Local crystalline structure in an amorphous protein dense phase

[Running title: Crystal order in a protein precipitate}

Daniel G. Greene[1], Shannon Modla[2], Norman J. Wagner[1], Stanley I. Sandler[1], Abraham M. Lenhoff[1*]

[1] Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19716

[2] Delaware Biotechnology Institute, 15 Innovation Way, Newark, DE 19711

Keywords: protein salting-out, protein crystallization, protein-protein interactions, small-angle neutron scattering

[*] Corresponding author. Phone +1 302 831 8989; e-mail lenhoff@udel.edu.

Abstract

Proteins exhibit a variety of dense phases ranging from gels, aggregates, and precipitates to crystalline phases and dense liquids. While the structure of the crystalline phase is known to atomistic detail, little attention has been paid to non-crystalline protein dense phases; in many cases the structures of these phases are assumed to be amorphous. In this work, we measured the structure of ovalbumin precipitate particles salted-out with ammonium sulfate using small-angle neutron scattering, electron microscopy and electron tomography. We found that ovalbumin phase-separates into core-shell particles with core radius ~ 2 μm and shell thickness ~ 0.5 μm. Within this shell region, nanostructures comprised of crystallites of ovalbumin self-assemble into a well-defined bicontinuous network with branches ~12 nm thick. These results demonstrate that the protein gel is comprised in part of nanocrystalline protein.

1 Introduction

Protein dense phases such as gels, aggregates, and precipitates arise both *in vivo* and during protein solution processing such as in the biotechnology or pharmaceutical industries. While certain forms may be desirable, for instance protein crystals (1) or monoclonal antibody clusters (2), others, like non-native aggregates, are not. The physical nature of the dense phase, whether crystal, dense liquid, or a non-equilibrium phase such as a gel, is strongly dependent on the solution conditions as well as the nature of the protein molecule. Consequently, tasks such as protein crystallization can be cumbersome and require extensive empirical exploration (1), whereas in other circumstances unwanted dense phases may appear unexpectedly.

Both equilibrium and non-equilibrium phase boundaries of protein solutions have been measured extensively (3–5) and the equilibrium results have been interpreted within the theoretical framework of suspensions of colloidal particles with short-ranged attractive interactions (3). In addition to crystallization, protein solutions can also undergo a metastable liquid-liquid phase separation when the attraction between protein molecules is sufficiently strong. Changing the temperature or adding salts or non-adsorbing polymers have been used to alter the attraction strength, and qualitatively similar trends among the three methods are observed in the phase behavior (3–6). As with colloidal suspensions, the liquid-liquid phase transition can be arrested by non-equilibrium phases such as gels, aggregates, and precipitates (3, 7–9).

Although crystal structures are known to atomistic detail and dense liquid phases can often be described using integral equation theories (10–12), the structure of non-equilibrium dense phases can be challenging to predict and measure. Despite simulations of colloids with a short-ranged interaction potential that show ordered regions developing within dense phases (13, 14), and experimental evidence that nominally amorphous solids can be used to seed crystal growth (15, 16), such non-equilibrium dense phases are typically assumed to be amorphous. A more detailed characterization of the structure of non-equilibrium protein dense phases would therefore aid in understanding the complexities of protein phase separation.

In this study, transmission electron microscopy (TEM), electron tomography, ultra-small- and small-angle neutron scattering (USANS and SANS), and computational tools are combined to measure the microstructure of a model protein precipitate produced by salting-out ovalbumin using ammonium sulfate. This provides the first reported description of a non-equilibrium protein dense phase characterized from the molecular to macroscopic length scales.

2 Materials and Methods

2.1 Protein and chemicals

Deuterium oxide (CAS 7789-20-0) was purchased from Acros Organics (Pittsburgh, PA) for use in improving the neutron scattering contrast. Sodium hydroxide (CAS 1310-73-2) and

sodium phosphate (CAS 7558-80-7) were purchased from Fisher Scientific (Pittsburgh, PA). Ammonium sulfate (CAS 7783-20-2) was purchased from Sigma (St. Louis, MO). Osmium tetroxide (CAS 20816-12-0), glass-distilled acetone (DA) (CAS 67-64-1), ethylene glycol diglycidyl ether (CAS 2224-15-9) resin (Quetol 651), nonenylsuccinic anhydride (NSA) (CAS 28928-97-4), n-butyl glycidyl ether (n-BGE) (CAS 2426-08-6), uranyl acetate (CAS 541-09-3), lead nitrate (CAS 10099-74-8), sodium citrate (CAS 6132-04-3), and 200 mesh formvar carbon-coated copper grids were purchased from Electron Microscopy Services (Hatfield, PA). Gold fiducial particles (15 and 20 nm) were purchased from Ted Pella (Redding, CA).

