Reexamining protein–protein and protein–solvent interactions from Kirkwood-Buff analysis of light scattering in multi-component solutions

Marco A. Blanco,[1] Erinc Sahin,[1] Yi Li,[1,a)] and Christopher J. Roberts[1,2,b)]
[1]*Department of Chemical Engineering and Center for Molecular and Engineering Thermodynamics, University of Delaware, Newark, Delaware 19716, USA*
[2]*National Institute of Standards and Technology Center for Neutron Research, Gaithersburg, Maryland 20899, USA*

(Received 26 January 2011; accepted 13 May 2011; published online 10 June 2011)

The classic analysis of Rayleigh light scattering (LS) is re-examined for multi-component protein solutions, within the context of Kirkwood-Buff (KB) theory as well as a more generalized canonical treatment. Significant differences arise when traditional treatments that approximate constant pressure and neglect concentration fluctuations in one or more (co)solvent/co-solute species are compared with more rigorous treatments at constant volume and with all species free to fluctuate. For dilute solutions, it is shown that LS can be used to rigorously and unambiguously obtain values for the osmotic second virial coefficient (B_{22}), in contrast with recent arguments regarding protein interactions deduced from LS experiments. For more concentrated solutions, it is shown that conventional analysis over(under)-estimates the magnitude of B_{22} for significantly repulsive(attractive) conditions, and that protein-protein KB integrals (G_{22}) are the more relevant quantity obtainable from LS. Published data for α–chymotrypsinogen A and a series of monoclonal antibodies at different pH and salt concentrations are re-analyzed using traditional and new treatments. The results illustrate that while traditional analysis may be sufficient if one is interested in only the sign of B_{22} or G_{22}, the quantitative values can be significantly in error. A simple approach is illustrated for determining whether protein concentration (c_2) is sufficiently dilute for B_{22} to apply, and for correcting B_{22} values from traditional LS regression at higher c_2 values. The apparent molecular weight $M_{2,app}$ obtained from LS is shown to generally not be equal to the true molecular weight, with the differences arising from a combination of protein-solute and protein-cosolute interactions that may, in principle, also be determined from LS. © *2011 American Institute of Physics*. [doi:10.1063/1.3596726]

I. INTRODUCTION

Light scattering (LS) from protein solutions has long been used to provide information regarding protein-protein interactions[1–7] via the protein osmotic second virial coefficient (B_{22}). In keeping with common practice, the Scatchard notation for aqueous protein solutions is adopted here, with 1 denoting water, 2 denoting protein, and $i = 3, 4, \ldots$ denoting any co-solvent and co-solute species.[8] B_{22} is formally related to protein–protein interactions in the limit of low protein concentration (c_2), averaged over the spatial degrees of freedom of the solvent and any co-solute or co-solvent species – i.e., the protein–protein potential of mean force W_{22} in a grand-canonical ensemble[9] via

$$B_{22} = -\frac{1}{2}\int (e^{-W_{22}/k_BT} - 1)\, 4\pi r^2 dr, \qquad (1)$$

where k_B is the Boltzmann constant, T is the absolute temperature, and r denotes distance between centers-of-mass. Although one can derive a similar equation to Eq. (1) in any thermodynamic ensemble, the osmotic second virial coefficient is recovered only when the protein–protein potential of mean force comes from a grand-canonical ensemble.[9] Otherwise the integral has a different thermodynamic meaning. For example, by using W_{22} from a canonical ensemble (fixed temperature, composition, and volume) to solve the integral in Eq. (1), one obtains the trivial solution[9] that the integral is independent of the magnitude or sign of the interactions described by W_{22}. The interested reader is referred to detailed discussions and derivations elsewhere (e.g., Refs. 9 and 10).

B_{22} and other osmotic virial coefficients play central roles in both qualitative and quantitative models and theories relating colloidal protein-protein interactions to protein crystallization and fluid-fuid phase behavior,[11–18] protein aggregation,[19–24] and protein purification.[25–27] In some cases, only the sign of B_{22} is considered important, as a negative (positive) B_{22} corresponds to net attractive (repulsive) protein-protein interactions relative to an ideal solution.[9,10,28] More recent work highlights the importance of considering an equivalent hard-sphere (HS) or purely steric protein-protein interaction (i.e., B_{22}^{HS}) as a more appropriate reference point.[24,29–31] That is, systems with $B_{22}/B_{22}^{HS} > 1$ have non-steric interactions that are dominated by non-steric repulsions,

a)Present address: Merck Research Laboratories, West Point, Pennsylvania 19446, USA.
b)Author to whom correspondence should be addressed. Electronic mail: cjr@udel.edu.

while those with $B_{22}/B_{22}^{HS} < 1$ are dominated by attractive interactions.

Recent experimental comparisons between LS and alternative means to estimate B_{22} have led some authors to conclude that LS fundamentally does not yield B_{22} as an independent quantity – i.e., that the apparent virial coefficient (A_2) from traditional LS analysis is instead a convolution of B_{22} with protein-cosolute and/or (co)solvent virial coefficients.[32,33] This is potentially problematic, as LS is arguably the historic method of choice for determining B_{22}, and is the basis for much if not the majority of experimental B_{22} values reported in the literature.[34–37]

The original treatments by Stockmayer and others form the basis of most non-trivial derivations to relate Rayleigh scattering to deviations from ideal solution for any solute-solvent system.[38–41] The approach proposed by Stockmayer has been the most widely used and accepted.[42–46] For thermodynamic convenience and mathematical simplicity, Stockmayer's treatment and those of others make the simplifying assumption that fluctuations in pressure (p) and solvent (water) concentration are negligible within the scattering volume. That is, the scattering (sub)volume is treated as though it contains a constant number of water molecules, and it is implicitly assumed that one can neglect the differences between constant pressure and constant volume for the scattering region of the fluid. Corrections have also been applied to account for requirements of electroneutrality and resulting Donnan effects on A_2,[44] particularly for the case of protein solutions. As noted above, recent derivations[32,33] have been proposed that argue A_2 from LS contains contributions from other osmotic virial coefficients, or from Donnan contributions via constraints of electroneutrality.[44] Alternatively, Casassa and Eisenberg[47,48] corrected Stockmayer's treatment by imposing electroneutrality through the equilibration of protein with cosolutes within the scattering volume (a correction which was implicitly obtained herein). Nevertheless, in all of the above cases, fluctuations in pressure and solvent concentration are neglected, and fluctuations in the remaining species are related to LS intensity via thermodynamic transformations that express non-idealities in terms of activity coefficients,[38,39,42,44,49–51] implying, as it is shown below, that information about interactions between solvent and all the other components is not recovered explicitly, but rather it is convoluted within A_2 and the apparent molecular weight.

This report examines the result of relaxing these assumptions, as well as utilizing Kirkwood-Buff (KB) solution theory[52] to rigorously and more generally relate LS to protein-protein, protein-solvent, and protein-cosolute interactions. The results highlight errors or unnecessary approximations in alternative treatments, and also provide a new mathematical description of LS that in principle is valid at both low and high protein concentration, with a straightforward means to extract protein-protein KB integrals (G_{22}) from LS data. The remainder of the article is organized as follows. In Sec. II, a general equation for Rayleigh scattering from multicomponent solutions is developed in terms of KB integrals. Dilute and semi-dilute protein solutions are then considered in Sec. III, including a comparison with a more general version of the traditional, non-KB approach that is strictly valid only for non-ionizable solutes. A more general expression is also derived in the canonical framework that does not assume $c_2 \to 0$. Additionally, the supplementary material[53] provides a comparison with the simplest and the most traditional thermodynamic analysis with non-dissociable solutes to show equivalence between the two approaches as $c_2 \to 0$ if fluctuations in pressure and solvent species are not neglected, and if protein-solvent and protein-cosolute interactions can be neglected. Finally Sec. IV uses the working equations from the KB and more general canonical approaches, as well as the standard expression,[54] and analyzes published LS data for α–chymotrypsinogen A (aCgn) and a set of monoclonal antibodies (mAb)[24,31] to highlight differences in the resulting B_{22} (or G_{22}) and apparent molecular weight values as a function of solvent pH and NaCl concentration for experimental systems.

II. KIRKWOOD-BUFF THEORY APPLIED TO RAYLEIGH SCATTERING FROM MULTICOMPONENT SOLUTIONS

Laser light scattering from multicomponent single-phase solutions due to Rayleigh scattering is described by the theory of Einstein,[55] with each species treated as a point scatter so long as its characteristic dimension is sufficiently small compared to the wavelength of the incident light.[28,38,39,56,57] In the case of laser light scattering, light is scattered from a small but macroscopic volume V within a bulk solution, with time-averaged scattered intensity that is proportional to the magnitude of ensemble-averaged fluctuations in the dielectric constant, or equivalently fluctuations in refractive index (n),

$$R_{90} = \frac{4\pi^2 n^2 \langle (\Delta n)^2 \rangle V}{\lambda^4}, \tag{2}$$

where R_{90} is the Rayleigh ratio for a 90 degree scattering angle, and λ is the wavelength of incident light *in vacuo*. The brackets $\langle \cdots \rangle$ denote an ensemble average. The fluctuations in refractive index within the scattering volume are expressed based on the total differential for n as a function of temperature (T), volume (V), and the number of molecules of each species (N_j, $j = 1, 2, \ldots$). For fixed scattering volume, and assuming negligible fluctuations in temperature, this gives

$$\Delta n = (n - \bar{n}) \approx dn = \sum_j \left(\frac{\partial n}{\partial N_j} \right)_{T,V,N_{k \neq j}} dN_j. \tag{3}$$

