in Sec. III. In Sec. IV, we compare results from our approach to those from SAXS fitting for human serum albumin (HSA) and bovine serum albumin (BSA) in solution with trivalent salt ions, which were shown to be a versatile method to manipulate interactions,<sup>19,20</sup> before we conclude in Sec. V.

## II. THEORY

We start by considering a protein solution with concentration  $c$ . Its chemical potential  $\mu$ , relative to a dilute reference system with concentration  $c_0$  and chemical potential  $\mu_0$ , is given by

$$\mu = \mu_0 + RT \ln \left( \frac{\gamma c}{c_0} \right), \quad (2)$$

where  $\gamma$  is the activity coefficient that takes the deviations of the solution from ideal behavior into account. Here, we assume that all effects of the solvent, ions, and crowding are included by  $\gamma$ . By definition,  $\gamma$  is equal to 1 for ideal solutions, i.e., at low protein concentrations. If  $\gamma$  deviates from unity, it indicates a concentration that is sufficiently high so that particles interact. A value of  $\gamma$  smaller than 1 corresponds to overall interactions between particles that are attractive, while a value larger than 1 hints toward overall interactions that are repulsive. We wish to describe an experimental situation, where a protein solution is prepared at an initial concentration  $c_i$ , which lies within the LLPS region, so that the solution phase separates into a low density phase with concentration  $c_l$  and a high density phase with concentration  $c_h$ . For protein solutions, the LLPS is usually meta-stable, which makes the determination of  $c_l$  and  $c_h$  far from trivial. However, we know that

along the tie line of the phase separation, one can observe chemical and mechanical equilibrium, i.e., constant chemical potential and constant osmotic pressure. In the following, we will make use of the chemical equilibrium:  $\mu(c_i) = \mu(c_l) = \mu(c_h)$ . A schematic phase diagram describing this process can be found in Ref. 21. After the LLPS occurs, one can observe that the concentration of the low density phase reduces to its equilibrium value  $c_{eq}$ , as the solubility line is reached. During this step, the high density phase is totally or partially consumed if additional phases are present, i.e., precipitates or crystals. We assume that the resulting solution at concentration  $c_{eq}$  is sufficiently dilute that we can treat it as an ideal solution with  $\gamma_{eq} = 1$ . The change in the chemical potential during this process is given by  $\Delta\mu = \mu(c_{eq}) - \mu(c_i) = \mu(c_{eq}) - \mu(c_i)$ , where we have employed the assumption that the chemical potential of the low density phase equals that of the initial solution (chemical equilibrium along the tie line of the LLPS). With the chemical potential given in Eq. (2), we can express change in the chemical potential as

$$\Delta\mu_a = RT \ln \left( \frac{c_{eq}}{\gamma_i c_i} \right), \quad (3)$$

where the index  $a$  in Eq. (3) indicates that we employ the activity formula of the chemical potential. Note that in this expression,  $c_i$  and  $c_{eq}$  are experimentally accessible, while  $\gamma_i$  is unknown.

Following the work of McMillan and Mayer,<sup>22</sup> we can also treat our system as an effective two-component system of a solvent, including ions, which we denote as component 1, and proteins, which is component 2. The chemical potential of component 1,  $\mu_1$ , can be expanded into a virial series in the concentration  $c$  of the proteins to obtain

$$\mu_1 = \mu_1^\ominus - RT V_1 c_2 \left[ \frac{1}{M} + B_2 c_2 + B_3 c_2^2 + \dots \right], \quad (4)$$

where  $V_1$  is the molar volume of the solvent and  $M$  is the molecular weight of the protein.  $B_2, B_3, \dots$  are the second, third, and higher virial coefficients. If the protein solution is sufficiently dilute, as we have assumed in the low density phase, it is possible to truncate the virial expansion at the second term.<sup>23</sup> By differentiating  $\mu_1$  with respect to the protein concentration,

$$\left( \frac{d\mu_1}{dc_2} \right)_{p,T} = - \frac{RT V_{m,1}}{M} [1 + 2B_2 M c_2], \quad (5)$$

and employing the Gibbs–Duhem relation,<sup>24</sup> it is possible to connect the chemical potential of the solvent with the chemical potential of the proteins to obtain

$$\left( \frac{d\mu_2}{dc_2} \right)_{p,T} = - \frac{M}{c_2 V_{m,1}} \left( \frac{d\mu_1}{dc_2} \right)_{p,T}. \quad (6)$$

Combining Eqs. (5) and (6), we can calculate the chemical potential, relative to a low density reference state with concentration  $c_0$  as

$$\mu(c) = \mu_0 + RT \ln \left( \frac{c}{c_0} \right) + 2RT B_2 M (c - c_0), \quad (7)$$

which is an alternative to Eq. (2), as long as the virial expansion can be truncated after the second term. We are now in the position to

express the change in the chemical potential from the initial solution at concentration  $c_i$  to its final state with concentration  $c_{eq}$  by

$$\Delta\mu_{MM} = RT \ln \left( \frac{c_{eq}}{c_i} \right) + 2RTB_2M(c_{eq} - c_i), \quad (8)$$

which is an alternative expression to Eq. (3), where  $\Delta\mu_{MM}$  indicates the chemical potential obtained from the McMillan–Mayer theory. By demanding that  $\Delta\mu_a = \Delta\mu_{MM}$ , we find that the second virial coefficient takes the form

$$B_2 = -\frac{\ln(\gamma_i)}{2M(c_{eq} - c_i)}, \quad (9)$$

which still contains the unknown activity coefficient  $\gamma_i$ .

