A Critical Examination of the Decoupling Approximation for Small-angle Scattering from Hard Ellipsoids of Revolution

Daniel G. Greene, Daniel V. Ferraro, Abraham M. Lenhoff and Norman J. Wagner *

Center for Molecular & Engineering Thermodynamics, Department of Biomolecular and Chemical Engineering, University of Delaware, 150 Academy Street, Newark, DE 19716 USA. E-mail: wagnernj@udel.edu

Abstract

The decoupling approximation, proposed by Kotlarchyk & Chen (1983), is a first-order correction to the experimentally determined apparent structure factor that arises due to concentration effects in polydisperse and or non-spherical systems. While the approximation is considered accurate for spheres with low polydispersity ($< 10\%$), the corresponding limitations for non-spherical particles are unknown. We study the validity of this approximation for monodisperse dispersions of hard ellipsoids of revolution with aspect ratios ranging from 0.333 to 3 and provide a guide for its accuracy.

1. Introduction

Small-angle scattering (SAS) is an important tool for measuring the solution structure of colloidal suspensions and, more recently, of highly concentrated globular protein

solutions (Stradner *et al.*, 2004; Heinen *et al.*, 2012; Yearley *et al.*, 2013; Yearley *et al.*, 2014; Liu *et al.*, 2011; Godfrin *et al.*, 2016). Interpretation of the SAS signal from protein solutions is complicated the non-spherical shape of the protein molecule and by the complex, orientation-dependent, interaction potential (Neal *et al.*, 1998; Lomakin *et al.*, 1999) . These complexities are typically addressed by treating the by treating the protein shape as an ellipsoid or other non-spherical shape (e.g., a flexible 'Y' shape for a monoclonal antibody (Godfrin *et al.*, 2016)) and the interaction potential as an effective, isotropic potential (Heinen *et al.*, 2012; Liu *et al.*, 2011; Godfrin *et al.*, 2016). Here we rigorously evaluate standard methodologies for addressing the second issue, the non-spherical shape of the protein molecule, by examining SAS from hard ellipsoid configurations.

While the SAS spectrum for monodisperse spheres can be rigorously separated into independent contributions arising from particle shape and particle-particle interactions, the SAS spectrum for polydisperse spheres and for non-spherical objects cannot be similarly factored (Wagner *et al.*, 1991; Kotlarchyk & Chen, 1983). A variety of approximations have been developed to interpret SAS data from these systems (Pedersen, 1997; Hansen, 2013), but the most prevalent approximation used in the protein community is known as the decoupling approximation (DA) proposed by Kotlarchyk & Chen (1983).

The DA is a rigorous first-order correction to the structure factor that assumes for the case of non-spherical particles that particle orientations and positions are not correlated. Thus, using the nomenclature from Kotlarchyk & Chen (1983), the DA allows the scattering intensity to be written as

$$I(q) = n_p P(q) S'(q), \tag{1}$$

where n_p is the number density, $P(q)$ is the particle form factor and $S'(q)$ is an *apparent*

interparticle structure factor. The DA relates the apparent structure factor, $S'(q)$, to what Kotlarchyk & Chen (1983) refer to as the 'true structure factor', $S(q)$, via,

$$S'(q) = 1 + \beta(q)[S(q) - 1],$$
$$\beta(q) = \frac{|\langle F(q)\rangle|^2}{\left\langle |F(q)|^2 \right\rangle} \tag{2}$$

where $\left\langle |F(q)|^2 \right\rangle \equiv P(q)$. The true structure factor is defined as,

$$S(q) = \left\langle \frac{1}{N} \sum_{j,k=1}^{n} e^{-i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_k)} \right\rangle, \tag{3}$$

where the angled brackets indicate averaging over configurations and orientations.

The DA has been shown to be valid for polydisperse spherical systems with low polydispersity (Pedersen, 1997), but, to our knowledge, the limits of its validity for non-spherical particles have not been established. Here we use Monte Carlo (MC) simulations to quantitatively establish those limits in terms of aspect ratio and concentration for hard ellipsoids of revolution with aspect ratios typical for globular proteins.

2. Methods

2.1. Simulation Details

The surface of a hard ellipsoid of revolution is given by (Frenkel & Mulder, 2002),

$$\frac{z^2}{a^2} + \frac{x^2 + y^2}{b^2} = 1, \tag{4}$$

where $2a$ is the length of the molecule along the symmetry axis and $2b$ is the length of the axes perpendicular to the symmetry axis. The aspect ratio (AR) of the ellipsoid of revolution is defined as $AR \equiv \frac{a}{b}$. In this work, the maximum of either a or b is referred to as α while the minimum is β. 2α was set to unity.