Squaring Eq. (3) and ensemble averaging gives

$$\begin{aligned} \langle \Delta n^2 \rangle = & \sum_i \left(\frac{\partial n}{\partial N_i} \right)^2_{T,V,N_{k \neq i}} \left(\langle N_i^2 \rangle - \langle N_i \rangle^2 \right) \\ & + 2 \sum_{i<j} \left(\frac{\partial n}{\partial N_i} \right)_{T,V,N_{k \neq i}} \left(\frac{\partial n}{\partial N_j} \right)_{T,V,N_{k \neq j}} \\ & \times \left(\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle \right). \end{aligned} \tag{4}$$

Because the scattering volume is open to exchange of all species and is constant volume and effectively constant temperature, the fluctuations in species concentrations are most easily and naturally expressed in terms of Kirkwood-Buff (KB) integrals (G_{ij}) in the grand canonical

ensemble[9,52,58]

$$\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle = \langle N_i\rangle\left(\delta_{ij} + \frac{\langle N_j\rangle}{V}G_{ij}\right), \tag{5}$$

with δ_{ij} denoting the Kronecker delta function, the brackets denoting the average within the grand canonical ensemble for the scattering volume, and with G_{ij} defined by

$$G_{ij} = \int (\bar{g}_{ij}(r) - 1)\,4\pi r^2 dr, \tag{6}$$

where $\bar{g}_{ij}(r)$ is the molecular pair correlation function for component i with respect to component j. It gives the probability, relative to an ideal mixture, of finding an $i-j$ pair of molecules at a distance r between centers-of-mass, averaged over the orientations of species i and j, and averaged over the possible positions and orientations of all other species in the mixture. The overbar indicates that the averaging must be done in an open system, i.e., with fixed chemical potential (μ) of each of components i and j, and *not* with fixed N_i or N_j.[9,58] This is equivalent to a weighted average over the canonical pair correlation functions $g_{ij}(r;N_i,N_j)$ across all possible values of (N_i,N_j).[9] That is, $g_{ij}(r;N_i,N_j)$ is the pair correlation function in the closed ensemble at fixed (N_i,N_j). The overbar indicates an average over $p(N_i,N_j)g_{ij}(r;N_i,N_j)$, with $p(N_i,N_j)$ denoting the equilibrium probability of observing a particular set of (N_i,N_j) values within the open (i.e., grand canonical) ensemble.[58] The average values for each species are fixed by the bulk composition of the solution, within which the scattering volume is a small subsystem. Equations (5) and (6) are general; they do not require assumptions of pairwise additivity, and inherently incorporate restrictions of electroneutrality for ionic solutions, provided that the system volume is open to exchange of all ionized species.

Combining Eqs. (2), (4), and (5) gives upon rearrangement

$$\frac{R_{90}}{K'} = \sum_i \left(\frac{\partial n}{\partial c_i}\right)^2_{T,V,N_{k\neq i}} c_i(c_i G'_{ii} + 1) + 2\sum_{i<j}\left(\frac{\partial n}{\partial c_i}\right)_{T,V,N_{k\neq i}}\left(\frac{\partial n}{\partial c_j}\right)_{T,V,N_{k\neq j}} c_i c_j G'_{ij}, \tag{7}$$

where $c_i = N_i/(VN_A)$ is the concentration of component i on a mole/volume basis, and the prime on each G_{ij} indicates that the KB integrals have units of volume/mole. K' is given by $4\pi^2 n^2 N_A^{-1}\lambda^{-4}$, with N_A denoting Avogadro's number. The prime is to distinguish this from the closely related quantity K that appears in traditional treatments of LS.

Experimentally, changes in n with bulk solvent and solute concentrations are more easily evaluated at fixed pressure (p), rather than at fixed V. Transforming the derivatives of refractive index from constant V to constant p gives

$$\left(\frac{\partial n}{\partial \mathbf{n_i}}\right)_{T,V,\mathbf{n_{k\neq i}}} = \left(\frac{\partial n}{\partial \mathbf{n_i}}\right)_{T,p,\mathbf{n_{k\neq i}}} + \left(\frac{\partial n}{\partial p}\right)_{T,\{\mathbf{n_k}\}}\frac{\bar{v}_i}{\langle V\rangle\kappa_T}, \tag{8}$$

where $\{\mathbf{n_k}\}$ denotes the set of all mole numbers, $\bar{v}_i$ the partial molar volume of component i, κ_T the isothermal compressibility of the solution, and $\langle V\rangle$ the average volume of a solution with $\{\mathbf{n_k}\}$ at the T and p of interest.

Using Eq. (8) in Eq. (7) gives

$$\frac{R_{90}}{K'} = \sum_i \eta_i^2 c_i(c_i G'_{ii} + 1) + 2\sum_{i<j}\eta_i\eta_j c_i c_j G'_{ij}, \tag{9}$$

with

$$\eta_i = (1 - c_i\bar{v}_i)\left(\frac{\partial n}{\partial c_i}\right)_{T,p,c_{k\neq i}} + \left(\frac{\partial n}{\partial p}\right)_{T,\{\mathbf{n_k}\}}\frac{\bar{v}_i}{\kappa_T}. \tag{10}$$

At dilute conditions of component i, the term $c_i\bar{v}_i$ in Eq. (10) can be neglected. Equations (9) and (10) apply generally, and show that Rayleigh scattering as a function of solvent composition and solute concentrations can be described succinctly in terms of KB integrals, the dependence of n on the bulk concentrations of solvent and solute(s), the volumetric properties of the solution, and the dependence of n on pressure. The magnitude and sign of the difference between treating scattering as a constant pressure process instead of constant volume is then determined by the second term in Eq. (10), and this quantity is squared in the expression for Rayleigh scattering. In a sense, this indicates the magnitude of the errors introduced by neglecting pressure fluctuations, and thereby assuming one can replace η_i with $(1 - c_i\bar{v}_i)(\partial n/\partial c_i)_{T,p,c_{k\neq i}}$.

III. RAYLEIGH SCATTERING IN DILUTE PROTEIN SOLUTIONS

A. Protein-protein and protein-solvent interactions from KB analysis

Consider a ϑ - component mixture of water (component 1), protein (component 2), and $\vartheta - 2$ cosolute or cosolvent species (components 3, 4,...). Let R_0 be the scattering at 90 degrees from an equivalent solution at zero protein concentration, and let $R_{90}^{ex} = R_{90} - R_0$ denote the excess Rayleigh ratio. Assuming that c_2 is sufficiently small that G_{ij} $(i,j \neq 2)$ values are the same for the protein solution and the protein-free solution,[58] and converting to the more experimentally convenient concentration units of mass/volume, R_{90}^{ex} can be written with the aid of Eqs. (9) and (10) as

$$\frac{R_{90}^{ex}}{K'\left(\eta_2^m\right)^2} = M_{2,app}c_2^m + M_2 G_{22}^m\left(c_2^m\right)^2, \tag{11a}$$

$$\left.\frac{R_{90}^{ex}}{K'\left(\eta_2^m\right)^2}\right|_{c_2\to 0} = M_{2,app}c_2^m - 2M_2 B_{22}^m\left(c_2^m\right)^2, \tag{11b}$$

with the apparent molecular weight ($M_{2,app}$) given by

$$M_{2,app} = M_2\left[1 + 2\sum_{i\neq 2}\left(\frac{\eta_i^m}{\eta_2^m}\right)c_i^m G_{2i}^m\right]. \tag{12}$$

In the above expressions, M_2 is the protein molecular weight, G_{2j}^m $(= G'_{2j}/M_2)$ is G_{2j} in units of volume per mass-of-protein, and the superscript m denotes that concentrations and derivatives with respect to concentration are expressed on a w/v basis. The summation in Eq. (12) can be either positive or negative, therefore the apparent molecular weight can be either larger or smaller than M_2. In the limit of $c_2 \to 0$, G'_{22} can be replaced[9] by $-2B_{22}$ in Eq. (11a), leading to Eq. (11b).

For an ideal solution, G'_{22} and G'_{2i} are identically zero; yielding the classical ideal result $R_{90}^{ex} \propto M_2 c_2^m$.[38,39,42]

The results above clearly show that excess Rayleigh scattering is directly and unambiguously related to B_{22} (or more generally to G_{22}) if one utilizes the KB-based analysis of the scattering data, and there is no fundamental convolution of B_{22} or G_{22} with other osmotic virial coefficients or KB integrals. Subsection III B examines this further from the perspective of a more canonical treatment of Rayleigh scattering. Notably, there is no assumption regarding non-dissociable cosolutes or electroneutrality in Eq. (11). That is, scattering occurs in a grand canonical ensemble with the fluctuations of all species obeying the constraint of constant chemical potential for each species. The fluctuations within the system therefore cannot violate electroneutrality, as doing so would result in an effectively infinite chemical potential for one or more species within the scattering volume. As such, Eq. (11) holds for both ionizable and non-ionizable solutes.

Equation (11) also suggests it may be possible to assess protein-solvent/cosolute interactions via the dependence of $M_{2,app}$ on solvent composition. However, in the limit of $c_2^m \to 0$, the values of $c_{i\neq 2}^m$ are not independent. Thus, it may be difficult to determine independent values of G_{2j} (all $j \neq 2$) from light scattering alone. A similar conclusion holds in what follows in Subsection III B. Finally, Eq. (11) is not limited to highly dilute protein solutions, in that G'_{2j} depends on c_2^m at higher protein concentrations. The remainder of this report focuses on dilute or semi-dilute protein solutions, with a more detailed treatment of highly concentrated solutions left to a future report.