Ovalbumin was recrystallized three times from fresh single-comb white Leghorn eggs following the purification protocol used by Judge et al. (17). After purity was checked using silver-stained SDS-PAGE, the ovalbumin crystals were lyophilized. Before subsequent use, lyophilized ovalbumin was reconstituted in and extensively dialyzed against 5 mM sodium phosphate buffer, pH 7.0. For deuterated samples used for SANS, ovalbumin was re-lyophilized to remove residual water and reconstituted in D_2O containing 5 mM sodium phosphate, pD 7.0. On a standard glass electrode, the pD is the measured pH plus 0.4 (18). The ovalbumin solutions were concentrated to 100-150 mg/mL using a 10K MWCO Amicon Ultra-4 centrifugal filter from Millipore. Protein concentrations were measured using absorbance at 280 nm, where the ovalbumin extinction coefficient is $E^{1\%}_{1\,cm} = 7.35$ (19).

2.2 Phase diagram

The phase boundaries for ovalbumin in ammonium sulfate were determined by microbatch measurements (3) in which samples were prepared by mixing in an Eppendorf tube (Fisher Scientific, 05-408-120), in order, a concentrated salt solution, 5 mM sodium phosphate buffer at pH 7.0, and a concentrated ovalbumin solution. The same mixing procedure was used for deuterated samples except that all solutions were prepared in D_2O and adjusted to pD 7.0. Samples were immediately checked for phase separation by visual inspection and were subsequently monitored for 2 weeks to determine phase behavior. Boundary points were determined as the average of the lowest salt concentration at which phase separation occurred and the next lower salt concentration studied.

2.3 TEM and electron tomography

Precipitate samples were prepared as described in section 2.2 such that the final protein concentration was approximately 5 mg/mL and the final ammonium sulfate concentrations were 2.0, 2.4 and 2.8 M. The precipitates were spun down at 11,400 g for 2 minutes. Pellets were pipetted into 1.2 mm x 200 µm flat specimen carriers and high-pressure frozen in a Leica EMPACT. Samples were transferred under liquid nitrogen to a Leica AFS for freeze substitution in 2% osmium tetroxide containing 1% water and 99% acetone. Freeze-substituted samples were held at -85 °C for 105 hours, were subsequently warmed to -20 °C over 15 hours and held at -20 °C for an additional 3 hours. Samples were warmed to 4 °C over 2 hours and then taken out of the AFS and kept at room temperature for 1.5 hours. Room-temperature samples were washed

twice with DA and stored overnight at 4 °C. The next day, samples were washed 3 times for 15 minutes each with DA. After freeze-substitution, samples were infiltrated with a mixture of 1 part n-BGE to 1 part DA for 30 minutes. Samples were then infiltrated with 100% n-BGE for 30 minutes followed by infiltration with Quetol 651-NSA resin (QNSA) for 60 hours. Samples were then embedded in TAAB flat embedding capsules and polymerized at 60 °C for 48 hours. Following polymerization, the samples were sectioned on a Reichert Jung Ultracut E ultramicrotome using a Diatome ultra diamond knife. Sections 60-70 nm thick were flattened with a Pelco HP1 heat pen and collected onto 200 mesh formvar carbon-coated copper grids. Sections were stained for 5 minutes with 2% methanolic uranyl acetate, washed in methanol, and dried for 15-30 minutes. The sections were then stained with Reynolds' lead citrate for 5 minutes and washed with boiled, cooled nanopure water. Grids were imaged on a Zeiss Libra 120 transmission electron microscope operated at 120 kV. Images were acquired using a Gatan Ultrascan 100 CCD.

Sample fixation and staining for TEM tomography was performed in a similar fashion to that for TEM imaging, but samples were sectioned to 200 nm. Gold fiducial particles were applied to both sides of the grid, with 15 nm particles attached to one side and 20 nm particles to the opposite side. Dual-axis tomograms were collected on a Zeiss Libra 120 transmission electron microscope operating at 120 kV. Sections were pre-irradiated under the beam to minimize specimen shrinkage during acquisition of the tilt series. Samples were tilted over a +60° to -60° tilt range, and images were acquired at every 1° of tilt using a Gatan Ultrascan 100 CCD camera. The grid was then removed from the microscope and manually rotated 90°, the region of interest was relocated, and a second tilt series was collected over a +60° to -60° tilt range with images acquired at every 1° of tilt. The dual-axis tilt series was reconstructed using an R-weighted back projection in IMOD (20).