Ellipsoids with aspect ratios 0.333, 0.5, 0.75, 1, 1.5, 2, and 3 were studied at volume fractions ranging from 0.1 to 40%. All of the states studied here lie within the isotropic

region of the hard ellipsoid phase diagram (Frenkel & Mulder, 2002). Our choice for aspect ratios coincides with typical values for globular proteins of which the crystal structures are approximated as ellipsoids (e.g., bovine serum albumin: 0.37 (Heinen *et al.*, 2012), lysozyme: 1.5 (Liu *et al.*, 2011), ovalbumin: 2.6 (Greene *et al.*, 2015)). The maximum concentration studied here is higher than intracellular protein concentrations, which are about 200-300 mg/mL (Ellis, 2001), or, assuming an average specific molar volume of 0.736 mL/g (Mylonas & Svergun, 2007), about 15-22% by volume.

MC simulations (Frenkel & Smit, 2002) were performed within the canonical ensemble (NVT) in a cubic box with side length 40α to generate hard ellipsoid configurations. Ellipsoids were initially placed on a tetragonal lattice within the cubic box and allowed to equilibrate for 10^5 steps. Each step involved 3N attempted moves, where N is the number of particles. On average, half the moves were translations and half were rotations. The average move acceptance was 20 - 30%. After equilibration, configurations were recorded every 50 steps. 500 configurations were analyzed for scattering intensity, structure factor, center-to-center radial distribution function, and an orientational distribution function, $O(r)$.

2.2. Orientational Distribution Function, $O(r)$

The orientational distribution function is defined as,

$$O(r) = \langle |\cos(\theta)|_r \rangle \,, \tag{5}$$

where θ is the relative angle between the symmetry axes of two ellipsoids. The averaging is performed over pairs of ellipsoids whose centers are separated by distance r. The expected value of $O(r)$ for random orientations is $\frac{1}{2}$.

2.3. Scattering Calculations

The measured SAS intensity is the Fourier transform of the excess scattering-length density distribution, $\rho(\mathbf{r})$, averaged over all orientations, Ω,

$$I(q) = \left\langle \left| \int_V \rho(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \right|^2 \right\rangle_\Omega . \tag{6}$$

Eq. (6) was evaluated by discretizing ellipsoid configurations on a 350 $\times$ 350 $\times$ 350 grid and taking its three-dimensional Fourier transform using the FFTW3 software package (Frigo & Johnson, 2005). Orientational averaging was performed by integrating the three-dimensional intensity distribution over the polar and azimuthal angles so that the intensity was a function of only the magnitude of the scattering vector, q.

To establish rigorous limits of the DA, the true structure factor, $S(q)$, was calculated in a similar fashion to scattering intensities. Ellipsoid centers were discretized onto a 350 $\times$ 350 $\times$ 350 grid and the Fourier transform of the grid was taken using FFTW3. The signal was integrated such that it was a function of only q and normalized by the number of ellipsoids.

The accuracy and appropriateness of the DA was determined by comparing three independent calculations of the apparent structure factor, $S'(q)$. First, Eq. 1 was evaluated for $S'(q)$ by dividing the rotationally averaged particle form factor into the calculated scattering intensity. This structure factor is referred to as S'_{APP}. Simulation scattering data for 0.1% were used for the form factor in this work. Second, $S'(q)$ was calculated by modifying the true structure factor derived from the ellipsoid simulations by the DA (Eq. 2). This method is referred to as S'_{COM} as it is based on the ellipsoids' centers of mass. Finally, $S'(q)$ was calculated from the structure factor, $S(q)$, for monodisperse spheres calculated using the Percus-Yevick closure relation (Wertheim, 1963). An effective radius for the monodisperse spheres that conserves particle number density and volume fraction was adopted because this radius gives better agreement

with S'_{APP} than the radius that matches the second virial coefficient. For clarity, this third method is referred to as S'_{PY} as it arises from the Percus-Yevick closure.

The deviation of S'_{COM} from S'_{APP} was quantified by the χ^2 statistic,

$$\chi^2 = \sum_{i=1}^{n} \frac{(S'_{APP}(q_i) - S'_{COM}(q_i))^2}{\sigma^2(q_i)}. \tag{7}$$

Here, n is the number of points and $\sigma(q_i)$ is an assumed standard error in S'_{APP} that was introduced to generate a quantifiable boundary to denote where the DA is applicable and where it is not. Three typical values of standard error were used to determine boundaries denoting the levels of accuracy, 5, 10, and 20%, such that $\sigma(q_i) = 0.05S'_{APP}(q_i)$, etc. In each case, the boundary was set to $\chi^2 = 1$.