B. Revisiting the canonical treatment of Rayleigh scattering

In order to provide the relation between Rayleigh scattering and thermodynamic quantities such as activity coefficients, Eq. (4) can be expressed alternatively using an approach similar to that employed by Stockmayer[38] or Kirkwood and Goldberg,[39] but without making the simplifying assumption that one can neglect the differences between constant pressure and constant volume, and not neglecting fluctuations of any species in the mixture. By using the identity[9]

$$\frac{RT}{V}\left(\frac{\partial c_i}{\partial \mu_j}\right)_{T,V,\mu_{k\neq j}} = \langle c_i c_j\rangle - \langle c_i\rangle\langle c_j\rangle. \tag{13}$$

Equation (4) can equivalently be written as

$$\langle(\Delta n)^2\rangle = \frac{RT}{V}\sum_i\sum_j \eta_i\eta_j\left(\frac{\partial c_i}{\partial \mu_j}\right)_{T,V,\mu_{k\neq j}}, \tag{14}$$

where η_i is defined in Eq. (10), μ_k is the chemical potential of the kth component in the solution, and R is the ideal gas constant. It is possible to express the derivatives in Eq. (14) in terms of derivatives at constant $c_{k\neq i}$ rather than constant $\mu_{k\neq j}$. To do so, one can begin with the set of differential equations

$$d\mu_k = \sum_i a_{ik} dc_i$$

with

$$a_{ik} = \left(\frac{\partial \mu_i}{\partial c_k}\right)_{T,V,c_{j\neq k}} = \left(\frac{\partial \mu_k}{\partial c_i}\right)_{T,V,c_{j\neq i}} = a_{ki}. \tag{15}$$

For neutral, non-dissociable cosolutes (components 3,4,...),[38,44] the solution to this set of simultaneous equations can be expressed as[9,38,39]

$$\left(\frac{\partial c_i}{\partial \mu_j}\right)_{T,V,\mu_{k\neq j}} = \frac{A_{ij}}{|\,A\,|} \tag{16}$$

$$a_{ij} = \left(\frac{\partial \mu_i}{\partial c_j}\right)_{T,V,c_{k\neq j}},$$

where $|\,A\,|$ represents the determinant of the matrix $\{a_{ij}\}$, and A_{ij} is the co-factor of the element a_{ij} in this determinant. Combining Eqs. (2), (14), and (16) gives

$$\frac{R_{90}}{K'} = RT\sum_i\sum_j \eta_i\eta_j\frac{A_{ij}}{|\,A\,|}. \tag{17}$$

For concreteness, consider a three-component mixture, using the same notation as in Subsection III A. Equation (17) can then be rearranged to

$$\frac{R_{90}}{K'} = \frac{\left[\eta_1\left(\frac{A_{21}}{A_{22}}\right)+\eta_2+\eta_3\left(\frac{A_{32}}{A_{22}}\right)\right]^2}{\frac{1}{RT}\left[a_{12}\left(\frac{A_{21}}{A_{22}}\right)+a_{22}+a_{23}\left(\frac{A_{32}}{A_{22}}\right)\right]} + \frac{\left[\eta_3-\eta_1\left(\frac{a_{13}}{a_{11}}\right)\right]^2}{\frac{1}{RT}\left[a_{33}-\frac{a_{13}^2}{a_{11}}\right]} + \frac{\eta_1^2}{a_{11}/RT}. \tag{18}$$

Defining R_0 and R_{90}^{ex} as above, and making the same approximations as in Sec. II when subtracting contributions from the solvent background, gives

$$\frac{R_{90}^{ex}}{K'} = \frac{\left[\eta_1\left(\frac{A_{21}}{A_{22}}\right)+\eta_2+\eta_3\left(\frac{A_{32}}{A_{22}}\right)\right]^2}{\frac{1}{RT}\left[a_{12}\left(\frac{A_{21}}{A_{22}}\right)+a_{22}+a_{23}\left(\frac{A_{32}}{A_{22}}\right)\right]}. \tag{19}$$

Note the term in brackets in the numerator of the right hand side of Eq. (19) is equivalent to applying a Legendre transform to η_2 in order to obtain $(\partial n/\partial c_2)_{T,V,\mu_{j\neq 2}}$. By doing so, one recoverers a similar term to that proposed by Casassa and Eisenberg[47,48] as a correction to Stockmayer's derivation, with the difference being that fluctuations in water concentration are not neglected in the present case. From a practical perspective, this would be equivalent to neglecting terms accounting for protein-solvent interactions, and assuming that fluctuations in protein concentration and co-solute concentration are coupled, but those involving water are not coupled (see also the further discussion below on this point).

Defining the protein activity coefficient (γ_2) using a molar reference state gives

$$RT\ln\gamma_2 = \mu_2^{ex} = \mu_2 - \mu_2^{id} \tag{20}$$

with

$$\mu_2^{id} = \mu_2^0 + RT\ln c_2$$

and μ_2^0 denoting the protein standard state chemical potential. Doing so allows one to express the denominator on the right hand side of Eq. (19) as

$$\frac{1}{RT}\left[a_{12}\left(\frac{A_{21}}{A_{22}}\right)+a_{22}+a_{23}\left(\frac{A_{32}}{A_{22}}\right)\right] = \frac{1}{c_2}+\frac{A_{21}}{A_{22}}\left(\frac{\partial \ln \gamma_2}{\partial c_1}\right)_{T,V,c_{k\neq 1}}+\left(\frac{\partial \ln \gamma_2}{\partial c_2}\right)_{T,V,c_{k\neq 2}} + \frac{A_{32}}{A_{22}}\left(\frac{\partial \ln \gamma_2}{\partial c_3}\right)_{T,V,c_{k\neq 3}}. \quad (21)$$

Equation (21) can be formally related (see Appendix A) to G_{22}, or to the osmotic second virial coefficient (B_{22}) when protein (component 2) is sufficiently dilute.

Combining Eqs. (19)–(21) and using KB theory (see also below) gives

$$\frac{R_{90}^{ex}}{K'\eta_2^2}=\frac{\left[1+\left(\frac{\eta_1}{\eta_2}\right)\left(\frac{c_1G'_{12}}{1+c_2G'_{22}}\right)+\left(\frac{\eta_3}{\eta_2}\right)\left(\frac{c_3G'_{23}}{1+c_2G'_{22}}\right)\right]^2}{\frac{1}{c_2}-\frac{G'_{22}}{1+c_2G'_{22}}}, \quad (22)$$

where the ratio A_{2i}/A_{22} with $i=1,3$ was expressed with the aid of Eqs. (5) and (13) in terms of KB integrals as

$$\frac{A_{2i}}{A_{22}}=\frac{A_{2i}/|A|}{A_{22}/|A|}=\frac{(\partial c_i/\partial \mu_2)_{T,V,\mu_{j\neq 2}}}{(\partial c_2/\partial \mu_2)_{T,V,\mu_{j\neq 2}}} = \frac{c_iG'_{2i}}{1+c_2G'_{22}}. \quad (23)$$

Changing to w/v units, multiplying the numerator and denominator by $c_2^m\left(1+c_2^mG_{22}^m\right)$, and rearranging Eq. (22) gives

$$\frac{R_{90}^{ex}}{K'\left(\eta_2^m\right)^2}=M_2\left(1+2k_{13}\right)c_2^m+M_2G_{22}^m\left(c_2^m\right)^2 + \frac{M_2c_2^mk_{13}^2}{1+c_2^mG_{22}^m}, \quad (24)$$

where

$$k_{13}=\left(\frac{\eta_1^m}{\eta_2^m}\right)c_1^mG_{12}^m+\left(\frac{\eta_3^m}{\eta_2^m}\right)c_3^mG_{23}^m. \quad (25)$$

If k_{13} is sufficiently small, Eq. (11) is recovered from Eq. (24). Alternatively, multiplying numerator and denominator by c_2^m in Eq. (22), and converting to w/v units, gives

$$\frac{R_{90}^{ex}}{K'\left(\eta_2^m\right)^2}=\frac{c_2^mM_2\left[1+\frac{k_{13}}{1+c_2^mG_{22}^m}\right]^2}{1+\frac{c_2^mG_{22}^m}{1+c_2^mG_{22}^m}}, \quad (26a)$$

$$\frac{R_{90}^{ex}}{K'\left(\eta_2^m\right)^2}=\frac{c_2^mM'_{2,app}}{1+\frac{c_2^mG_{22}^m}{1+c_2^mG_{22}^m}}, \quad (26b)$$

with a somewhat different apparent molecular weight

$$M'_{2,app}=M_2\left(1+\frac{k_{13}}{1+c_2^mG_{22}^m}\right)^2.$$

Equation (26) has the same functional form as that of the now standard expression for analysis of static light scattering,[28,54,57]

$$\frac{R_{90}^{ex}}{K'}\propto\frac{M_2c_2^m}{1+2A_2^mc_2^m} \quad (27)$$

if one defines

$$A_2^m=-\frac{1}{2}\left(\frac{G_{22}^m}{1+c_2^mG_{22}^m}\right). \quad (28)$$

Historically, A_2 in Eq. (27) is treated as being identical to B_{22}.[18,42,59] Notably, $A_2 > B_{22}$ if one considers highly repulsive conditions ($G_{22} \ll 0$), and vice versa for highly attractive conditions ($G_{22} \gg 0$). Comparison of Eq. (27) with Eq. (26) also shows that the former erroneously replaces $M_{2,app}$ with the molecular weight, M_2. While it is true that Eq. (26) and (27) show that A_2 from LS is more complex than simply B_{22}, this is purely a consequence of not working at sufficiently low c_2, and is not due to convolution with other osmotic virial coefficients or KB integrals.[33] The same cannot be stated for $M_{2,app}$, as this clearly depends on all G_{2j}, independent of whether one uses Eq. (11) or ((26)) to show that $M_{2,app} \neq M_2$.

A similar issue exists for other treatments[42,45] that neglect fluctuations in N_1, as well as additional consequences for those that incorrectly equate B_{22} with a derivative of μ_2 or μ_2^{ex} with respect to c_2 without fixing $\mu_{k\neq 2}$.[24,32,33,44,50,51] When fluctuations in N_1 are neglected, one inherently changes the coupling of fluctuations in solvent with respect to fluctuations in all the other components. For example, if a protein molecule leaves the scattering volume but the system has a fixed number of water molecules, there would either be a large void or only co-solute molecules would be able to enter the system to fill that void; vice versa, adding a protein molecule to the system would result in preferentially "crowding" out co-solute molecules rather than water molecules if this constant-N_1 assumption holds. The supplementary material[53] provides additional, mathematical analysis of the consequences of assuming constant N_1, as well as if fluctuations in other solvent components or solutes are simultaneously neglected, akin to common treatments.[28,42] In addition, Appendix B shows that the same mathematical form as Eq. (27) can be obtained if one applies the classical approximations akin to that used by Stockmayer[38] (i.e., neglecting not only fluctuations in solvent concentration, but also in pressure), although B_{22} or G_{22} is not explicitly recovered.