In order to express the second virial coefficient solely by quantities that are accessible in the experiment, we consider again the chemical equilibrium between the initial solution and the low density phase along the tie line,  $\mu(c_i) = \mu(c_l)$ , from which the following equation is obtained:

$$\ln(\gamma_i) = \ln \left( \frac{\gamma_l c_l}{c_i} \right) \approx \ln \left( \frac{\gamma_{eq} c_{eq}}{c_i} \right) = \ln \left( \frac{c_{eq}}{c_i} \right), \quad (10)$$

where we have approximated the activity of the low density phase, which is not directly accessible, by that of the equilibrium solution, for which we have assumed that its activity coefficient  $\gamma_{eq} = 1$ . With this assumption, we reach our theoretical main result, an estimate of the second virial coefficient based on quantities that are easily accessible,

$$B_2 = -\frac{\ln \left( \frac{c_{eq}}{c_i} \right)}{2M(c_{eq} - c_i)}. \quad (11)$$

Note that assuming an interaction potential with a fixed interaction distance, like in the SAXS fitting procedure, or using a linear interpolation of the light scattering intensity, like in the DLS framework, the second virial coefficient should not depend on  $c_i$ , despite that dependence on initial concentration was already found in systems of cyclodextrins using a virial expansion with respect to the structure factor that follow the same concentration dependence described by Eq. (11).<sup>25,26</sup> In order to remove the concentration dependence, two assumptions can be adopted. The first one is to average the results over the concentration interval explored, keeping the salt–protein ratio constant, and the second one is to normalize the values with respect to the first concentration point at any given initial concentration where LLPS is showed, as described in Sec. IV D. The second virial coefficient obtained is normalized by the second virial coefficient of hard spheres.<sup>27</sup> In order to demonstrate the validity of Eq. (11), we apply it to different protein solutions that display LLPS, induced by trivalent salts.

### III. EXPERIMENTS AND METHODS

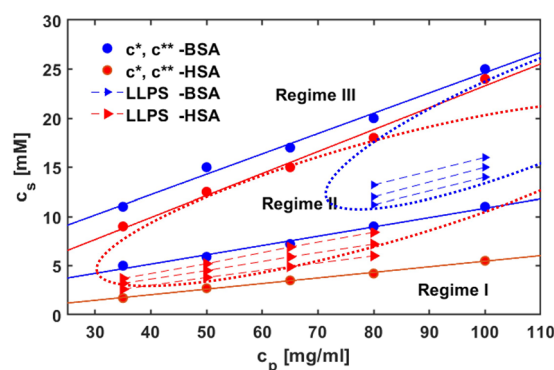
#### A. Materials and sample preparation

Proteins and salt were purchased from Merck and used as received. The purities were 98% for BSA (product no. A7906), 97% for HSA (product no. A9511), and 99.99% for  $\text{CeCl}_3$  (product no. 429406). Stock solutions were prepared by dissolving the protein

and salt in deionized (18.2 M $\Omega$ ), degassed Millipore water. The resulting concentration of protein was determined with an ultra-violet visible (UV–vis) spectrophotometer (Cary 50 UV–vis spectrometer, Varian Technologies) using an extinction coefficient of 0.667 ml mg<sup>−1</sup> cm<sup>−1</sup> for BSA and 0.531 ml mg<sup>−1</sup> cm<sup>−1</sup> for HSA at a wavelength of 278 nm.<sup>28</sup> All samples were prepared by mixing deionized, degassed Millipore water, protein stock solution, and salt stock solution. All samples had a pH (between 6.3 and 7.0) above the respective pI of HSA and BSA, measured with a pH-Meter from Mettler Toledo (Germany). No buffer was added as neutral trivalent salts (i.e.,  $\text{CeCl}_3$ ) do not induce significant pH variation. All samples were prepared, stored, and investigated at  $21 \pm 1^\circ\text{C}$ . When the LLPS is approached in the phase diagram (Fig. 1), the solution phase separates into a dense, yellowish phase and in a low density, dilute phase. In addition, for HSA, crystallization was also observed;<sup>29</sup> during this process, the dense phase is almost totally consumed.<sup>8</sup> A macroscopic picture of the vials during LLPS (BSA) is available in Ref. 19. After 14 days, the concentration of the resulting dilute equilibrium phase was determined via UV–vis spectroscopy.