2.4 Neutron scattering

2.4.1 Experimental

Precipitate samples were prepared as described in Section 2.2 at a final protein concentration of about 5 mg/mL. The final ammonium sulfate concentrations ranged from 2.4 to 2.9 M and samples were aged for 28 days, 10 days, 3 days, 2 days, hours and minutes before measurement. Measurements on samples aged 2 days and longer were made on the NG-7 30 m SANS instrument at the National Institute of Standards and Technology (NIST) Center for Neutron Research (NCNR) using three sample-to-detector distances, 1 m, 4 m, and 13.5 m with lenses, and two neutron wavelengths, 0.6 nm for the 1 m and 4 m distances and 0.8 nm for the 13.5 m distance, allowing coverage of a q-range of 0.009–5.19 nm^{-1}. For early-time samples (20 minutes and 3 hours), measurements were made on the CG-3 beamline at the High Flux Isotope Reactor (HFIR) at Oak Ridge National Laboratory using two sample-to-detector distances, 30 cm and 6 m, with 0.6 nm neutrons, to cover a q-range of 0.070–7.37 nm^{-1}. USANS measurements on samples prepared in 2.4, 2.5 and 2.8 M ammonium sulfate and aged 2 days prior to measurement were made on the BT-5 USANS instrument at the NCNR. Neutrons with a

wavelength of 0.24 nm were used to cover a q-range of 0.0004–0.003 nm^{-1}. For the NCNR measurements, samples were loaded into 1 mm thick demountable titanium cells with quartz windows, and for the HFIR measurements, samples were loaded into 1 mm thick quartz banjo cells.

NCNR data were placed on an absolute scale and were reduced using IGOR Pro NCNR SANS software (21) and HFIR data were reduced and placed on an absolute scale using IGOR Pro HFIR SANS software. Intensities were corrected for background due to incoherent scattering by fitting the high-q region ($q > 2.5$ nm^{-1}) to Porod's Law, $I(q) = Aq^{-4} + B$, in MATLAB (The MathWorks, Natick, MA) and subtracting B, the background, from the intensities. Unless otherwise stated, all data fitting was performed using the IGOR Pro NCNR SANS software (21).

2.4.2 SANS structure analysis

Crystalline clusters of ovalbumin were generated *in silico* from the Protein Data Bank (PDB) file, 1OVA, by propagating the unit cell along the three principal crystallographic axes (22). A list of the 21 crystalline clusters considered is given in Table S1 in the Supporting Information. Their form factors were calculated using the CRYSON software package, which calculates the spherically averaged form factor from atomic coordinates in a PDB file (23). The highest-resolution implementation of the software was used. We assumed that 90% of labile hydrogens in the protein exchange with deuterium from the solvent, resulting in a scattering length density of 2.95×10^{-4} nm^{-2} for the ovalbumin monomer, in reasonable agreement with the experimental value of 3.25×10^{-4} nm^{-2} (24). Since neutron scattering from the hydration layer is negligible, its contribution was set to zero in the calculations (12). Using MATLAB, the mid- to high-q region of the scattering data was fit to all possible linear combinations of the 21 clusters and a free monomer term according to

$$I(q) - B = \sum_{i=1}^{n} \beta_i P_i(q) + \beta_m P_m(q) + \varepsilon \qquad (1)$$

where B is the background scattering, n is the total number of clusters in consideration, β_i is the number density of cluster i, β_m is the number density of free monomer, P_i is the scattering cross-section of cluster i, P_m is the scattering cross-section of free monomer and ε is an error term. Eq. 1 assumes that cluster-cluster structure factors are unity and cross-correlation terms are negligible. Model fits containing negative coefficients were excluded as these are unphysical, and the best model was selected from the remaining candidates using the Akaike information criterion (AIC) (25).

Clusters of randomly oriented ellipsoids, representing ovalbumin monomers, were generated using an event-driven molecular dynamics simulation package, PackLSD (26, 27). The dimensions of the ellipsoids were set by fitting scattering data from a dilute suspension of ovalbumin monomers. The ellipsoids were packed to the protein volume fraction in the ovalbumin crystal, 0.64. The scattering length density distribution, $\rho(\mathbf{r})$, was generated by

randomly sampling 100,000 points within the clusters and the cluster scattering cross-sections were calculated by taking the Fourier transform of $\rho(\mathbf{r})$. The scattering data were fit, using MATLAB, to an expression similar to Eq. 1,

$$I(q) - B = \beta_c P_c(q) + \beta_m P_m(q) \tag{2}$$

where the subscript c refers to the random cluster. Standard models for the scattering form factors used in this work are found in the Supporting Information.