IV. EXPERIMENTAL VIRIAL COEFFICIENTS AND APPARENT MOLECULAR WEIGHTS

The preceding sections clearly show that B_{22} and G_{22} arise naturally, and without convolution by protein–solvent or protein–cosolute virial coefficients, if one properly accounts for fluctuations in all species simultaneously. Nevertheless, there are differences in the *de novo* mathematical forms presented herein if one considers the KB analysis (i.e., an open ensemble) (Eq. (11)) or upon imposing a restriction of non-dissociable or unionized solutes (Eq. (24) or (26)). Fundamentally, one might then expect differences in the results

for experimental systems when comparing G_{22} or B_{22} values regressed from LS data using one form versus the other. Equation (11) would be expected to hold more generally than Eq. (24) or (26) for protein solutions, as proteins are necessarily ionized species in most practical instances, even if one can eliminate buffer salts from the solution via extensive dialysis.

In addition, because Eq. (27) is the traditional form used to regress LS data to obtain A_2 values, it remains an open question whether A_2 and B_{22} or G_{22} will differ significantly from a quantitative and qualitative perspective. The results below consider this question by comparing B_{22} and A_2, as well as $M_{2,app}$ values, obtained by regression of experimental SLS data as a function of c_2^m for monoclonal antibodies (mAb) and α-chymotrypsinogen A (aCgn). The experimental details and the scattering data were published previously,[24,31] along with fitted A_2 values using the equivalent of Eq. (27). Those results are combined here with re-analysis of the same data using Eqs. (11) and (26). Before doing so, however, it is useful to examine the implications of some of the standard assumptions when LS experiments are conducted.

A. Experimental implications

Two aspects of LS experiments that are related to the standard assumptions of classical LS analysis are revisited in this subsection. This first is regarding the assumption that the derivative of n with respect to protein concentration at constant V is equal to the same derivative at constant p. Equation (10) provides a rigorous relationship between $\left(\partial n/\partial c_i^m\right)_{T,V,c_{k\neq i}^m}$ $(=\eta_i^m)$ and the more experimentally convenient $(\partial n/\partial c_i^m)_{T,p,c_{k\neq i}^m}$. For water, the term $(\partial n/\partial p)_{T,\mathbf{N}}/\kappa_T$ is of order of 0.2. For proteins, typical values of $(\partial n/\partial c_2)_{T,p,c_{k\neq 2}}$ are of the order of 0.2 mL/g, while typical values of partial specific volume are $\approx$0.75 mL/g. Thus, both terms on the right hand side of Eq. (10) are expected to be quantitatively significant, and assuming that $(\partial n/\partial c_2^m)_{T,p,c_{k\neq 2}^m} \approx \eta_2^m$ is likely a poor assumption when dealing with protein solutions.

Table I summarizes the values for solution conditions used here for aCgn and the four antibodies at different pH and cosolvent concentration. For illustrative proposes, the partial specific volume of protein $(\hat{v}_2)$ was assumed constant and equal to 0.75 mL/g for all cares. In the case of the other thermodynamic quantities involved in Eq. (10), κ_T and $(\partial n/\partial p)_{T,\mathbf{N}}$ were taken from Ref. 60 for aqueous solution of NaCl. The results in Table I show that typical values of η_2^m are approximately twice the value of $(\partial n/\partial c_2^m)_{T,p,c_{j\neq 2}^m}$ for all of the proteins and conditions, demonstrating a potentially large source of quantitative error in values of $M_{2,app}$ and B_{22} or G_{22} fitted with the assumption of $\eta_2 \approx (\partial n/\partial c_2)_{T,p,c_{k\neq 2}}$.

The second aspect which needs consideration is that LS experiments require the Rayleigh ratio of a reference (pure) liquid (R_{90}^{Ref}, e.g., toluene or benzene) in order to measure absolute values of R_{90}^{ex}.[61] The relation between the excess Rayleigh ratio at 90° for a protein solution and that for the reference liquid is given by[28]

$$R_{90}^{ex} = \frac{(I_{90} - I_{90}\,|_{c_2=0})}{I_{90}^{Ref}} \left(\frac{n}{n_{ref}}\right)^2 R_{90}^{Ref}, \tag{29}$$

TABLE I. Comparison of $(\partial n/\partial c_2)$ under the assumption of constant volume or constant pressure.

Protein	pH	NaCl [mM]	$\left(\partial n/\partial c_2^m\right)_{T,p,c_{k\neq 2}^m}$ $[mL/g]$	η_2^m $[mL/g]$
aCgn	2.5 - 4.5	0	0.192	0.407
		100		0.410
		200		0.413
IgG1.1	6.5	54	0.181	0.398
	5.5		0.181	0.398
	4.5		0.183	0.400
	3.5		0.182	0.399
IgG1.2	6.5	54	0.183	0.400
	5.5		0.184	0.401
	4.5		0.184	0.401
	3.5		0.184	0.401
IgG1.3	6.5	54	0.182	0.399
	5.5		0.187	0.404
	4.5		0.183	0.400
	3.5		0.183	0.400
IgG1.4	6.5	54	0.180	0.397
	5.5		0.181	0.398
	4.5		0.181	0.398
	3.5		0.182	0.399

where I_{90} and $I_{90}\,|_{c_2=0}$ are, respectively, the scattered intensities of the protein solution and the solution without protein. I_{90}^{Ref} is the scattered intensity of the reference pure liquid, and n and n_{ref} are the refractive indexes of the solution and the reference liquid, respectively.

In the literature, one can find R_{90}^{Ref} measured from several methods, though they are commonly classified as "high" or "low" values.[62] "High" R_{90} values correspond to those values (methods) which return the average molecular weight of a standard polymer when they are used in Eq. (29) for a dilute solution of the same polymer,[63–68] whereas "low" values do not correlate well with the expected M_2 values for standard–polymer solutions. Apparently, because it is now accepted that LS can be used to obtain the true molecular weight, "low" R_{90} values are often omitted or neglected from later literature, and the use of "high" values has become standard. Thus, if the differences between $(\partial n/\partial c)_T$ at fixed p versus V for the reference macromolecule-solvent system are similar to those for the protein-solvent system of interest, then the errors in approximating $\eta_2 \approx (\partial n/\partial c_2)_{T,p,c_{k\neq 2}}$ may be smaller than the statistical uncertainty in the fitted parameters.

This may explain the historical observations that utilizing $(\partial n/\partial c_2)_{T,p,c_{k\neq 2}}$ instead of η_2 yields at least physically reasonable magnitudes for $M_{2,app}$, although this can differ significantly from the known value of M_2 in either a positive or negative direction.[24,31,69] It remains an open question of how valid this approximation is if one finds significant deviations of $M_{2,app}$ from M_2, particularly if there is no evidence of oligomerization. Therefore, for parity with current practice, in Subsection IV B the reported experimental R_{90}^{ex} values utilize the currently accepted values of R_{90}^{Ref}, and therefore also employ $\eta_2 \approx (\partial n/\partial c_2)_{T,p,c_{k\neq 2}}$. While doing so does not affect the comparison or conclusions below, it behooves the LS community to more carefully consider the historically

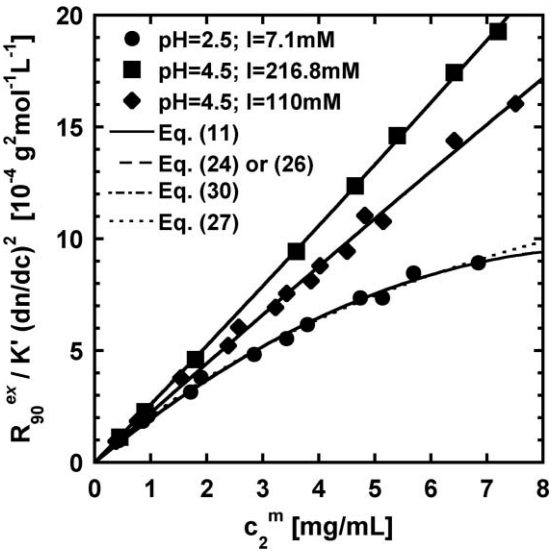

FIG. 1. Representative LS data and fits for aCgn solutions at pH of 2.5 and 4.5, and different ionic strength. The symbols, solid lines, dashed lines, and dotted lines correspond to experimental data,[31] and mathematical fits to Eqs. (11), (26), (27), and (30) (using constant A_2^m), respectively. The fits to Eqs. (11), (24), (26), or (30) are indistinguishable on the scale of the plot.

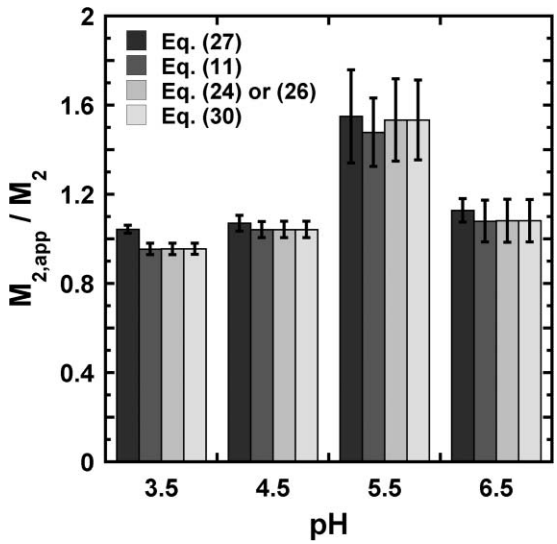

FIG. 2. Apparent protein molecular weight $M_{2,app}$ for IgG1.3 in 54mM NaCl as a function of pH. The values are obtained from regressing experimental data to the working equations. For Eq. (26), G_{22}^m and k_{13} were regressed, and $M_{2,app}$ calculated from the expression for $M'_{2,app}$ defined in Eq. (26). Error bars are based on 95% confidence intervals for the fitted parameters. Analogous results are shown for IgG1.1, IgG1.2, and IgG1.4, as well as for aCgn, in the supplementary material.[53]

accepted assumptions of constant pressure in LS analysis, and in this case its impact on the Rayleigh ratio of the reference liquid, and what is the correct value for the differential of n with respect to solute and/or protein concentration.