Hansen (2013) approximated the apparent structure factor for ellipsoids of revolution with $AR = 0.5$ to 2 by the structure factor for polydisperse spheres. The web application BayesApp, available at http://bayesapp.org/, was used to calculate this structure factor (Hansen, 2012), which is referred to as S'_{Hansen}.

3. Results and Discussion

Figure 1 shows the apparent and true structure factor for ellipsoids with $AR = 0.333$, 0.5, 2, and 3 at volume fractions 10, 20, 30, and 40%. The profiles are shifted vertically for clarity by the respective amounts indicated in Fig. 1. The apparent structure factor, S'_{APP}, for oblate ellipsoids contains two peaks (Fig. 1 A and B), while $S(q)$, S'_{COM} and S'_{PY}, contain only one peak. The peak in $S(q)$, S'_{COM}, and S'_{PY} sharpens, increases in magnitude, and shifts to higher q with increasing concentration. This behavior is observed in the structure factor for hard spheres and can be interpreted to reflect a decrease in the average center-to-center distance between ellipsoids. The first peak in S'_{APP} occurs at approximately the same q value as the peak in S'_{COM} and $S(q)$, so it is likely associated with the average center-to-center distance between ellipsoids. In

contrast, the second peak in S'_{APP} is not associated with any peak in S'_{COM} or in $S(q)$, indicating that it does not correspond to a real correlation length in the system. In all cases, the peak in S'_{COM} and $S(q)$ occurs at smaller q values than the first peak in S'_{APP} and is larger in magnitude.

S'_{APP} differs qualitatively for prolate ellipsoids (Fig. 1 C and D) in that it contains only one primary peak. The peak sharpens, increases in magnitude, and shifts to higher q with increasing concentration, suggesting that it is associated with the average center-to-center distance between ellipsoids. However, as for the case of oblate ellipsoids, the peak in $S(q)$ and S'_{COM} always occurs at lower q values than the peak in S'_{APP}. In contrast, the magnitude of the peak in S'_{APP} is larger than the magnitude of the peak in $S(q)$ and S'_{COM}.

From these data one can observe that the apparent structure factor can be significantly distorted as compared to the true structure factor, especially at high concentrations, indicating that the DA is inadequate in modifying the true structure factor. In general, the primary peak in the apparent structure factor is always located at larger q values than the primary peak in $S(q)$ and S'_{COM}. This is particularly important because the peak position is commonly utilized to estimate key length scales in protein systems (Stradner *et al.*, 2004; Liu *et al.*, 2011). An overestimation of the peak location would result in an underestimation of these length scales. Additionally, the secondary peak in S'_{APP} of oblate ellipsoids does not appear to correspond to a real correlation length in the system and may be an artefact of modelling the intensity as a product of the form factor and apparent structure factor.

Fig. 2 depicts the χ^2 values that quantify the difference between S'_{APP} and S'_{COM} as a function of concentration and aspect ratio. Because the DA is exact for monodisperse spheres, the χ^2 value is zero for $AR = 1$. The solid black contour indicates $\chi^2 = 1$ assuming a standard error of 10%, so state points below this curve have deviations

between S'_{APP} and S'_{COM} of less than 10%. The dashed and dotted contours indicate the corresponding $\chi^2 = 1$ boundary assuming that the standard error in Eq. 7 is 5% or 20% respectively. As shown in Fig. 2, the limits of applicability of the DA depend strongly on particle shape and volume fraction. In general, the DA is valid for higher concentrations for oblate ellipsoids than for prolate ellipsoids, and in both cases, there is a monotonic increase in χ^2 as the aspect ratio approaches unity. The monotonic increase is expected as the DA is exact in the limit $AR \to 1$.

As noted in the introduction, numerous experimental reports of SAS from globular proteins have employed various anisometric shape models in the analysis. Our work can provide guidance as to the level of accuracy afforded by the DA. The effective protein concentration assuming an average specific molar volume of 0.736 mL/g (Mylonas & Svergun, 2007) is shown on the secondary y-axis in Fig. 2 and the aspect ratios of some model globular proteins are indicated on the secondary x-axis. The cross-hatched region in Fig. 2 indicates the approximate intracellular protein concentration, 200-300 mg/mL (Ellis, 2001). As can be observed, the DA is effective only for the ellipsoids deviating least from a spherical shape at these high concentrations. We note that proteins interact via a complex and 'patchy' potential (Neal *et al.*, 1998; Lomakin *et al.*, 1999); however, some cases have been effectively modeled as the sum of an isotropic short-ranged attraction and long-ranged repulsion (Stradner *et al.*, 2004; Liu *et al.*, 2011). Depending on the relative strengths of the attraction and repulsion we expect the boundaries of validity for the DA will change, but nonetheless propose that these boundaries for hard ellipsoids of revolution provide a useful guide for interpreting SAS from protein solutions and from other anisotropic colloidal and nanoparticle suspensions.