#### B. Salt vs protein concentration phase diagram

Samples of protein concentrations ( $c_p$ ) at 35, 50, 65, 80, and 100 mg/ml were prepared for BSA and HSA, varying salt concentrations ( $c_s$ ). The mean value of  $c_s$  of the last clear and first turbid sample is referred to as  $c^*$ , and that of the last turbid and first clear sample is referred to as  $c^{**}$ . The macroscopic phase separation was ensured by visual inspection. No additional investigation with respect to the distribution of the salt in the two phases was made, and the concentration of the high density phase was not determined. Therefore, the points depicted in the phase diagram refer to the preparation conditions. For the complete knowledge of the LLPS, loop additional analysis is needed, such as those performed in Refs. 10, 15, and 31. The phase diagram depicted in Fig. 1 shows that for BSA, the amount of salt required to reach the  $c^*$  border is higher



**FIG. 1.** The experimental phase diagram displays triangular points that indicate the conditions where the samples inside the LLPS were prepared. Points that share the same  $c_s$  over  $c_p$  ratio are connected by dashed lines. The phase diagram boundaries are marked by full dots connected by solid lines. The  $c^*$  border is indicated by lower solid lines, while the upper solid lines represent the  $c^{**}$  border. Moreover, the phase diagram includes red and blue elliptical regions that represent a hypothetical complete LLPS loop for HSA and BSA.<sup>30</sup>

than for HSA. Such behavior could be due to the different surface charge of the proteins at such condition. BSA is slightly more negative with an overall charge of  $-11$  compared to the value of HSA, which is around  $-9$ .<sup>30</sup> Therefore, more salt has to be added to neutralize the excess negatively charged residues with counterions and trigger the aggregation. In addition, the different locations of the  $c^{**}$  borders could be explained as a consequence of an increased amount of counterions in BSA compared to HSA to trigger the pair repulsion (reentrant condensation).

### C. Small-angle x-ray scattering

SAXS measurements were performed at the Petra III beamline P12 (Hamburg, Germany). All the details related to the experimental procedures and the data analysis are available in Ref. 30.

For the determination of the reduced osmotic second virial coefficient  $B'_2 = B_2/B_{HS}$ , the data were fitted with a sticky hard sphere (SHS) model, provided in the NIST macros, using IGOR Pro 6.37.

In this model, the potential of mean force between proteins is assumed to be

$$\beta U(r) = \begin{cases} \infty, & r < \sigma, \\ -\beta U_o \ln \left( \frac{12\tau\Delta}{\sigma + \Delta} \right), & \sigma < r < \sigma + \Delta, \\ 0, & r > \sigma + \Delta, \end{cases} \quad (12)$$

where  $\tau$  is the stickiness parameter,  $\beta$  is the inverse temperature, and  $\Delta$  is the width of the square well. A perturbative solution of the Percus–Yevick closure relation was used to calculate the structure factor.<sup>32</sup> The analysis was performed using HSA and BSA solution with an initial concentration of 50 mg/ml. For BSA, no LLPS or crystallization was observed at such concentration; HSA instead displays a rich phase behavior with both crystallization and LLPS. These differences were already investigated in our previous studies.<sup>30</sup> The normalized second virial coefficient is thus calculated in the limit of  $\Delta \rightarrow 0$  using

$$\lim_{\Delta \rightarrow 0} \frac{B_2}{B_{HS}} = 1 - \frac{1}{4\tau}. \quad (13)$$

## IV. RESULTS AND DISCUSSION

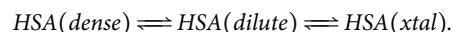
### A. Discussion of salt vs protein concentration phase diagram

The determination of the phase diagram was carried out by visual inspection. For small amounts of salt, the overall interactions between proteins are repulsive, resulting in a clear solution without macroscopic aggregation (regime I). Increasing the salt concentration the system reaches the boundary called  $c^*$  where the last clear sample and the first turbid were observed (regime II). Here, the interactions between proteins are mostly attractive due to the ion bridging effect.<sup>33</sup> Further addition of trivalent salt ( $\text{CeCl}_3$ ) moves the system toward the second boundary  $c^{**}$  where the resulting solution becomes clear again (regime III or reentrant). A possible explanation of this behavior is the charge inversion of the proteins due to the complete saturation of the binding sites with trivalent ions, leading to strong repulsive interactions.<sup>33</sup> The experimental data points are fitted using a linear regression. It is interesting to note that the criti-

cal threshold value of protein concentration for macroscopic LLPS is shifted by about two-three times if we compare HSA and BSA (35–40 and 72–80 mg/ml) for metastable LLPS.<sup>7,30</sup> This behavior reflects the trend observed in the calculation of the osmotic second virial coefficient from supersaturation measurements explained in the next chapter, where it is observed that the values of the second virial coefficient for BSA is 2–3 times bigger than the values of the HSA system, in agreement with the magnitude of the attractive interactions found in this work. If the interactions are not sufficiently attractive at low concentration of protein, an increasing protein concentration greatly shortens the distance between proteins, enhancing the likelihood of formation of a critical cluster large enough to allow the nucleation of a new phase. An additional threshold in concentration in regime II is reached at high protein concentration, where the system does not undergo phase separation. This might be due to additional short range repulsion that can affect the overall attractive interactions in crowded environments as well as different entropic contribution.<sup>17</sup> Additional investigations are needed to understand protein interactions in regime II at higher protein volume fraction, where additional phenomena are observed for several systems, such as kinetic aggregation and amorphous precipitation or gelation.<sup>34,35</sup>

### B. Real time observation of phase separation

The phase separation was followed by visual inspection in glass vials and on glass slides under the microscope. The general behavior observed reveals that close to the binodal, the solutions form a stable high density phase, in the case of BSA, and a meta-stable high density phase, in the case of HSA. This is connected with a different crystallization behavior of the two proteins. In fact, as shown in our previous work, a more stable crystalline phase exist for HSA at such condition, while for BSA, crystallization does not take place. This is due to the different intermolecular interactions between these two proteins; since the BSA is more hydrophilic, additional hydrophobic interactions necessary for crystallization are suppressed. Therefore, for the HSA solution, the LLPS is formed as a kinetically driven process upon salt addition followed by the formation of the crystals from the low density phase that is the thermodynamic stable phase as follows:



The system first phase separates in a high density and low density phase until a dynamic equilibrium is reached between them. They remain stable until the first crystals appear in the low density phase. Therefore, the variation of the osmotic pressure between the two previously coexistent phases leads to the consumption of the dense liquid droplets to preserve the initial equilibrium. However, if the driving force of the crystallization process is sufficiently strong, additional three phases equilibrium cannot be reached and the process continues until the complete high density phase is consumed and the equilibrium concentration  $c_{eq}$  is reached, which corresponds to the solubility limit.<sup>30</sup>

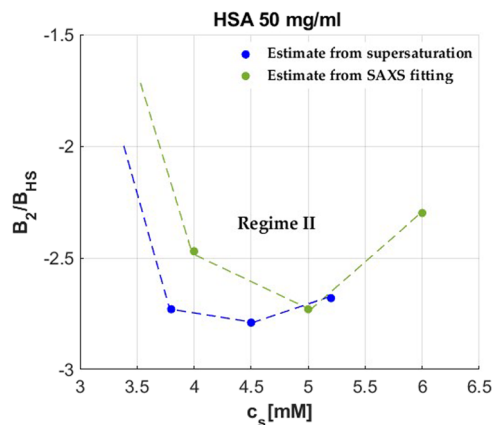
The growth of the crystals generally is completed between 7 and 10 days at these conditions. In order to ensure that the phase conversion was completed, a period of 14 days from the sample preparation was chosen for our supersaturation measurements.



### C. Calculation of the second virial coefficient from the supersaturation

Here, the experimental setup for the calculation of the second virial coefficient and the results are discussed. Upon the addition of  $\text{CeCl}_3$ , the samples in glass vials, sealed with parafilm, were kept at a temperature of  $21 \pm 1^\circ\text{C}$  for 14 days. This was done to ensure complete phase separation and crystallization. If HSA samples after preparation are centrifuged at 5000 rpm for 10 min and the high density phase is precipitated, the resulting concentration of the low density phase is always higher than the equilibrium concentration obtained 14 days after preparation. This behavior is essentially due to the crystallization that affects the final equilibrium concentration. For the BSA samples, which only show phase separation but no crystallization, the equilibrium concentration after 14 days is close to the values obtained after centrifuging. Therefore, to consider the total contribution (phase separation and crystallization) on the partitioning, only the values of the concentration obtained after 14 days are employed for the estimation of the osmotic second virial coefficient. Considering the normalized second virial coefficient obtained as a ratio of Eq. (11) and  $B_{HS}$  in  $\text{ml mol g}^{-2}$ , the only parameter not known experimentally is the hydrodynamic hard sphere radius. Hence, for the calculation of the hydrodynamic radius, a previously resolved crystallographic structure in the presence of  $\text{YCl}_3$  was used as an input in the software HullRad from Fluidic Analytics, which uses a convex hull model to estimate the hydrodynamic volume of a macromolecule.<sup>36</sup> Further details about this method are given in Refs. 36 and 37. The resulting value is estimated to be  $R_h = 3.55$  nm and was used together with a molecular weight of 66.4 kDa for the calculation of the normalized second virial coefficient of both HSA and BSA from the supersaturation measurements. A similar value for  $R_h$  was found experimentally using SAXS on BSA dilute solutions in Ref. 38, where the scattering curves are well fitted by a prolate ellipsoid with an effective hydrodynamic radii of  $R_h = 3.62$  nm. In the same reference, a value of  $R_h = 3.66$  nm obtained via DLS is also found. For the calculation of the second virial coefficient from the SAXS fitting procedure, an ellipsoid form factor was chosen. The ellipsoid form factor chosen was the same for both the proteins, with axes fixed to  $r_a = 1.8$  nm and  $r_b = 6.1$  nm. The results for the normalized osmotic second virial coefficient for HSA obtained from our approach and from SAXS measurements are plotted in Fig. 2. Both experiments were performed under similar conditions, and the overall agreement, despite the fact that in the different experiments  $B_2$  is extracted from different data, is good.

For BSA, the experimental conditions explored with SAXS were far from the LLPS border, and additional analysis shall be performed for direct comparison of the  $B'_2$  calculated with these two methods for the BSA- $\text{CeCl}_3$  system. However, a good agreement was found comparing the results obtained at 80 mg/ml from supersaturation measurements of BSA with  $\text{CeCl}_3$  with those obtained with the SAXS fitting procedure at 80 mg/ml BSA concentration with  $\text{LaCl}_3$  from Ref. 19, as shown in Table II. Clearly, the effect of salt should produce differences in the magnitude of  $B'_2$  due to effects such as differences in binding sites affinity on the surface of the protein or ions polarizability that leads to different intermolecular interactions.<sup>39,40</sup> We have chosen solutions with  $\text{CeCl}_3$  for comparison because these exhibit LLPS with respect both HSA and BSA, while for  $\text{LaCl}_3$ , we obtain LLPS with HSA, but not for BSA. In BSA with  $\text{LaCl}_3$ ,