B. Classical analysis vs. KB analysis

Using KB theory and a more general canonical treatment, two *de novo* expressions (Eq. (11) and each of the equivalent Eq. (24) or (26)) have been derived to describe LS data. In addition, if one neglects the c_2 dependence of $M'_{2,app}$ of Eq. (26b), this yields a third expression

$$\frac{R_{90}^{ex}}{K'\left(\eta_2^m\right)^2} = \frac{c_2^m M_{2,app}}{1+\frac{c_2^m G_{22}^m}{1+c_2^m G_{22}^m}} \tag{30}$$

that is reminiscent of the canonical expression for LS (Eq. (27), with A_2^m held fixed) except that the denominator does not assume $A_2^m = -2G_{22}^m$. In Eq. (30), $M_{2,app} = (1+k_{13})^2 \approx M_2$, and therefore this expression is expected to hold only when protein-solvent and protein-cosolute interactions are relatively weak. In what follows, the results from using each of the three expressions (Eq. (11), (26), or (30)) with experimental data for protein solutions are compared with that from a canonical treatment. In the remainder of the report, Eq. (27) is used with A_2^m as a constant, making it equivalent to the canonical expression that historically equates B_{22} with A_2^m, factoring in the appropriate conversion between different conventions for the units of B_{22}.[18,42,59] This provides a means to test not only whether the *de novo* expressions can provide reasonable fits to the experimental data, but also to assess differences in the resulting values of G_{22} and $M_{2,app}$ compared to A_2 and M_2 in the classical LS analysis.

Figure 1 shows a comparison of illustrative fits for each *de novo* expression and the classical equation to representative LS data for aCgn at solvent conditions that span from attractive (upward curvature) to repulsive (downward curvature). For all the conditions, all the equations capture the qualitative behavior of R_{90}^{ex} as a function of protein concentration, as expected since each of Eqs. (11), (26), (30), and (27) produce the same sign for the curvature of R_{90}^{ex} if one use a positive or negative value for G_{22} or A_2. When protein–protein interactions are near ideal (i.e., linear R_{90}^{ex} vs. c_2^m), no difference can be observed between these expressions. However, for large deviations, differences between the *de novo* equations (Eqs. (11)), (26), and (30)), and the classical expression (Eq. (27) with constant A_2^m) become more evident. However, these differences are minor, and likely would not give one cause to conclude that one expression is inherently better able to capture the LS data per se. Rather, the comparison of the fitted parameter values reveals the differences more clearly.

Figure 2, as well as Figs. S1 and S2 in the supplementary material,[53] summarize differences in $M_{2,app}$ values obtained from regressing SLS data as a function of pH and I, using the classical analysis (Eq. (27) or (B6)) and the new expressions (Eqs. (11), (26), and (30)), for aCgn, and each antibody. These values are normalized by the true protein molecular weight M_2 derived from amino acid analysis. The true molecular weights for the five proteins are: 146.7, 146.9, 146.5, 143.0, and 25.7 kDa for IgG1.1, IgG1.2, IgG1.3, IgG1.4, and aCgn, respectively. For the four antibodies, LS data correspond to solutions at one cosolvent concentration ([NaCl] = 54 mM) and four different pH (3.5, 4.5, 5.5, and 6.5), whereas for aCgn the data correspond to solutions at three different cosolvent concentrations ([NaCl] = 0, 100, and 200 mM) and five different pH values (2.5, 3, 3.5, 4, and 4.5).

Inspection of Fig. 2, and S1 and S2 in the supplementary material[53] show that the fitted $M_{2,app}$ values are the same from Eqs. (11), (26), (27), and (30) within 95% confidence intervals, across a wide range of conditions.

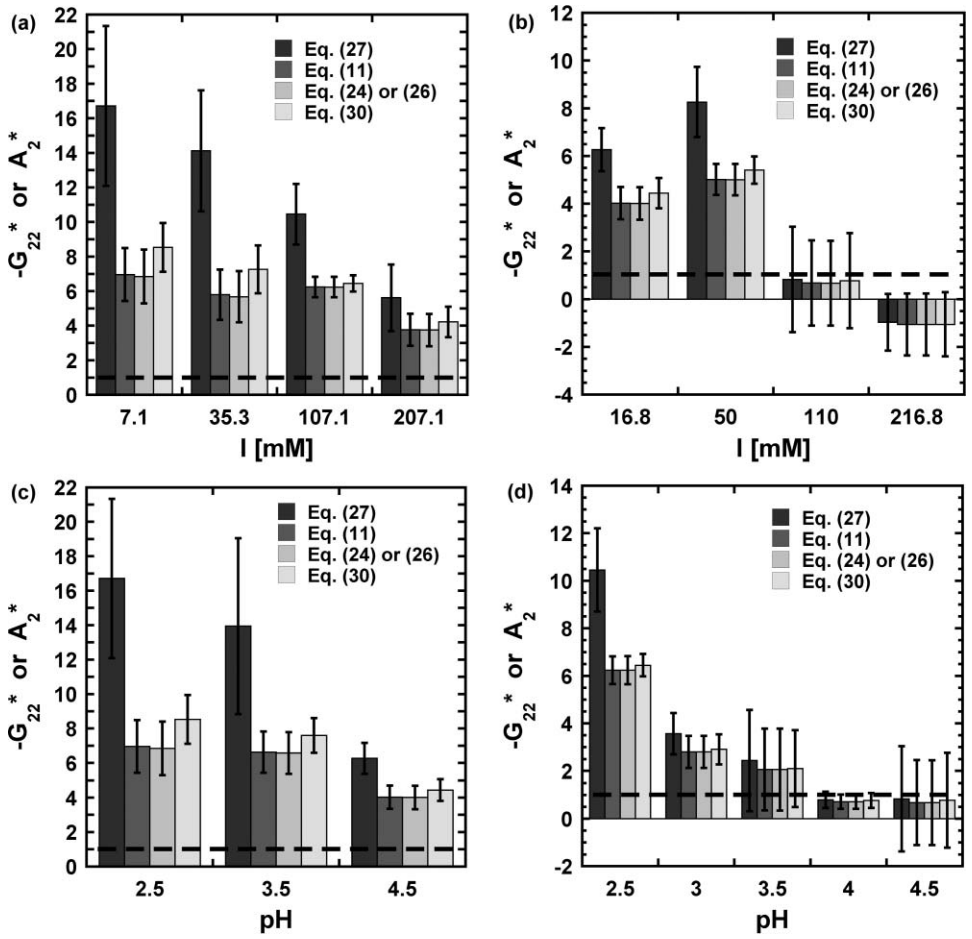

FIG. 3. B_{22} values for aCgn obtained from classical (Eq. (27)) and KB (Eqs. (11), (24), (30) or (26)) analysis at different pH and ionic strength (I): (a) pH = 2.5; (b) pH = 4.5; (c) low ionic strength ([NaCl] = 0 mM); (d) medium ionic strength ([NaCl] = 100 mM). Dashed line indicates value for ideal hard spheres. Error bars are based on 95% confidence intervals for the fitted parameters.

Additionally, $M_{22,app}/M_2$ is not greatly different from 1 for most examples tested here. This suggests that $|k_{13}| \ll 1$, as required for Eq. (24) or (26) to be equivalent to the more general expression (Eq. (11)). In the case of IgG1.2 (Fig. S1b) and pH 5.5 in Fig. 2, protein dimerization was suspected,[24] consistent with $M_{2,app} \gg M_2$. A more systematic and broader range of cosolute compositions may need to be tested to assess how large the deviations of $M_{2,app}$ from M_2 may become for real systems.

In terms of protein–protein interactions, Figs. 3 and 4, and S3 (see supplementary material[53]) summarize the values of G_{22} or A_2 obtained from regressing the experimental LS data for aCgn and each of mAb. These values are reported relative to the hard-sphere second virial coefficient ($-G^*_{22} = -G_{22}/2B_2^{HS}$ and $A_2^* = A_2/B_2^{HS}$) for easier comparison across different proteins; recall that $G_{22} = -2B_{22}$ at low c_2.[9] The estimated B_2^{HS} was calculated as $B_2^{HS} = (2/3)\pi\sigma_{HS}^3$, where the hard-sphere diameter (σ_{HS}) was taken as 10 nm for each of the four mAb[70–72] and 4 nm for aCgn.[31]

At low pH (≤ 3.5), repulsive protein interactions are expected ($B_{22} > B_2^{HS}$, $G^*_{22} \ll 0$), since all the charged side chains have positive charges and the contribution from electrostatic interactions to the protein–protein interactions is significant at these relatively low ionic strength values. As pH increases, there are both repulsive and attractive protein–protein interactions because there are both positively and negatively charged side chains, and G_{22} (A_2) is expected to increase (decrease). Similarly, at low ionic strength, elec-

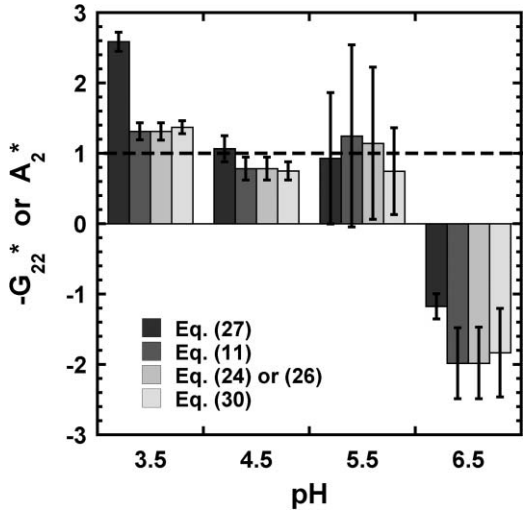

FIG. 4. B_{22} values for IgG1.3 in 54 mM NaCl as a function of pH. The values are obtained from regressing experimental data to the working equations. Dashed line indicates value for ideal hard spheres. Error bars are based on 95% confidence intervals for the fitted parameters.

trostatic interactions are relatively unscreened, and become more screened with increasing I.[73] If one is concerned about only the sign of B_{22}, A_2, or G_{22}, then Eqs. (11), (24) or (26), (30), and (27) provide equivalent results. That is, linear behavior for R_{90}^{ex} gives $A_2 = B_{22} = G_{22} = 0$, and positive (negative) deviations from linearity requires A_2, B_{22}, or $-G_{22}$ to be negative (positive).