Scherrer's relation(28),

$$S = \frac{K\lambda}{W\cos\theta} \tag{3}$$

provides an estimate for the crystal domain size of a given Bragg reflection and was used here to estimate the domain size of nanocrystalline ovalbumin clusters. Here S is the domain thickness, K is the Scherrer constant, λ is the incident wavelength, W is the full width at half-maximum (FWHM) for the Bragg peak and θ is the Bragg peak angle. We assumed $K = 1$, but it can range from 0.9 to 2 depending on the crystallite shape and the Miller indices of the reflection (29). W was estimated by fitting a Voigt function, i.e., a convolution of a Lorentzian and a Gaussian function, to a peak in the scattering data. The Gaussian function accounts for instrumental smearing while the Lorentzian accounts for peak broadening due to the crystallite domain size. We did not take into account broadening due to crystallite strain because neutron data do not have the resolution to perform the necessary analysis.

3 Results and Discussion

3.1 Ovalbumin phase diagram in H_2O and D_2O

The phase diagram of ovalbumin in ammonium sulfate solutions that was measured here, shown in Figure 1, is consistent with previously reported measurements (3), but differs slightly due to the solvent effect for samples prepared in D_2O for SANS measurements. Similar to Dumetz et al. (3) we found two kinetically distinct regimes where ovalbumin phase-separates into a protein-rich dense phase and a protein-lean light phase (Fig. 1). In the first regime, which occurs at salt concentrations ~1.8 M – 2.0 M, dense ovalbumin gel beads that are of order 100 μm in diameter (Fig. 1b) form on the order of hours to days, whereas for higher salt concentrations, > ~2.0 M, phase separation into droplets ~5 μm in diameter (Fig. 1c) occurs instantaneously. The scale on the ordinate in Fig. 1a is inverted to mimic the general colloidal phase diagram for short-ranged attractive potentials. Dumetz et al. interpreted the two boundaries as being a liquid-liquid binodal and spinodal respectively (3).

There is a well-documented isotope effect associated with D_2O that shifts phase boundaries due to a difference in hydrogen bonding between H_2O and D_2O (30, 31). Despite the

shift in the boundary, we still observe equivalent microstructures of dense phases formed in H_2O and D_2O, as confirmed by small-angle x-ray scattering data (Fig. S1 in Supporting Information). Also as in H_2O, crystals eventually nucleate and grow out of the dense phases formed in D_2O. Specifically, for samples prepared at 5 mg/mL ovalbumin and at ammonium sulfate concentrations greater than 2.4 M and less than 2.8 M, crystals were observed to nucleate and grow within 30 days, but no crystals were observed in samples prepared at 2.8 M ammonium sulfate or higher.

3.2 TEM

TEM measurements reveal the internal structure of the ovalbumin precipitate particles. The micrograph in Fig. 2a, in which the protein is stained and appears dark, shows that ovalbumin precipitate particles are a few microns in diameter and exhibit a dense shell surrounding a less dense core. Elongated, crescent-shaped objects, henceforth termed "clusters", are present throughout the particle, but are thicker and present in higher quantities in the shell than in the core (Fig. 2b and c). Within the shell, clusters are 10 – 20 nm thick and are spaced on the order of 100 nm (Fig. 2c), while in the core, clusters are thinner than in the shell and spaced on a larger length scale (Fig. 2b). The average of 106 measurements on the shell yields a cluster thickness of 12.1 ± 3.4 nm.

Although Figures 2b and c provide an indication of the cluster dimensions, they do not indicate the shape or extent of the structures perpendicular to the plane of the images. The three-dimensional structure was therefore observed using electron tomography on a 200 nm thick section taken from the shell region. A surface rendering (Fig. 2d) shows that the interior of the shell region is tortuous and highly porous, as is characteristic of the microstructure of coarsened colloidal gels (32). It is also apparent from the tomographic data that the TEM clusters have a complex shape that is sheet-like. A video of the reconstructed tomogram is available in the Supporting Information.

One possible mechanism for the formation of the core-shell gel particles is that upon quenching into the two-phase region, the solution undergoes a liquid-liquid phase separation that results in the formation of many dense, protein-rich, liquid droplets. Nucleation occurs at the droplet interface and, due to the lack of protein outside the droplets, growth proceeds from the edge of the droplet inwards. Eventually, most of the protein in the droplet is incorporated into the porous network and growth stops, resulting in a dense shell surrounding a less dense core. We will examine the network formation kinetics and thus the mechanism in a separate publication.