**FIG. 2.** Normalized second virial coefficient estimate from supersaturation and SAXS fitting procedure; in the SAXS data points, the errors bars are smaller than the symbols shown here. For the points obtained by supersaturation measurements, the error is estimate to be around 10%—largely due to the uncertainty of the concentration measurements, determined via UV-vis spectroscopy at equilibrium. The dotted lines act as a guide for the eyes. Although the different routes to  $B_2$  make use of different data, the agreement between the estimates is good.

we observed transitions between regime I (clear)—regime II (turbidity without LLPS) and regime II (turbidity without LLPS)—regime III (reentrant). In principle, we could induce partitioning in BSA with  $\text{LaCl}_3$  by centrifuging the samples, while with  $\text{CeCl}_3$ , this is not necessary. Since our method relies only on thermodynamical assumptions, we apply the method exclusively on systems that show spontaneous driving force for LLPS. Independent of the type of protein used in this work, the general trend observed in both protein systems was an increase in the second virial coefficient as the protein concentration is increased, as shown in Tables I and II. It is not clear if this behavior is directly derived from the mass transfer treatment discussed in Sec. II or if it is because there are deviations from ideality at higher protein concentration and the approximation  $\gamma_{eq} = 1$  might not be valid even at the protein volume fraction of the low density equilibrium phase. However, this trend seems to describe the phase behavior of colloidal fluids near coexistence.<sup>41–43</sup> For HSA with  $\text{CeCl}_3$  at this condition, the solubility line is located at  $\approx 10$  mg/ml and the LLPS border is around  $\approx 30$  mg/ml. In comparison, the BSA solution is far more soluble than HSA and a smaller equilibrium concentration found was around 67 mg/ml. The values of the solubility line and LLPS borders depend of the initial protein concentration, salt concentration, and their ratio. Using the data obtained in Table I, as well as the initial concentration of HSA and the equilibrium one, it was possible to rebuild an experimental and qualitative phase diagram that resembles the theoretical diagram for colloidal fluids Fig. 3. For the complete phase diagram additional samples at higher protein concentration could be investigated as this might provide additional information for the observed trend of the second virial coefficient. In fact, from the theory of colloidal fluids, a turning point should be reached, corresponding to the critical point, and after that, a decrease in  $B'_2$  should be observed at higher protein volume fractions. However, this could not be detectable with

**TABLE I.** Normalized second virial coefficient values for HSA obtained via supersaturation. The results at 50 mg/ml are compared with those obtained from SAXS fitting. The error of the second virial coefficient obtained from supersaturation is about 10%.  $B'_2$ . Values of  $c_s/c_p$  are approximated by an integer number.

HSA				
$c_p$ (mg/ml)	$c_s$ (mM)	$c_s/c_p$	$B'_2(\text{supersaturation})$	$B'_2(\text{SAXS})$
35	2.6	5	−3.85	
35	3.2	6	−3.92	
35	3.7	7	−3.81	
50	3.8	5	−2.73	−2.47 ( $c_s/c_p = 5.3$ )
50	4.5	6	−2.79	−2.72 ( $c_s/c_p = 6.7$ )
50	5.2	7	−2.69	−2.31 ( $c_s/c_p = 8.0$ )
65	4.9	5	−2.24	
65	5.9	6	−2.31	
65	6.9	7	−2.28	
80	6	5	−1.86	
80	7.2	6	−1.94	
80	8.4	7	−1.88	

**TABLE II.** Normalized second virial coefficient values for BSA obtained via supersaturation. Note that no phase separation was observed at 50 mg/ml. The estimated error on the second virial coefficient is about 10%. For 80 mg/ml, the results are similar to those obtained with  $\text{LaCl}_3$  from Ref. 19. Values of  $c_s/c_p$  are approximated by an integer number.

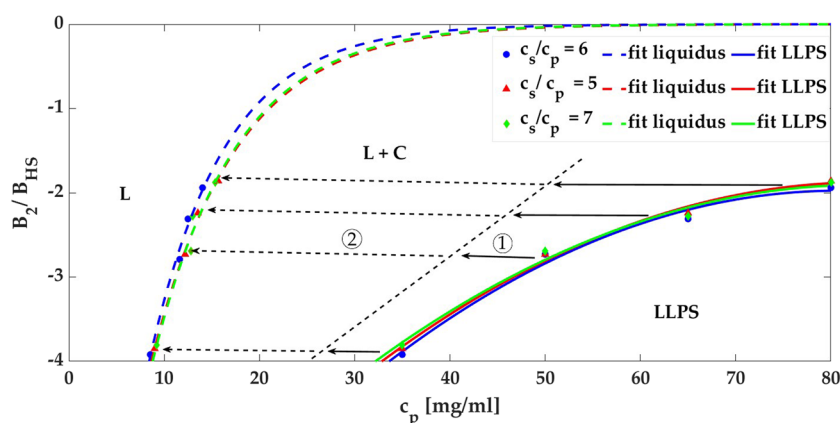
BSA				
$c_p$ (mg/ml)	$c_s$ (mM)	$c_s/c_p$	$B'_2(\text{supersaturation})$	$B'_2(\text{SAXS})$
50	7	9	NO	−2.41
50	8	11	LLPS	−2.43
50	9	12		−2.43
80	11	9	−1.0	
80	12	10	−1.0	
80	13	11	−0.96	
80	8	7		−1.0 ( $\text{LaCl}_3$ )
80	12	10		−1.5 ( $\text{LaCl}_3$ )
80	17	14		−1.2 ( $\text{LaCl}_3$ )
100	14	9	−0.79	
100	15	10	−0.8	
100	16	11	−0.83	