On the other hand, if one is concerned with the magnitude of protein–protein interactions, then Figs. 3 and 4, and S3 in the supplementary material[53] show that the classical LS treatment is in error. As Eq. (28) shows, the errors are most pronounced at high c_2 and/or $|G_{22}|$. The now-standard expression to analyze LS (Eq. (27)) implicitly assumes $|c_2 G_{22}| \ll 1$ by considering $A_2 = B_{22}$. However, within the experimentally accessible range of protein concentration, that product is not necessarily small. Thus, for very repulsive (attractive) conditions, $G_{22} \ll 0$ ($\gg 0$), the classical expression is no longer valid. This causes $-G_{22}$ or B_{22} to be over (under) estimated under strongly repulsive (attractive) conditions. The main difference in deriving Eq. (11) or (26) is the assumption of non-dissociable solutes for Eq. (26), i.e., via the neglect of Donnan contributions in Eq. (26). The agreement between G_{22}^* for fits to Eq. (11) vs. Eq. (26) or (30) in Figs. 3 and 4, and S2 (supplementary material[53]) suggests that contributions from Donnan equilibria involving the proteins and the counterions are not significant for the present examples, within experimental uncertainty. As shown above, A_2 from traditional analysis is not equal to B_{22} unless the magnitude of $c_2 B_{22}$ (or $c_2 G_{22}$) is small compared to 1. However, this cannot explain the difference in sign argued elsewhere.[32,33] The source of the discrepancies between B_{22} obtained from different experimental methods[33,74] is not apparent at this point, but the analysis of LS provided here clearly shows that B_{22} or G_{22} is rigorously obtained from proper analysis of LS data.

Furthermore, the fact that differences between A_2 and B_{22} (or G_{22}) in Eq. (26) and (27) are the artifact of a mathematical manipulation and the assumption of $c_2 \to 0$ suggests A_2 may be corrected without a need to regress data retrospectively. Figure 5 shows A_2^* values regressed from Eq. (27) vs. $-G_{22}^*$ values regressed from Eq. (11) for all the data analyzed here, along with a curve that represents Eq. (28) – i.e., the curve is not a fit to the data. The average protein concentration among all the LS data was used in Eq. (28) for Fig. 5. Together with the analysis provided in Sec. III, this clearly demonstrates that A_2 regressed from Eq. (27) is not the osmotic second virial coefficient unless c_2 and/or $|G_{22}|$ are sufficiently small, but one can relate A_2 from classical analysis to G_{22} via Eq. (28). However, the difference between A_2 and G_{22} is not due to a convolution with other osmotic virial coefficients; it occurs because of the erroneous approximation of $c_2 \to 0$ when analyzing LS data. For the data considered here, differences between A_2 and $-G_{22}/2$ are significant when $|G_{22}^m c_2^m| \geq 3\text{x}10^{-2}$, even if the average concentration is low (≈ 3 mg/mL). Nevertheless, Eq. (28) may provide a way to recover accurate values of G_{22} from previously determined A_2 values, and therefore also test whether the approximation $A_2 \approx B_{22}$ is valid.

Interestingly, Asthagiri *et al.*[44] also found protein–protein interactions were over-estimated at repulsive conditions by comparing A_2 obtained from experimental LS data with those values obtained from molecular simulations. By acknowledging ionizable species in a derivation starting from Stockmayer approach,[38] they suggested that the A_2 parameter is the combination of a protein–protein interaction term and a Donnan effect term arising from a need to impose electroneutrality. Thus, for conditions dominated by electrostatic interactions (very high or low pH, and/or low ionic strength), the Donnan term becomes important, leading to overestimated positive values of A_2. The analysis here provides an alternative explanation, without a need to invoke argument of electroneutrality, as the grand canonical ensemble inherently maintains electroneutrality by the imposition of constant chemical potential. The empirical observation that G_{22}^* fit from Eqs. (11) and (26) were equivalent in the present work argues that Donnan contributions to Eq. (26) could be neglected for the experimental data considered here, as Eq. (11) is not limited by such restrictions.

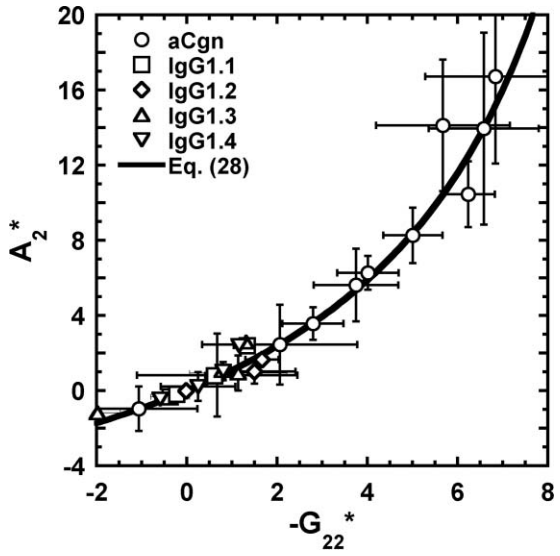

FIG. 5. Normalized osmotic virial coefficients obtained from Eq. (27) (A_2^*) and Eq. (11) ($-G_{22}^*$) for aCgn and each of the mAb's. Symbols corresponds to (○) aCgn; (□) IgG1.1; (◇) IgG1.2; (△) IgG1.3; (▽) IgG1.4; solid line corresponds to Eq. (28).

Strictly, if one considers the scenario of dissociable or ionizable species, particularly protein, the canonical treatment (Eq. (26)) is no longer valid, since Eq. (21) (or Eq. (A4)) only applies to non-dissociable components. However, Eq. (11) can be used whether or not protein and/or cosolutes dissociate or ionize if one acknowledges the concentration of some dissociated or ionized species is proportional to c_2. Nevertheless, special care must be taken in using the KB analysis for such situations, as G_{22} may be a linear combination of the KB integrals for the interactions between protein and some of the dissociated components (e.g., counterions). This is a common issue in other types of experiments for measuring osmotic virial coefficients or KB integrals such as sedimentation equilibrium and classical osmometry experiments, since there is thermodynamic coupling between protein and some of the dissociated or ionized components in solution. This does not preclude the possibility that one would obtain significantly different results at very low ionic strength and/or high

protein charge if one used Eq. (26) instead of Eq. (11), as well as the importance of considering Donnan contributions when large strong electrostatic contributions are present. Correcting for additional factors such as Donnan equilibrium in deriving Eq. (26) or (30) is left to a future report, and instead Eq. (11) is recommended for use more generally.

V. SUMMARY AND CONCLUSIONS

Classic analysis of Rayleigh light scattering in terms of concentration fluctuations is revisited and three alternative versions of the final working equation are derived and compared with the traditional form. The new formulations are based on either KB solution theory, or on a more general canonical treatment applicable to solutions with non-dissociable solutes. The former is more general in that it is not restricted in terms of the nature of the solutes or solvents, or assumptions of independent cross fluctuations of charged species. The analysis shows that B_{22} arises naturally and without convolution by other virial coefficients if one considers sufficiently dilute protein concentrations and/or low $|B_{22}|$ values – disputing recent arguments that are based on a less general analysis. However, comparison to the now-standard expression for regressing LS data shows that the standard analysis may significantly over(under)-estimate the magnitude of the osmotic second virial coefficient for strongly repulsive(attractive) protein-protein interactions. The KB integral G_{22} is the more relevant quantity as one considers larger protein concentrations and/or strong repulsions or attractions. Expressions are derived that in principle allow one to correct previous results, as well as to unambiguously determine whether one is sufficiently dilute to recover B_{22} rather than G_{22} from LS experiments. The analysis is applied to a number of experimental systems, illustrating that the magnitude of the errors from the traditional analysis can be significant (a factor of 2 or more) for estimating protein-protein virial coefficient or KB integral values; while the errors in assuming the apparent molecular weight from LS analysis is equal to the true molecular weight may be appreciably smaller but still measurable. The present work also lays a foundation for extending LS to model protein-protein and protein-solvent or protein-cosolute interactions in more concentrated solutions.

ACKNOWLEDGMENTS

P. Butler, V. Shen, D. Siderius, and W. Krekelberg are thanked for many helpful and stimulating discussions, and the National Institute of Standards and Technology is gratefully acknowledged for hosting C.J.R. during his sabbatical leave. D. Asthagiri is thanked for helpful discussions and comments on the manuscript, and the National Institutes of Health (Grant No. R01EB006006) and National Science Foundation (Grant No. CBET0931173) are gratefully acknowledged for financial support.