3.3 Neutron scattering

Neutron scattering measurements complement the TEM results in that they provide statistically accurate measurements of the average shell thickness, core radius, and size of the clusters. Importantly, neutron scattering also provides information regarding the molecular-level structure that is unresolvable through TEM.

USANS data cover small values of q, representing dimensions on the scale of overall precipitate particles. In Figure 3, USANS intensities for precipitate suspensions prepared in 2.4, 2.5 and 2.8 M ammonium sulfate are shown along with the best fit to a polydisperse core-shell model, where the polydispersity in the core is equal to that in the shell, for which fitted parameters are given in Table 1. The data in Fig. 3 are not desmeared and are compared to properly smeared models. Both the 2.5 and 2.8 M samples show well-defined minima in their scattering spectra while the 2.4 M sample does not. We set the polydispersity equal to 10% for the 2.5 and 2.8 M cases and allowed fitting of the polydispersity in the 2.4 M case. Despite the absence of minima in the 2.4 M data, the fitted values for the core radius (722, 1864, 1834 nm for 2.4, 2.5 and 2.8 M respectively) and shell thickness (231, 750, 235 nm respectively) are reasonable considering the length scales observed by TEM.

SANS probes length scales from nanometers to hundreds of nanometers and so is ideally suited to measuring the local structure of the shell network. SANS measurements at 2.5 M ammonium sulfate show that the precipitate microstructure evolves initially over a period of order one hour but is then stable over the course of days to weeks (Fig. 4a), until crystals form. Scattering in the latter samples is dominated at low q by scattering from crystal surfaces, as indicated by the q^{-4} scaling. A strong scattering peak at $q \sim 1.6$ nm^{-1} develops as the samples crystallize (Fig. 4a inset). Surprisingly, the peak is observed in samples that are *not macroscopically crystalline*; the samples appear as those in Fig. 1c and do not exhibit birefringence or sharp edges. Samples at other salt concentrations over the range of examined salt conditions show similar microstructures on intermediate time scales (Fig. 4b), but no crystals appear within 28 days in samples prepared in 2.8 M ammonium sulfate and higher, and their microstructure remains constant over the 28-day experiment. A more detailed investigation of the dynamics will be reported in future work.

In the absence of macroscopic crystals, structural information is found on three length scales in the SANS measurements (Fig. 4b); two regions correspond to features observed by TEM and a third occurs on a length scale similar to the size of the protein monomer, which was unresolvable in TEM. First, there is a low-q peak at 0.08 nm^{-1} and a secondary peak at 0.15 nm^{-1} which, using Bragg's Law, $d = 2\pi/q$, correspond to real-space length scales, d, of 78 nm and 41 nm respectively. The peak location ratio, 0.15 nm^{-1}/0.08 nm^{-1}, is 1.9, which is close to the value of 2 expected for lamellae (33) and confirms the clusters' sheet-like nature. The presence of two scattering peaks and their locations confirm that ovalbumin precipitate particles contain void space with characteristic spacing similar to that measured by TEM.

The second features of interest are a downturn and a secondary peak in the mid-q region from q = 0.15 to 1.0 nm^{-1}. The primary minimum in this region occurs at q = 0.5 nm^{-1} and corresponds to a length scale of 12.6 nm, in agreement with the width of the clusters measured by TEM. To estimate the dimensions of the clusters, various form-factor models, including those for lamellae, cylinders, and ellipsoids, were fit to the mid-q region of the 2- and 3-day data, i.e., in the absence of macroscopic crystals. The models were able to capture the location of the

minimum as well as the location of the peak. However, no model was capable of describing the width of the peak. The fitted dimensions (Table 2) are in good qualitative agreement with the dimensions measured via TEM.