the previous assumptions for the limitation due to the approximations made. Furthermore, for a complete description of the phase diagram, the concentration in the high density phase should be determined so that the border at higher protein concentrations of the LLPS can also be calculated. In Fig. 3, a straight dotted line is used to describe the concentration of the low density phase after phase separation. This line could, in principle, be determined experimentally if the sample is centrifuged at 5000 rpm for 10 min immediately after salt addition to separate the contribution of LLPS (a kinetically driven process) to the contribution of the crystallization (a thermodynamically driven process).

#### D. Variation of the chemical potential near coexistence

In order to take into account variations of the chemical potential and to extend the range of application of the derived formula outside phase coexistence, it is possible to normalize the previous relation [Eq. (11)] with respect to the minimum concentration where LLPS is shown  $\tilde{c}$ ,

$$B_2 = -\frac{\ln\left(\frac{c_{eq}}{c_i}\right)}{2M(c_{eq} - c_i)} \frac{c_i}{\tilde{c}}, \quad (14)$$



**FIG. 3.** Experimental phase diagram of HSA with  $\text{CeCl}_3$ . The position of the solubility line (liquidus) follows the magnitude of the interactions and the different partitioning effect at different protein salt-ratios. For  $c_s/c_p = 6$ , the solubility of the protein diminishes due to the stronger attractions. The overall mechanism of LLPS and crystallization of HSA in the presence of  $\text{CeCl}_3$  might be explained as follows: the system from the initial preparation conditions phase separates in a low and in a high density phase (1). From the low density phase, the crystals start to nucleate and the consumption of material proceed until the solubility line is reached (2). The amount of high density and low density phase produced follow the arrows intensities.

where  $\tilde{c}$  is the lowest protein concentration for which LLPS is observed at a given temperature and pressure. Taking into account this change in the chemical potential, it is also possible to rescale the second virial coefficient in such a way that at constant supersaturation, it corresponds always to the same value of the second virial coefficient. Using this method, the concentration dependence vanishes. However, this is an empirical correction, and additional studies are needed to explore the validity of the assumptions done. The minimum concentration  $\tilde{c}$  where LLPS is shown for HSA at a salt-protein ratio of 6 is around 35–40 mg/ml, and for BSA, at a salt-protein ratio of 10, the minimum concentration  $\tilde{c}$  is around 72–80 mg/ml. By using those values, the normalized second virial coefficient becomes  $-3.4$  for HSA at initial concentrations of 35 and 50 mg/ml. The HSA system shows also an increased supersaturation by increasing protein concentration; by taking into account this behavior, the resulting second virial coefficient normalized for the minimum concentration where LLPS is shown  $\tilde{c}$  at initial concentrations of 65 and 80 mg/ml is  $\sim -3.7$ . For BSA, the initial-equilibrium concentration ratio is approximately the same at each condition explored, around 1.2. Normalizing such a system with respect to the critical concentration of 72 mg/ml, we obtain a normalized second virial coefficient of about  $-1.12$  close to the theoretical limit for phase separation of sticky hard sphere if specific energies routes are followed.<sup>10</sup> With this information, we speculate that for the BSA system, we are moving close to the critical point in the phase diagram, while for HSA, we observed the optimal crystallization window below the critical point, as predicted by Vliegthart and Lekkerkerker.<sup>9</sup>

## V. CONCLUSIONS

In conclusion, we calculated the normalized second virial coefficient for a protein solution close to LLPS using a thermodynamic approach for a two component system under the assumption that salt addition only changes the surface charge density of

the proteins. Furthermore, we linked the macroscopic behavior of protein partitioning with the magnitude of the second virial coefficient that is a free model calculation and experimentally accessible. The magnitude of the second virial coefficient follows the partitioning ratio between two phases as previously observed for a different system.<sup>10,19</sup> The mass transfer consideration might be, in principle, applied also outside the LLPS region when the system reaches the solubility line. However, this strongly depends of the protein phase behavior, the molecular weight and the precipitation agent are employed. Our expression was developed with the aim of being applicable to solutions containing two to three components, which exhibit both stable and metastable liquid-liquid phase separation (LLPS). This includes globular proteins with molecular weights ranging from 10 to 400 KDa, which covers all the proteins investigated in Ref. 39. We have compared our results with those obtained from a different method finding a good agreement. The main constraint of the thermodynamical relation adopted is its two-component approach. In fact, we access the chemical potential of the protein using the chemical potential of the solvent and the Gibbs-Duhem equation. As many solutions contain multiple components, such as buffers, salts, proteins, PEG (polyethylene glycol), water, preservatives, impurities, and surfaces of glass vials or containers, the LLPS process and the supersaturation measurements may be influenced by the presence of those compounds, decreasing the accuracy on the estimation of the second virial coefficient. Another constraint stems from the prerequisite understanding of the phase diagram and accurate identification of the binodal region. Additionally, smaller negatively charged proteins may experience the effects of bridging, resulting in significant kinetic amorphous aggregation or a greater driving force for crystallization than observed in the current system. When the interactions between proteins are too strong, the dynamic aspect of the interactions plays a crucial role on determining the final state of the system. At this condition, Monte Carlo simulations of particles with hard cores and isotropic, square-well interactions,