Ultimately, the failure of the DA is due to a breakdown in the assumption that particle positions and orientations are not correlated. To show this, the relative par-

ticle orientations and real-space particle distributions were quantified using $O(r)$ and $g(r)$ (Fig. 3 and Fig. 4 respectively). Fig. 3 shows the orientational distribution function for all non-spherical ellipsoids. The separation distance is normalized by 2β, the closest two ellipsoids can approach without overlapping. The horizontal line is drawn at $r/2\beta = 2\alpha$. 2α is an important length scale in the system because it represents the collisional cross-section of the ellipsoids. Particles separated by distances less than this length scale experience an excluded volume effect that hinders certain rotations and effectively couples the particles' orientations with their relative positions. This is evidenced by the systematic increase in $O(r)$ as r decreases below 2α. As the concentration increases a secondary peak develops in $O(r)$ around 2α for $AR = 0.333$ and $AR = 2$ and slightly below 2α for $AR = 3$. The secondary peak in these cases suggests that next-nearest neighbors also have a preferred orientation. This cooperative, multi-body behavior is known to be a signature of pre-nematic swarms in liquid crystals (de Gennes, 1971).

The nearest-neighbor peak in $g(r)$ for ellipsoids is located at a length scale smaller than or equal to 2α (Fig. 4). Technically, the particles separated by this length scale violate the key assumptions of the DA because below 2α, particles adopt preferred relative orientations (Fig. 3). At the lowest concentrations, these configurations are apparently too few in number to significantly affect the DA result. However, as the concentration increases, the nearest-neighbor peak grows and moves to smaller r. These trends in $g(r)$ result in increased neighboring particle alignment that leads to an unacceptable error in the DA.

The model calculations presented here provide guidance in analyzing SAS data from anisotropic particles, such as often performed with proteins. It is demonstrated that caution is required in interpreting SAS from protein solutions or anisotropic colloidal suspensions that are near or above the boundary in Fig. 2, especially under

physiological conditions, or other conditions that promote attraction. Any association between particles could more strongly couple their relative positions and orientations and therefore further limit the validity of the DA. This situation would require a more detailed model, such as a coarse-grained or full atomistic simulation.

For systems where the DA breaks down, a potential alternative is the technique developed by Hansen (2013), who showed that the apparent structure factor for hard ellipsoids with aspect ratios ranging from 0.5 to 2.0 may be approximated by the structure factor for polydisperse spheres. The method explicitly accounts for excluded volume effects not considered in the DA. Here we have used Hansen's method to calculate $S'(q)$ and compare it to S'_{APP}. The results for $AR = 0.5$ and $AR = 2$ at volume fractions of 10, 20, 30% are shown in Fig. 5. For $AR = 0.5$, Hansen's method performs worse than the DA, while for $AR = 2$, it outperforms the DA. Interestingly, Hansens's method does not account for the secondary peak observed in S'_{APP} for oblate ellipsoids, indicating that the excluded volume contributions to the apparent structure factor are not adequate in reproducing this peak. The origin of the peak is still unclear.

The deviation between S'_{Hansen} and S'_{APP}, χ^2_{Hansen} was computed using an expression identical to Eq. 7 where S'_{COM} has been replaced with S'_{Hansen}. The standard error was assumed to be 10% of S'_{APP}. The ratio, of χ^2_{Hansen} to the $\chi^2_{10\%}$ reported in Fig. 2, is shown in Fig. 6. Ratio values less than unity indicate that Hansen's method performs better than the DA, while values greater than unity indicate the converse. The ratio calculation shows that Hansen's method works better than the DA for prolate ellipsoids, up to a factor of $\sim$2, and also performs better than the DA for $AR = 0.75$. We are unaware of any literature applying this technique to protein solutions, but, based on the comparisons here, it appears more appropriate to use than the DA for the interpretation of SAS from protein solutions.

4. Conclusions

The DA is valid only within the contours established in Fig. 2. The most significant error that is not corrected by the DA is an overestimation of the primary peak location in the structure factor, which leads to an underestimation of the correlation length. Additionally, spurious peaks can be introduced into the apparent structure factor that do not correspond to any real length correlation length scale in the system. The DA fails when there are a significant number of ellipsoidal pairs that are separated by less than 2α. These model calculations provide a useful starting point for the analysis of SAS from protein solutions and anisotropic colloidal and nanoparticle dispersions.