Finally, a Bragg-like reflection emerges at q ~ 1.6 nm^{-1}, which corresponds to a length scale, d = 4 nm, that is commensurate with the molecular size of ovalbumin and is likely due to monomer-monomer spatial correlations within the clusters. The high-q peak is consistent with 6 of the allowed reflections in the ovalbumin unit cell (Fig. S2 in Supporting Information). The peak broadens for two additional reasons: 1) The peak width is inversely proportional to the size of the crystallite (28), which here is of order 10 nm and introduces a broadening of ~ 0.7 nm^{-1} beyond the broadening due to instrumental smearing. 2) The SANS measurement gives rise to peak broadening due to the wavelength spread, finite collimation of the neutron beam, detector resolution, and radial averaging of the scattered intensity (34), which here would lead to broadening of ~ 0.1 nm^{-1} at q = 1.6 nm^{-1}. Other than the peak at q = 1.6 nm^{-1}, reflections at lower q values are not seen because of interference from scattering from the lamellae, and higher-order reflections are not seen because they either do not scatter above background or are effectively broadened out to the point where they do not scatter by the mechanisms previously described. SAXS data (Fig. S2 in Supporting Information), where the contrast mechanism is different to that in SANS, show strong scattering peaks at q-values of 1.08, 1.57 and 2.04 nm^{-1}, which are also consistent with allowed reflections. SAXS data are subject to much less instrumental smearing than in SANS, which explains why higher-order reflections are apparent in the SAXS data but not in the SANS data. Since both data sets contain strong peaks that are consistent with the unit cell reflections it is likely that they are caused by a highly ordered, microcrystalline structure.

We rigorously test the hypothesis that the high-q peak in the SANS data is the result of ordered packing as opposed to random packing by examining two ideal cases: 1) a crystal packing of ovalbumin in clusters, and 2) a random packing of ellipsoids in clusters, in both of which the cluster size is constrained by the dimensions derived from the mid-q fitting. As explained in section 2.4.2, crystal packings of ovalbumin were generated by assembling unit cells along the three principal crystallographic axes and their form factors were calculated using CRYSON. The best model was selected using the AIC. For ammonium sulfate concentrations less than 2.6 M the best model consists of two cluster terms and a monomer term; the cluster sizes are 5 x 2 x 2 unit cells and 3 x 5 x 2 unit cells, which have approximate dimensions 31.5 x 16.9 x 14.3 nm and 18.9 x 42.4 x 14.3 nm respectively. For ammonium sulfate concentrations greater than or equal to 2.6 M, the best model consists of the 3 x 5 x 2 cluster term and the monomer term (Fig. 5a). The models account for 79-93% of the total protein in the samples. The crystalline cluster models quantitatively capture the mid-q upturn, the location of the mid-q peak and the location and width of the high-q peak, but because the cluster shape is an idealization the models do not accurately capture the width of the mid-q peak.

In the interest of model discrimination we also consider whether the fits are due primarily to the packing density of proteins within the clusters or if the crystalline ordering is a necessary condition as well. We therefore use as a model a randomly packed cluster, for which the form factor was calculated based on randomly packed ellipsoids of dimensions equivalent to ovalbumin monomers. Ellipsoids were packed into a box with the same dimensions as the 3 x 5 x 2 cluster and to the molecular volume fraction in ovalbumin crystals, 0.64 (22). Figure 5b shows the best fit of Eq. 2 to scattering from samples prepared in 2.7 M ammonium sulfate and aged 2 days. The scattering from randomly packed clusters does not contain a high-q peak, indicating that a degree of order is an essential feature of monomer packing within the network.

The presence of ordered clusters indicates that a nucleation-and-growth type mechanism is likely responsible for their formation as opposed to spinodal decomposition or the fractal-like growth commonly seen in colloid gels. TEM clearly shows preferential growth along two directions and a distinct lack of growth in the third. Proteins are known to bind improperly to growing crystals, which poisons growth along certain crystallographic planes and thus stunts growth relative to others (35). We propose that a similar self-poisoning mechanism is responsible for the relatively monodisperse clusters observed in ovalbumin precipitates. In addition to limiting growth along certain crystal axes, self-poisoning events may introduce defects in the growing crystal that could in turn promote branching and bicontinuous network formation.

Considering that ordered packings describe the high-q data better than disordered packings, we use Scherrer's Law, Eq. 3, to estimate the size of the crystal-like domains. The domain size, averaged over the two- and three-day data, is 9 ± 2 nm, which agrees with the cluster thickness measured via TEM reasonably well. Note that this size also agrees with the SANS fitting in the mid-q regime and the smallest dimension of the best-fit crystalline clusters. It was difficult to determine the cluster size by fitting SANS data to idealized models and we found that the measured thickness depended on the model. The estimated thickness was 17 nm for cylinders and ellipsoids, 13 nm for sheets, 7-11 nm from Scherrer analysis of the high-q peak, and 9-15 nm from TEM measurements. The difference between the largest and smallest estimate is 10 nm, which is approximately 2-4 ovalbumin molecules. SANS shows that the clusters are monodisperse as polydispersity would act to smear out both the primary minimum at $q = 0.5\ nm^{-1}$ and the maximum at $q = 0.8\ nm^{-1}$.