APPENDIX A: B_{22} IN TERMS OF ACTIVITY COEFFICIENTS

In order to establish the relation between the osmotic second virial coefficient, B_{22}, and the activity coefficients, we can start from the definition of fluctuations in the concentration of protein for a multi component mixture in a grand-canonical ensemble. That is

$$\langle(\Delta N_2)^2\rangle = kT\left(\frac{\partial N_2}{\partial \mu_2}\right)_{T,V,\mu_{k\neq 2}} = kT\frac{A_{22}}{|\ A\ |}, \tag{A1}$$

where A_{22} and $|\ A\ |$ are defined in Eq. (16). In addition, these fluctuations can be expressed rigorously in terms of the *KB* integral, G_{22}, as[9]

$$\frac{kT}{V}\left(\frac{\partial N_2}{\partial \mu_2}\right)_{T,V,\mu_{k\neq 2}} = c_2\left(1 + c_2 G_{22}\right) = kT\left(\frac{\partial c_2}{\partial \mu_2}\right)_{T,V,\mu_{k\neq 2}}. \tag{A2}$$

Combining Eqs. (A1) and (A2), and defining the protein activity coefficient (γ_2) as Eq. (20), we can express G_{22} as

$$G_{22} = \frac{-\left[\left(\frac{\partial \ln \gamma_2}{\partial c_2}\right)_{T,V,c_{k\neq 2}} + \sum_{i\neq 2}\frac{A_{2i}}{A_{22}}\left(\frac{\partial \ln \gamma_2}{\partial c_i}\right)_{T,V,c_{k\neq i}}\right]}{1 + c_2\left[\left(\frac{\partial \ln \gamma_2}{\partial c_2}\right)_{T,V,c_{k\neq 2}} + \sum_{i\neq 2}\frac{A_{2i}}{A_{22}}\left(\frac{\partial \ln \gamma_2}{\partial c_i}\right)_{T,V,c_{k\neq i}}\right]}. \tag{A3}$$

Equation (A3) is formally exact assuming non-dissociable species. In the limit of infinite dilution of protein ($c_2 \to 0$) this becomes

$$G_{22} = -\left[\left(\frac{\partial \ln \gamma_2}{\partial c_2}\right)^{(\infty)}_{T,V,c_{k\neq 2}} + \sum_{i\neq 2}\frac{A_{2i}}{A_{22}}\left(\frac{\partial \ln \gamma_2}{\partial c_i}\right)^{(\infty)}_{T,V,c_{k\neq i}}\right], \tag{A4}$$

where the superscript (∞) denotes that the derivatives are evaluated in the limit of low protein concentration. In addition, $G_{22}^{(\infty)} = -2B_{22}$,[9] therefore

$$B_{22} = \frac{1}{2}\left[\left(\frac{\partial \ln \gamma_2}{\partial c_2}\right)^{(\infty)}_{T,V,c_{k\neq 2}} + \sum_{i\neq 2}\frac{A_{2i}}{A_{22}}\left(\frac{\partial \ln \gamma_2}{\partial c_i}\right)^{(\infty)}_{T,V,c_{k\neq i}}\right]. \tag{A5}$$

Equations (A3)–(A5) give formal relations between protein–protein interactions in terms of G_{22} and B_{22} (i.e., an open ensemble), and the activity coefficients of the different component in the solution (i.e., a Helmholtz framework).

APPENDIX B: CONSTANT PRESSURE ASSUMPTION

Most of the theories[28,38,39,42,44] which relate Rayleigh scattering to molecular interactions assume that the system is at constant pressure. In addition, fluctuations on N_1 (water), or in all solvent species, are often neglected. The impact of these approximations are illustrated here, for comparison to the results in Sections II and III. For concreteness, first consider a ternary mixture at constant T, p, and with μ_k of all the components but solvent (component 1) held fixed, i.e., constant $\left(p, T, N_1, \mu_{k\neq 1}\right)$. In this ensemble, fluctuations in refractive

index are due to fluctuations in V and $N_{k\neq 1}$

$$\langle(\Delta n)^2\rangle = \left(\frac{\partial n}{\partial V}\right)^2_{T,\mathbf{N}} \langle(\Delta V)^2\rangle + 2\sum_{i=2}^{3}\left(\frac{\partial n}{\partial V}\right)_{T,\mathbf{N}}\left(\frac{\partial n}{\partial N_i}\right)_{T,V,N_{k\neq i}} \langle\Delta V\Delta N_i\rangle + \sum_{i=2}^{3}\sum_{j=2}^{3}\left(\frac{\partial n}{\partial N_i}\right)_{T,V,N_{k\neq i}}\left(\frac{\partial n}{\partial N_j}\right)_{T,V,N_{k\neq j}} \langle\Delta N_i\Delta N_j\rangle, \tag{B1}$$

where fluctuations in volume and number of molecules are expressed as

$$\begin{aligned}
\beta\langle(\Delta V)^2\rangle &= -\left(\frac{\partial V}{\partial p}\right)_{T,N_1,\mu_{k\neq 1}} \\
&= \kappa_T V - \sum_{i=2}^{3}\overline{v}_i\left(\frac{\partial N_i}{\partial p}\right)_{T,N_1,\mu_{k\neq 1}} \\
\beta\langle\Delta V\Delta N_i\rangle &= \left(\frac{\partial V}{\partial \mu_i}\right)_{T,P,N_1,\mu_{k\neq 1,i}} \\
&= -\left(\frac{\partial N_i}{\partial p}\right)_{T,N_1,\mu_{k\neq 1,i}} \\
&= \sum_{j=2}^{3}\overline{v}_j\left(\frac{\partial N_j}{\partial \mu_i}\right)_{T,p,N_1,\mu_{k\neq 1,i}} \\
\beta\langle\Delta N_i\Delta N_j\rangle &= \left(\frac{\partial N_j}{\partial \mu_i}\right)_{T,p,N_1,\mu_{k\neq 1,i}} \\
&= \left(\frac{\partial N_i}{\partial \mu_j}\right)_{T,p,N_1,\mu_{k\neq 1,j}}.
\end{aligned}$$

Replacing the above definitions in Eq. (B2), and substituting derivatives at constant V for their equivalents derivatives at constant p, we obtain

$$\langle V\rangle\langle(\Delta n)^2\rangle = kT\langle V\rangle\sum_{i=2}^{3}\sum_{j=2}^{3}\xi_i\xi_j\left(\frac{\partial N_i}{\partial \mu_j}\right)_{T,p,N_1,\mu_{k\neq j}} - \frac{kT}{\kappa_T}\left(\frac{\partial n}{\partial p}\right)^2_{T,\mathbf{N}}, \tag{B2}$$

where

$$\xi_i = \left(\frac{\partial n}{\partial N_i}\right)_{T,p,N_{k\neq i}}.$$

In a similar way to that for obtaining Eq. (16), the derivative of the number of molecules of the i'th component with respect the chemical potential of the j'th component can be written as

$$\left(\frac{\partial N_i}{\partial \mu_j}\right)_{T,p,N_1,\mu_{k\neq j}} = \frac{\Psi_{ij}}{|\Psi|}, \tag{B3}$$

where $|\Psi|$ and Ψ_{ij} are the determinant and the ij–cofactor in the determinant, respectively, of the matrix formed by the elements

$$\psi_{ij} = \psi_{ji} = \left(\frac{\partial \mu_i}{\partial N_j}\right)_{T,p,N_{k\neq j}} \qquad i,j = 2,3.$$

With Eqs. (B2) and (B3), Eq. (2) in mole/volume units becomes

$$\frac{R_{90}}{K'} = \frac{RT\left[\xi_2 - \xi_3\left(\frac{\psi_{23}}{\psi_{33}}\right)\right]^2}{\langle V\rangle\left[\psi_{22} - \psi_{23}\left(\frac{\psi_{23}}{\psi_{33}}\right)\right]} + \frac{RT\xi_3^2}{\langle V\rangle\psi_{33}} - \frac{RT}{\kappa_T}\left(\frac{\partial n}{\partial p}\right)^2_{T,\{\mathbf{n_k}\}}. \tag{B4}$$

One can then express $\left(R_{90}^{ex}\right)$ as

$$\frac{R_{90}^{ex}}{K'\xi_2^2} = \frac{\left[1 - \left(\frac{\xi_3}{\xi_2}\right)\left(\frac{\psi_{23}}{\psi_{33}}\right)\right]^2}{\frac{\langle V\rangle}{RT}\left(\frac{\partial \mu_2}{\partial \mathbf{n_2}}\right)_{T,p,\mathbf{n_1},\mu_3}}, \tag{B5}$$

where the denominator follows from

$$\left(\frac{\partial \mu_2}{\partial \mathbf{n_2}}\right)_{T,p,\mathbf{n_1},\mu_3} = \left(\frac{\partial \mu_2}{\partial \mathbf{n_2}}\right)_{T,p,\mathbf{n_{j\neq 2}}} + \left(\frac{\partial \mu_2}{\partial \mathbf{n_3}}\right)_{T,p,\mathbf{n_{j\neq 3}}}\left(\frac{\partial \mathbf{n_3}}{\partial \mathbf{n_2}}\right)_{T,p,\mathbf{n_1},\mu_3}.$$

Finally, by analogy with the derivation of Eq. (27) in Sec. III, one can express Eq. (B5) at low protein concentration as

$$\frac{R_{90}^{ex}}{K'\left(\xi_2^m\right)^2} = \frac{c_2^m M_{2,app}}{1 + 2c_2^m A_{2,app}^m}, \tag{B6}$$

where

$$M_{2,app} = M_2\left\{1 + \frac{\xi_3}{\xi_2}c_3^m\left[\frac{G_{23}}{M_2} - \frac{M_1}{M_2}\left(\frac{G_{13}}{M_1} + \frac{G_{12}}{M_1} - \frac{1}{c_1^m} - \frac{G_{11}}{M_1}\right)\right]\right\}^2, \tag{B7}$$

$$A_{2,app}^m = -\frac{1}{2}\left[G_{22}^m + \frac{M_1}{M_2}\left(\frac{1}{c_1^m} + \frac{G_{11}}{M_1} - \frac{2G_{12}}{M_1}\right)\right]. \tag{B8}$$