We are grateful to Dr. Laura Filion, who provided the simulation code used in this work. We are grateful for support from the National Science Foundation under award number DMR-1209424. Support from the University of Delaware Center for Bioinformatics and Computational Biology Core Facility and use of the BioHen compute cluster was made possible through funding from the Delaware INBRE (NIGMS 2P20GM103446), Delaware EPSCoR (NSF EPS-0814251, NSF IIA-1330446), the State of Delaware, and the Delaware Biotechnology Institute.

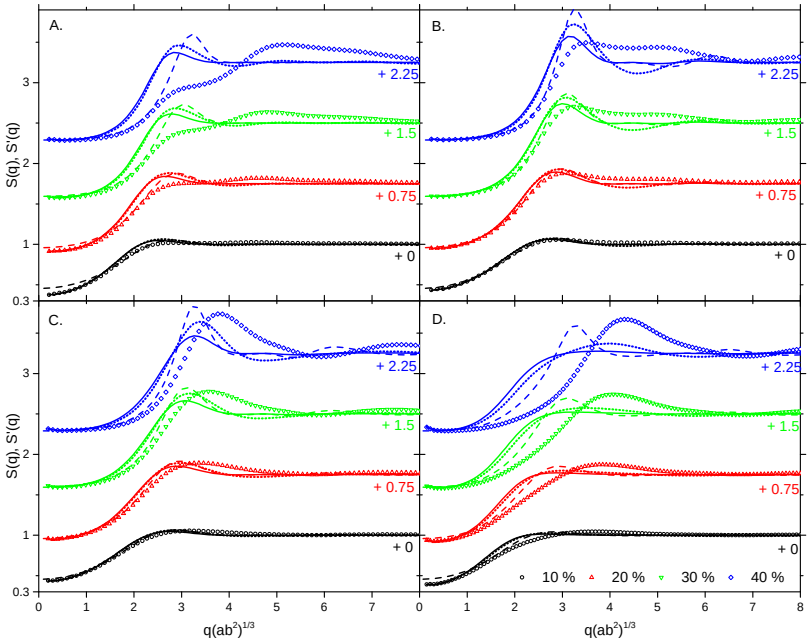

Fig. 1. The true structure factor, $S(q)$, (dotted lines) and apparent structure factors, S'_{APP} (open symbols), S'_{COM} (solid lines), and S'_{PY} (dashed lines). Values of AR: A, 0.333; B, 0.5; C, 2; D, 3. Values of volume fraction: black, 10%; red, 20%; green, 30%; blue, 40%.