4 Conclusions

The structure of ovalbumin precipitate particles as measured by TEM, USANS, and SANS presented here is, to our knowledge, the first reported molecular-level structure of a salted-out protein precipitate. We found that ovalbumin packs into a bicontinuous structure with significant void space and that ovalbumin is highly ordered within the dense network. Our results illustrate that salted-out protein precipitates can and do give rise to highly ordered structures in agreement with the experimental observation that precipitate phases can be used to seed protein crystal growth (15, 16). Our results also suggest that a careful revision of the

thermodynamic treatment of protein salting-out is necessary as current treatments implicitly assume that the dense phase is amorphous (36, 37).

5 Author Contributions

NJW, SIS, and AML initiated the project, helped to design and interpret experiments and reviewed the manuscript. DGG helped to design experiments, ran neutron scattering measurements, performed data reduction, generated scattering models, fit data and wrote the manuscript. SM prepared and ran TEM and TEM tomography experiments.

6 Acknowledgements

The authors thank Scott Jacobs for his help in creating the surface rendering of the tomographic data. This work was supported in part by the National Science Foundation under award number DMR-1209424. NJW acknowledges additional support from cooperative agreements 70NANB10H256 and 70NANB7H6178 from the National Institute of Standards and Technology (NIST), U.S. Department of Commerce. We acknowledge the support of the National Institute of Standards and Technology, U.S. Department of Commerce, in providing neutron research facilities used in this work. The research at ORNL's High Flux Isotope Reactor was sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, US Department of Energy. The microscopy was performed at was performed at the Delaware Biotechnology Institute Bioimaging Facility, which is supported in part by NIH grants P30 GM103519 and P20 GM103446 from the IDeA program of the National Institute of General Medical Sciences. The SAXS data in the Supporting Information were collected at the Cornell High Energy Synchrotron Source (CHESS) on the G-1 beamline. CHESS is supported by the NSF and NIH/NIGMS via NSF award DMR-1332208, and the MacCHESS resource is supported by NIH/NIGMS award GM-103485.

References

1. McPherson, A. 2004. Introduction to protein crystallization. Methods. 34: 254–265.

2. Johnston, K.P., J.A. Maynard, T.M. Truskett, A.U. Borwankar, M.A. Miller, B.K. Wilson, A.K. Dinin, T.A. Khan, and K.J. Kaczorowski. 2012. Concentrated dispersions of equilibrium protein nanoclusters that reversibly dissociate into active monomers. ACS Nano. 6: 1357–1369.

3. Dumetz, A.C., A.M. Chockla, E.W. Kaler, and A.M. Lenhoff. 2008. Protein phase behavior in aqueous solutions: crystallization, liquid-liquid phase separation, gels, and aggregates. Biophys. J. 94: 570–583.

4. Muschol, M., and F. Rosenberger. 1997. Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization. J. Chem. Phys. 107: 1953–1962.

5. Benedek, G.B., J.I. Clark, E.N. Serrallach, C.Y. Young, L. Mengel, T. Sauke, A. Bagg, and K. Benedek. 1979. Light scattering and reversible cataracts in the calf and human lens. Philos. Trans. R. Soc. London. Ser. A. Math. Phys. Sci. 293: 329–340.

6. Dumetz, A.C., A.M. Chockla, E.W. Kaler, and A.M. Lenhoff. 2009. Comparative effects of salt, organic, and polymer precipitants on protein phase behavior and implications for vapor diffusion. Cryst. Growth Des. 9: 682–691.

7. Eberle, A.P.R., N.J. Wagner, and R. Castañeda-Priego. 2011. Dynamical arrest transition in nanoparticle dispersions with short-range interactions. Phys. Rev. Lett. 106: 105704.

8. Foffi, G., G.D. McCullagh, A. Lawlor, E. Zaccarelli, K.A. Dawson, F. Sciortino, P. Tartaglia, D. Pini, and G. Stell. 2002. Phase equilibria and glass transition in colloidal systems with short-ranged attractive interactions: application to protein crystallization. Phys. Rev. E. 65: 031407.

9. Cheng, Y.-C., R.F. Lobo, S.I. Sandler, and A.M. Lenhoff. 2006. Kinetics and equilibria of lysozyme precipitation and crystallization in concentrated ammonium sulfate solutions. Biotechnol. Bioeng. 94: 177–188.