[1] W. H. Stockmayer and H. E. Stanley, J. Chem. Phys. **18**, 153 (1950).
[2] S. N. Timasheff, H. M. Dintzis, J. G. Kirkwood, and B. D. Coleman, J. Am. Chem. Soc. **79**, 782 (1957).
[3] N. L. Abbott, D. Blankschtein, and T. A. Hatton, Macromolecules **25**, 3917 (1992).
[4] Y. U. Moon, R. A. Curtis, C. O. Anderson, H. W. Blanch, and J. M. Prausnitz, J. Solution Chem. **29**, 699 (2000).
[5] R. A. Curtis, H. W. Blanch, and J. M. Prausnitz, J. Phys. Chem. B **105**, 2445 (2001).
[6] J. Narayanan and X. Y. Liu, Biophys. J. **84**, 523 (2003).
[7] W. Liu, D. Bratko, J. M. Prausnitz, and H. W. Blanch, Biophys. Chem. **107**, 289 (2004).
[8] G. Scatchard, J. Am. Chem. Soc. **68**, 2315 (1946).
[9] A. Ben-Naim, *Statistical Thermodynamics for Chemists and Biochemists* (Plenum Press, New York, 1992).
[10] D. A. McQuarrie, *Statistical Mechanics* (University Science Books, Sausalito, CA, 2000).

[11]A. George and W. W. Wilson, Acta Crystallogr., Sect. D: Biol. Crystallogr. **50**, 361 (1994).
[12]D. F. Rosenbaum and C. F. Zukoski, J. Cryst. Growth **169**, 752 (1996).
[13]B. L. Neal, D. Asthagiri, and A. M. Lenhoff, Biophys. J. **75**, 2469 (1998).
[14]B. L. Neal, D. Asthagiri, O. D. Velev, A. M. Lenhoff, and E. W. Kaler, J. Cryst. Growth **196**, 377 (1999).
[15]B. Guo, S. Kao, H. McDonald, A. Asanov, L. L. Combs, and W. W. Wilson, J. Cryst. Growth **196**, 424 (1999).
[16]P. M. Tessier and A. M. Lenhoff, Curr. Opin. Biotechnol. **14**, 512 (2003).
[17]P. M. Tessier, V. J. Verruto, S. I. Sandler, and A. M. Lenhoff, Biotechnol. Bioeng. **87**, 303 (2004).
[18]A. C. Dumetz, A. M. Snellinger-O'Brien, E. W. Kaler, and A. M. Lenhoff, Protein Sci. **16**, 1867 (2007).
[19]H. M. Schaink and J. A. M. Smit, Phys. Chem. Chem. Phys. **2**, 1537 (2000).
[20]J. Zhang and X. Y. Liu, J. Chem. Phys. **119**, 10972 (2003).
[21]J. R. Alford, B. S. Kendrick, J. F. Carpenter, and T. W. Randolph, Anal. Biochem. **377**, 128 (2008).
[22]W. F. WeissIV, T. M. Young, and C. J. Roberts, J. Pharm. Sci. **98**, 1246 (2009).
[23]Y. Li, W. F. Weiss IV, and C. J. Roberts, J. Pharm. Sci. **98**, 3997 (2009).
[24]E. Sahin, A. O. Grillo, M. D. Perkins, and C. J. Roberts, J. Pharm. Sci. **99**, 4830 (2010).
[25]S. Ruppert, S. I. Sandler, and A. M. Lenhoff, Biotechnol. Prog. **17**, 182 (2001).
[26]E. Dickinson, M. G. Semenova, L. E. Belyakova, A. S. Antipova, M. M. Il'in, E. N. Tsapkina, and C. Ritzoulis, J. Colloid Interface Sci. **239**, 87 (2001).
[27]P. J. Loll, M. Allaman, and J. Wiencek, J. Cryst. Growth **232**, 432 (2001).
[28]P. C. Hiemenz and R. Rajagopalan, *Principles of Colloid and Surface Chemistry*, 3rd ed. (Marcel Dekker Inc., New York, 1997).
[29]N. Kern and D. Frenkel, J. Chem. Phys. **118**, 9882 (2003).
[30]A. C. Dumetz, A. M. Chockla, E. W. Kaler, and A. M. Lenhoff, BBA-Proteins Proteomics **1784**, 600 (2008).
[31]Y. Li, B. A. Ogunnaike, and C. J. Roberts, J. Pharm. Sci. **99**, 645 (2010).
[32]M. Deszczynski, S. E. Harding, and D. J. Winzor, Biophys. Chem. **120**, 106 (2006).
[33]D. J. Winzor, M. Deszczynski, S. E. Harding, and P. R. Wills, Biophys. Chem. **128**, 46 (2007).
[34]J. G. S. Ho, A. P. J. Middelberg, P. Ramage, and H. P. Kocher, Protein Sci. **12**, 708 (2003).
[35]W. W. Wilson, J. Struct. Biol. **142**, 56 (2003).
[36]J. J. Valente, R. W. Payne, M. C. Manning, W. W. Wilson, and C. S. Henry, Curr. Pharm. Biotechnol. **6**, 427 (2005).
[37]R. A. Curtis and L. Lue, Chem. Eng. Sci. **61**, 907 (2006).
[38]W. H. Stockmayer, J. Chem. Phys. **18**, 58 (1950).
[39]J. G. Kirkwood and R. J. Goldenberg, J. Chem. Phys. **18**, 54 (1950).
[40]R. K. Bullough, Proc. R. Soc. London, Ser. A **275**, 271 (1963).
[41]P. Munk and M. Tian, Polymer **36**, 1975 (1995).
[42]A. P. Minton, Biophys. J. **93**, 1321 (2007).
[43]C. Fernandez and A. P. Minton, Biophys. J. **96**, 1992 (2009).
[44]D. Asthagiri, A. Paliwal, D. Abras, A. M. Lenhoff, and M. E. Paulaitis, Biophys. J. **88**, 3300 (2005).
[45]Y. Einaga, F. Abe, and H. Yamakawa, J. Phys. Chem. **96**, 3948 (1992).
[46]P. Fleming, A. R. Peacocke, and R. G. Wallis, J. Polym. Sci. **19**, 495 (1956).
[47]E. F. Casassa and H. Eisenberg, J. Phys. Chem. **64**, 753 (1960).
[48]E. F. Casassa and H. Eisenberg, J. Phys. Chem. **65**, 427 (1961).
[49]C. Fernandez and A. P. Minton, Anal. Biochem. **381**, 254 (2008).
[50]J. M. Mollerup and M. P. Breil, Fluid Phase Equilib. **286**, 78 (2009).
[51]J. M. Mollerup and M. P. Breil, Fluid Phase Equilib. **276**, 18 (2009)
[52]J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **19**, 774 (1951).
[53]See supplementary material at http://dx.doi.org/10.1063/1.3596726 for these figures.
[54]C. I. Carr and B. H. Zimm, J. Chem. Phys. **18**, 1616 (1950).
[55]A. Einstein, Ann. Phys.-Berlin **33**, 1275 (1910).
[56]B. H. Zimm, J. Chem. Phys. **16**, 1099 (1948).
[57]B. H. Zimm, J. Chem. Phys. **16**, 1093 (1948).
[58]A. Ben-Naim, A. M. Navarro, and J. M. Leal, Phys. Chem. Chem. Phys. **10**, 2451 (2008).
[59]E. Nordmeier and W. Dauwe, Polym. J. **24**, 229 (1992).
[60]*CRC Handbook of Chemistry and Physics*, 81st ed., edited by D. P. Lide (CRC Press, Boca Raton, Florida, 2000).
[61]H. Wu, Chem. Phys. **367**, 44 (2010).
[62]J. P. Kratohvil, G. Dezelic, E. Matijevic, and M. Kerker, J. Polym. Sci. **57**, 59 (1962).
[63]H. Wahid, J. Opt. **26**, 109 (1995).
[64]E. Moreels, W. Deceuninck, and R. Finsy, J. Chem. Phys. **86**, 618 (1987).
[65]E. R. Pike, W. R. M. Pomeroy, and J. M. Vaughan, J. Chem. Phys. **62**, 3188 (1975).
[66]A. Bello and G. Guzman, Eur. Polym. J. **2**, 79 (1966).
[67]J. Kremer and J. J. Shapiro, J. Opt. Soc. Am. **44**, 500 (1954).
[68]L. M. Kushner, J. Opt. Soc. Am. **44**, 155 (1954).
[69]E. Y. Chi, S. Krishnan, B. S. Kendrick, B. S. Chang, J. F. Carpenter, and T. W. Randolph, Protein Sci. **12**, 903 (2003).
[70]W. K. Hartmann, N. Saptharishi, X. Y. Yang, G. Mitra, and G. Soman, Anal. Biochem. **325**, 227 (2004).
[71]R. K. Brummitt, D. P. Nesta, L. Chang, S. F. Chase, T. M. Laue, and C. J. Roberts, "Non-native aggregation of an igg1 antibody in acidic conditions: 1. unfolding, colloidal interactions, and formation of amyloid-like high molecular weight aggregates," J. Pharm. Sci. (in press).
[72]R. K. Brummitt, D. P. Nesta, L. Chang, A. M. Kroetsch, and C. J. Roberts, "Non-native aggregation of an igg1 antibody in acidic conditions: 2. nucleation-and-growth kinetics with competing growth mechanisms," J. Pharm. Sci. (in press).
[73]O. D. Velev, E. W. Kaler, and A. M. Lenhoff, Biophys. J. **75**, 2682 (1998).
[74]A. Saluja, R. M. Fesinmeyer, S. Hogan, D. N. Brems, and Y. R. Gokarn, Biophys. J. **99**, 2657 (2010).