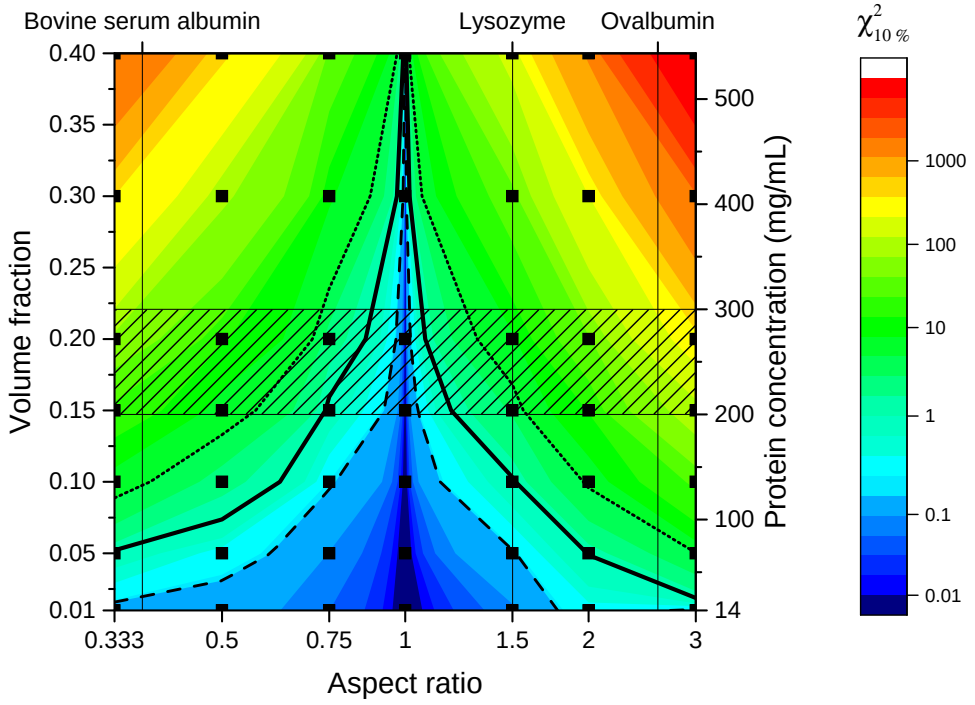

Fig. 2. χ^2 contours characterizing the deviation of S'_{APP} from S'_{COM}. Points studied in this work are marked with squares. The solid line indicates the $\chi^2 = 1$ boundary assuming a standard error of 10% in Eq. 7, the dashed line is the boundary assuming a standard error of 5% and the dotted line is the boundary assuming a standard error of 20%. The colors are based on the 10% standard error. The equivalent protein concentration, assuming a specific molar volume of 0.736 mL/g (Mylonas & Svergun, 2007), is shown on the right axis. The cross-hatched region indicates intracellular protein concentrations (Ellis, 2001) and the aspect ratios of several model proteins are shown on the top axis.

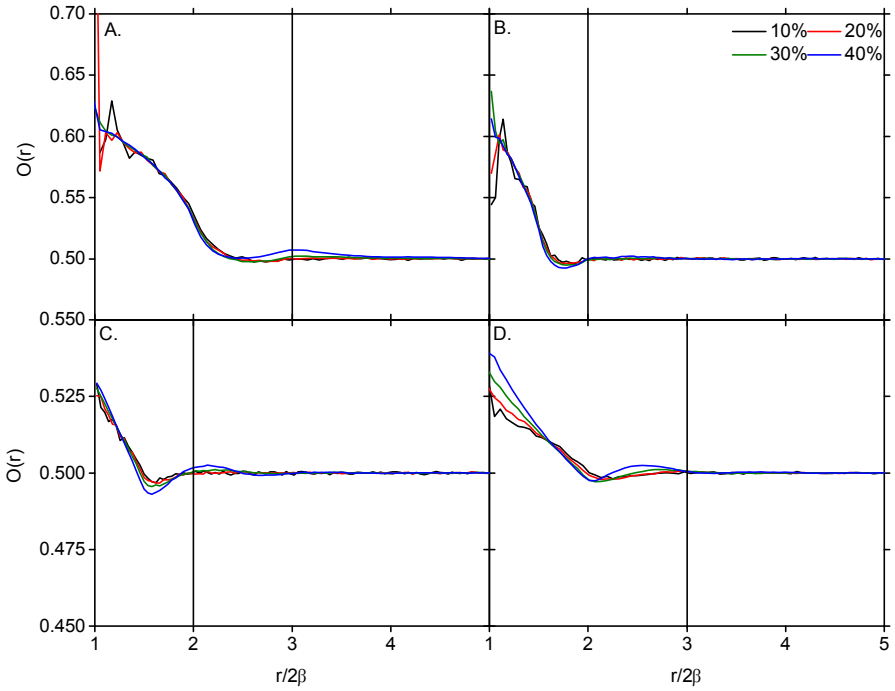

Fig. 3. Orientational distribution function for hard ellipsoids. Values of AR: A, 0.333; B, 0.5; C, 2; D, 3. Values of volume fraction: black, 10%; red, 20%; green, 30%; blue, 40%. Horizontal lines are drawn at $r/2\beta = 2\alpha$.