using the fluctuation–dissipation theorem, show that if no ordered bounds are present, like amorphous aggregation, the system do not relax in the more favorable thermodynamical state. Consequently, if not ordered clusters are formed, the bonding breaking energy is too high that the system is confined in an energetic trap.<sup>44</sup> In fact, the pathway for the formation of a new phase can be near equilibrium or far from it. The liquid–liquid phase separation process always competes with other forms of self-organization, including kinetic intermediates as previously discussed by Whitelam and Jack.<sup>45</sup> Under these conditions, the treatment used could become inaccurate. However, by taking into account the distance from the LLPS border, with the normalization used in Eq. (14), it might be possible to estimate the second virial coefficient also outside the binodal region. The accuracy of the treatment strongly decreases if after phase separation, the concentration of the protein in the low density phase is still high, probably due to protein–protein interactions still present in this phase that should be also taken into account. In this work, we have shown that under the conditions investigated, the magnitude of the second virial coefficient calculated via supersaturation measurements is in agreement with the values obtained via the SAXS fitting procedure. Working on the basic ideas presented here might lead to a different approach to estimate the second virial coefficient that do not make use of expansive and advanced techniques. The method, in fact, could be employed routinely in any protein crystallization laboratories, without the need of training to perform complicated experiments. Furthermore, the method could be useful not only for proteins–salt solutions but also for different colloidal systems that show crystallization or stable and metastable liquid–liquid phase separation.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support from the DFG and the allocation of beamtime as well as the support of Madeleine Fries and Alexander Gerlach. They appreciate fruitful discussions with Anita Girelli. They thank Hadra Banks, Cara Buchholz, and Lara Reichart for experimental assistance.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

## Author Contributions

**Furio Surfaro:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Ralph Maier:** Investigation (equal); Writing – review & editing (equal). **Kai-Florian Pastryk:** Data curation (supporting); Investigation (supporting). **Fajun Zhang:** Funding acquisition (supporting); Project administration (supporting); Supervision (supporting); Visualization (supporting); Writing – review & editing (supporting). **Frank Schreiber:** Funding acquisition (lead); Project administration (lead); Resources (lead); Supervision (supporting); Writing – review & editing (supporting).

**Roland Roth:** Conceptualization (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available within the article.

## REFERENCES

- <sup>1</sup> S. Alberti, A. Gladfelter, and T. Mittag, *Cell* **176**, 419 (2019).
- <sup>2</sup> A. Zbinden, M. Pérez-Berlanga, P. De Rossi, and M. Polymenidou, *Dev. Cell* **55**, 45 (2020).
- <sup>3</sup> K. You, Q. Huang, C. Yu, B. Shen, C. Sevilla, M. Shi, H. Hermjakob, Y. Chen, and T. Li, *Nucleic Acids Res.* **48**, D354 (2020).
- <sup>4</sup> F. Zhang, F. Roosen-Runge, A. Sauter, R. Roth, M. W. A. Skoda, R. M. J. Jacobs, M. Sztucki, and F. Schreiber, *Faraday Discuss.* **159**, 313 (2012).
- <sup>5</sup> A. A. Hyman, C. A. Weber, and F. Jülicher, *Annu. Rev. Cell Dev. Biol.* **30**, 39 (2014).
- <sup>6</sup> H. Zhang, X. Ji, P. Li, C. Liu, J. Lou, Z. Wang, W. Wen, Y. Xiao, M. Zhang, and X. Zhu, *Sci. China: Life Sci.* **63**, 953 (2020).
- <sup>7</sup> R. Maier, G. Zocher, A. Sauter, S. Da Vela, O. Matsarskaia, R. Schweins, M. Sztucki, F. Zhang, T. Stehle, and F. Schreiber, *Cryst. Growth Des.* **20**, 7951 (2020).
- <sup>8</sup> R. Maier, B. Sohmen, S. Da Vela, O. Matsarskaia, C. Beck, R. Schweins, T. Seydel, F. Zhang, and F. Schreiber, *Cryst. Growth Des.* **21**, 6971 (2021).
- <sup>9</sup> G. A. Vliegthart and H. N. W. Lekkerkerker, *J. Chem. Phys.* **112**, 5364 (2000).
- <sup>10</sup> M. Wolf, F. Roosen-Runge, F. Zhang, R. Roth, M. W. A. Skoda, R. M. J. Jacobs, M. Sztucki, and F. Schreiber, *J. Mol. Liq.* **200**, 20 (2014).
- <sup>11</sup> F. Platten, N. E. Valadez-Pérez, R. Castañeda-Priego, and S. U. Egelhaaf, *J. Chem. Phys.* **142**, 174905 (2015).
- <sup>12</sup> A. George and W. W. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 361 (1994).
- <sup>13</sup> D. J. Winzor, M. Deszczynski, S. E. Harding, and P. R. Wills, *Biophys. Chem.* **128**, 46 (2007).
- <sup>14</sup> M. Deszczynski, S. E. Harding, and D. J. Winzor, *Biophys. Chem.* **120**, 106 (2006).
- <sup>15</sup> F. Zhang, R. Roth, M. Wolf, F. Roosen-Runge, M. W. A. Skoda, R. M. J. Jacobs, M. Sztucki, and F. Schreiber, *Soft Matter* **8**, 1313 (2012).
- <sup>16</sup> J. Narayanan and X. Y. Liu, *Biophys. J.* **84**, 523 (2003).
- <sup>17</sup> F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, R. A. Martin, C. M. Martin, and F. Schreiber, *J. Phys. Chem. B* **111**, 251 (2007).
- <sup>18</sup> C. G. De Kruijff, P. W. Rouw, W. J. Briels, M. H. G. Duits, A. Vrij, and R. P. May, *Langmuir* **5**, 422 (1989).
- <sup>19</sup> O. Matsarskaia, F. Roosen-Runge, G. Lotze, J. Möller, A. Mariani, F. Zhang, and F. Schreiber, *Phys. Chem. Chem. Phys.* **20**, 27214 (2018).
- <sup>20</sup> F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, S. Zorn, R. A. Martin, C. M. Martin, G. F. Clark, S. Weggler, A. Hildebrandt, O. Kohlbacher *et al.*, *Phys. Rev. Lett.* **101**, 148101 (2008).
- <sup>21</sup> F. Zhang, J. A. Gavira, G. W. Lee, and D. Zahn, *Crystals* **11**, 174 (2021).
- <sup>22</sup> W. G. McMillan, Jr. and J. E. Mayer, *J. Chem. Phys.* **13**, 276 (1945).
- <sup>23</sup> B. Guo, S. Kao, H. McDonald, A. Asanov, L. L. Combs, and W. William Wilson, *J. Cryst. Growth* **196**, 424 (1999).
- <sup>24</sup> P. Duhem, *J. Phys. Chem.* **2**, 1 (1898).
- <sup>25</sup> A. Kusmin, R. E. Lechner, M. Kammel, and W. Saenger, *J. Phys. Chem. B* **112**, 12888 (2008).
- <sup>26</sup> A. Kusmin, Ph.D. thesis, Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, 2007.
- <sup>27</sup> F. Schreiber, F. Zanini, and F. Roosen-Runge, available at [www.soft-matter.uni-tuebingen.de/teaching/Tutorial\\_Virial\\_Expansion.pdf](http://www.soft-matter.uni-tuebingen.de/teaching/Tutorial_Virial_Expansion.pdf); accessed 5 June 2018, 2011.
- <sup>28</sup> G. D. Fasman, *CRC Handbook of Biochemistry and Molecular Biology* (CRC Press, 2018).