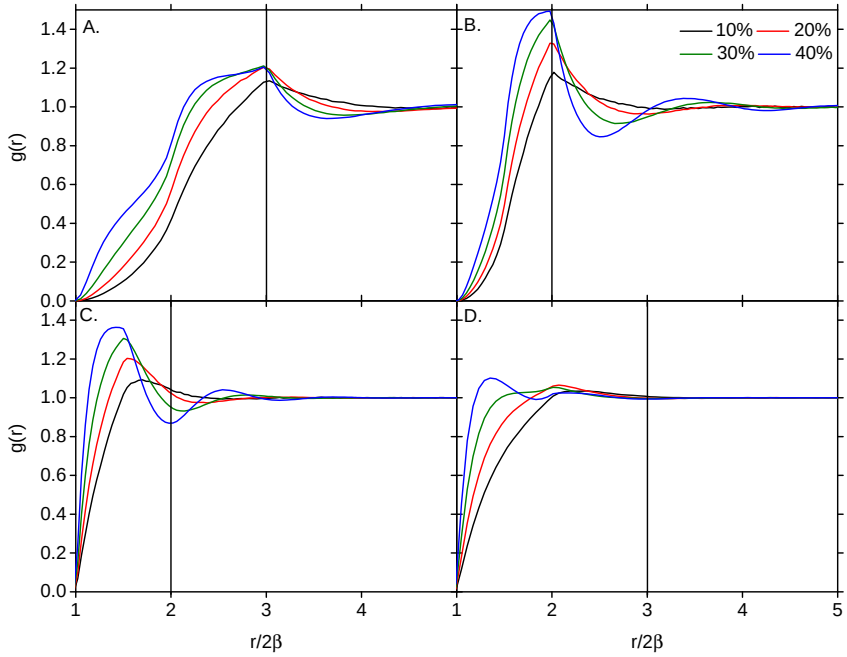

Fig. 4. Radial distribution functions for hard ellipsoids. Values of AR: A, 0.333; B, 0.5; C, 2; D, 3. Values of volume fraction: black, 10%; red, 20%; green, 30%; blue, 40%. Horizontal lines are drawn at $r/2\beta = 2\alpha$.

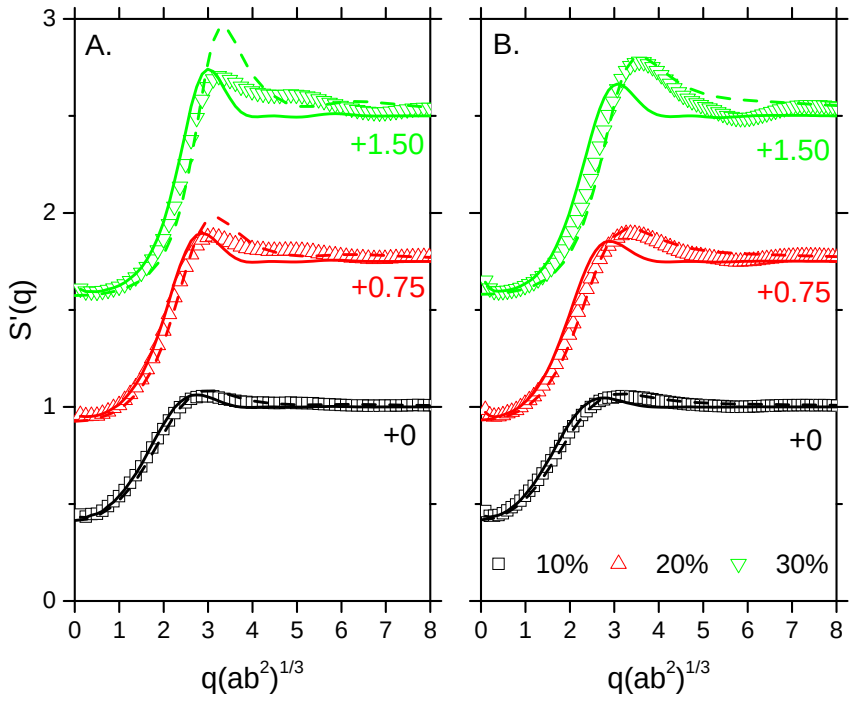

Fig. 5. Apparent structure factors, S'_{HANSEN} (dashed line), S'_{COM} (solid line), and S'_{APP} (open symbols). Values of AR: A, 0.5; B, 2. Values of volume fraction: black, 10%; red, 20%; green, 30%.

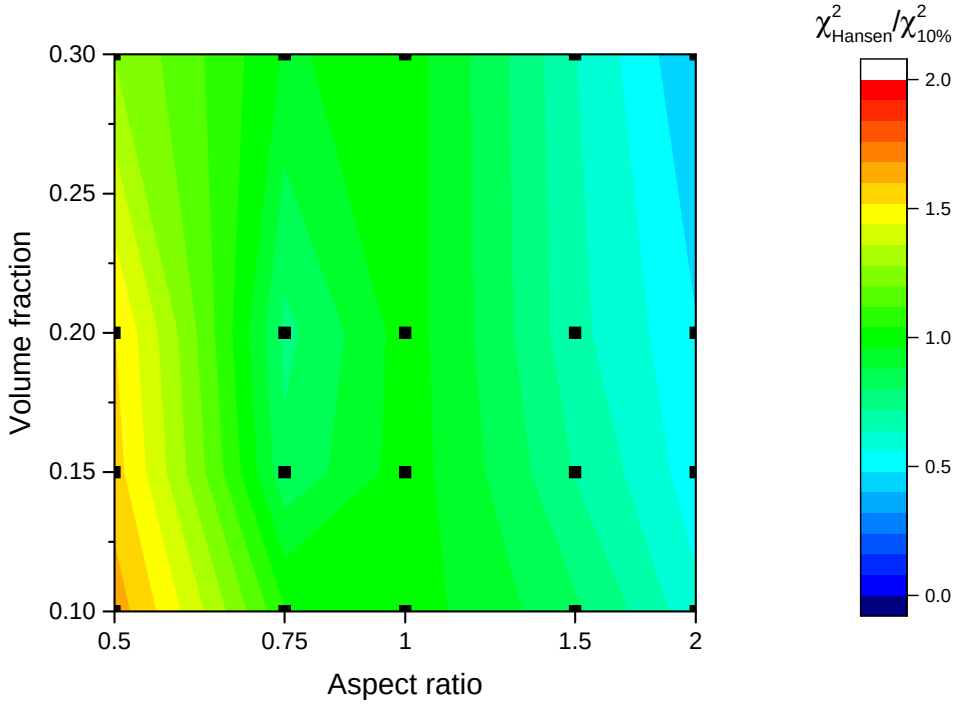

Fig. 6. Ratio of χ^2_{Hansen} to $\chi^2_{10\%}$ as a function of volume fraction and aspect ratio. Points studied in this work are marked with squares.

References

Ellis, R. J. (2001). *Curr. Opin. Struct. Biol.* **11**(1), 114–119.

Frenkel, D. & Mulder, B. M. (2002). *Mol. Phys.* **100**(1), 201–217.

Frenkel, D. & Smit, B. (2002). *Understanding molecular simulation: From algorithms to applications*. New York, New York: Academic Press, 2nd ed.

Frigo, M. & Johnson, S. G. (2005). *Proc. IEEE*, **93**(2), 216–231.

de Gennes, P. (1971). *Mol. Cryst. Liq. Cryst.* **12**, 193–214.

Godfrin, P. D., Zarraga, I. E., Zarzar, J., Porcar, L., Falus, P., Wagner, N. J. & Liu, Y. (2016). *J. Phys. Chem. B*, **120**(2), 278–291.

Greene, D. G., Modla, S., Wagner, N. J., Sandler, S. I. & Lenhoff, A. M. (2015). *Biophys. J.* **109**(8), 1716–1723.

Hansen, S. (2012). *J. Appl. Crystallogr.* **45**(3), 566–567.

Hansen, S. (2013). *J. Appl. Crystallogr.* **46**, 1008–1016.

Heinen, M., Zanini, F., Roosen-Runge, F., Fedunová, D., Zhang, F., Hennig, M., Seydel, T., Schweins, R., Sztucki, M., Antalík, M., Schreiber, F. & Nägele, G. (2012). *Soft Matter*, **8**(5), 1404.

Kotlarchyk, M. & Chen, S. (1983). *J. Chem. Phys.* **79**, 2461–2469.

Liu, Y., Porcar, L., Chen, J., Chen, W.-R., Falus, P., Faraone, A., Fratini, E., Hong, K. & Baglioni, P. (2011). *J. Phys. Chem. B*, **115**(22), 7238–47.

Lomakin, A., Asherie, N. & Benedek, G. B. (1999). *Proc. Natl. Acad. Sci. U.S.A.* **96**(17), 9465–8.

Mylonas, E. & Svergun, D. I. (2007). *J. Appl. Crystallogr.* **40**(s1), s245–s249.

Neal, B. L., Asthagiri, D. & Lenhoff, A. M. (1998). *Biophys. J.* **75**(5), 2469–77.

Pedersen, J. S. (1997). *Adv. Colloid Interface Sci.* **70**, 171–210.

Stradner, A., Sedgwick, H., Cardinaux, F., Poon, W. C. K., Egelhaaf, S. U. & Schurtenberger, P. (2004). *Nature*, **432**(7016), 492–5.

Wagner, N. J., Krause, R., Rennie, A. R., D'Aguanno, B. & Goodwin, J. (1991). *J. Chem. Phys.* **95**(1), 494.

Wertheim, M. S. (1963). *Phys. Rev. Lett.* **10**(8), 321–323.

Yearley, E. J., Godfrin, P. D., Perevozchikova, T., Zhang, H., Falus, P., Porcar, L., Nagao, M., Curtis, J. E., Gawande, P., Taing, R., Zarraga, I. E., Wagner, N. J. & Liu, Y. (2014). *Biophys. J.* **106**(8), 1763–70.

Yearley, E. J., Zarraga, I. E., Shire, S. J., Scherer, T. M., Gokarn, Y. & Wagner, N. J. (2013). *Biophys. J.* **105**(3), 720–731.

Synopsis

The limitations of the popular decoupling approximation used for the analysis of small-angle scattering from dispersions of anisotropic particles are established for ellipsoids of revolution of aspect ratio 0.333 to 3 as a function of particle concentration. The results provide valuable guidance for the analysis and interpretation of SAS from colloidal, nanoparticle, and protein solutions and suspensions.