- <sup>29</sup>C. Buchholz, L. F. Reichart, F. Surfaro, R. Maier, F. Zhang, A. Gerlach, and F. Schreiber, *J. Cryst. Growth* **603**, 126959 (2023).
- <sup>30</sup>R. Maier, M. R. Fries, C. Buchholz, F. Zhang, and F. Schreiber, *Cryst. Growth Des.* **21**, 5451 (2021).
- <sup>31</sup>O. Matsarskaia, M. K. Braun, F. Roosen-Runge, M. Wolf, F. Zhang, R. Roth, and F. Schreiber, *J. Phys. Chem. B* **120**, 7731 (2016).
- <sup>32</sup>R. J. Baxter, *J. Chem. Phys.* **49**, 2770 (1968).
- <sup>33</sup>F. Roosen-Runge, F. Zhang, F. Schreiber, and R. Roth, *Sci. Rep.* **4**, 7016 (2014).
- <sup>34</sup>M. A. Dessau and Y. Modis, *J. Visualized Exp.* **16**, e2285 (2011).
- <sup>35</sup>L. R. De Young, A. L. Fink, and K. A. Dill, *Acc. Chem. Res.* **26**, 614 (1993).
- <sup>36</sup>P. J. Fleming and K. G. Fleming, *Biophys. J.* **114**, 856 (2018).
- <sup>37</sup>C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, *ACM Trans. Math. Software* **22**, 469 (1996).
- <sup>38</sup>F. Roosen-Runge, M. Hennig, F. Zhang, R. M. J. Jacobs, M. Sztucki, H. Schober, T. Seydel, and F. Schreiber, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11815 (2011).
- <sup>39</sup>A. C. Dumetz, A. M. Snellinger-O'Brien, E. W. Kaler, and A. M. Lenhoff, *Protein Sci.* **16**, 1867 (2007).
- <sup>40</sup>R. A. Curtis, J. Ulrich, A. Montaser, J. M. Prausnitz, and H. W. Blanch, *Biotechnol. Bioeng.* **79**, 367 (2002).
- <sup>41</sup>S. V. G. Menon, C. Manohar, and K. S. Rao, *J. Chem. Phys.* **95**, 9186 (1991).
- <sup>42</sup>H. Verduin and J. K. G. Dhont, *J. Colloid Interface Sci.* **172**, 425 (1995).
- <sup>43</sup>M. A. Miller and D. Frenkel, *J. Chem. Phys.* **121**, 535 (2004).
- <sup>44</sup>D. Klotsa and R. L. Jack, *Soft Matter* **7**, 6294 (2011).
- <sup>45</sup>S. Whitelam and R. L. Jack, *Annu. Rev. Phys. Chem.* **66**, 143 (2015).