10. Heinen, M., F. Zanini, F. Roosen-Runge, D. Fedunová, F. Zhang, M. Hennig, T. Seydel, R. Schweins, M. Sztucki, M. Antalík, F. Schreiber, and G. Nägele. 2012. Viscosity and diffusion: crowding and salt effects in protein solutions. Soft Matter. 8: 1404–1419.

11. Liu, Y., L. Porcar, J. Chen, W.-R. Chen, P. Falus, A. Faraone, E. Fratini, K. Hong, and P. Baglioni. 2011. Lysozyme protein solution with an intermediate range order structure. J. Phys. Chem. B. 115: 7238–7247.

12. Ianeselli, L., F. Zhang, M.W.A. Skoda, R.M.J. Jacobs, R.A. Martin, S. Callow, S. Prévost, and F. Schreiber. 2010. Protein-protein interactions in ovalbumin solutions studied by small-angle scattering: effect of ionic strength and the chemical nature of cations. J. Phys. Chem. B. 114: 3776–3783.

13. Fortini, A., E. Sanz, and M. Dijkstra. 2008. Crystallization and gelation in colloidal systems with short-ranged attractive interactions. Phys. Rev. E. 78: 041402.

14. Khan, S.J., C.M. Sorensen, and A. Chakrabarti. 2009. Kinetics and morphology of cluster growth in a model of short-range attractive colloids. J. Chem. Phys. : 194908.

15. Ng, J.D., B. Lorber, J. Witz, A. Théobald-Dietrich, D. Kern, and R. Giegé. 1996. The crystallization of biological macromolecules from precipitates: evidence for Ostwald ripening. J. Cryst. Growth. 168: 50–62.

16. Bergfors, T. 2003. Seeds to crystals. J. Struct. Biol. 142: 66–76.

17. Judge, R.A., M.R. Johns, and E.T. White. 1995. Protein purification by bulk crystallization: the recovery of ovalbumin. Biotechnol. Bioeng. 48: 316–323.

18. Glasoe, P.K., and F.A. Long. 1960. Use of glass electrodes to measure acidities in deuterium oxide. J. Phys. Chem. 64: 188–190.

19. Sober, H.A. 1968. Handbook of biochemistry. Selected data for molecular biology. Cleveland, OH: The Chemical Rubber Company.

20. Kremer, J.R., D.N. Mastronarde, and J.R. McIntosh. 1996. Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. 116: 71–76.

21. Kline, S.R. 2006. Reduction and analysis of SANS and USANS data using IGOR Pro. J. Appl. Crystallogr. 39: 895–900.

22. Stein, P.E., A.G.W. Leslie, J.T. Finch, and R.W. Carrell. 1991. Crystal structure of uncleaved ovalbumin at 1.95 A resolution. J. Mol. Biol. 221: 941–959.

23. Svergun, D.I., S. Richard, M.H.J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proc. Natl. Acad. Sci. 95: 2267–2272.

24. Pipich, V., M. Balz, S.E. Wolf, W. Tremel, and D. Schwahn. 2008. Nucleation and growth of $CaCO_3$ mediated by the egg-white protein ovalbumin: a time-resolved in situ study using small-angle neutron scattering. J. Am. Chem. Soc. 130: 6879–6892.

25. Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control. 19: 716–723.

26. Donev, A., S. Torquato, and F.H. Stillinger. 2005. Neighbor list collision-driven molecular dynamics simulation for nonspherical hard particles. I. Algorithmic details. J. Comput. Phys. 202: 737–764.

27. Donev, A., S. Torquato, and F.H. Stillinger. 2005. Neighbor list collision-driven molecular dynamics simulation for nonspherical hard particles. II. Applications to ellipses and ellipsoids. J. Comput. Phys. 202: 765–793.

28. Scherrer, P. 1918. Bestimmung der Größe und der inneren Struktur von Kolloidteilchen mittels Röntgenstrahlen. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Math. Klasse. 1918: 98–100.

29. Langford, J.I., and A.J.C. Wilson. 1978. Scherrer after sixty years: a survey and some new results in the determination of crystallite size. J. Appl. Crystallogr. 11: 102–113.

30. Lopez-Barron, C.R., and N.J. Wagner. 2011. Solvent isotope effect on the microstructure and rheology of cationic worm-like micelles near the isotropic-nematic transition. Soft Matter. 7: 10856–10863.

31. Paesani, F. 2011. Hydrogen bond dynamics in heavy water studied with quantum dynamical simulations. Phys. Chem. Chem. Phys. 13: 19865–19875.

32. Zia, R.N., B.J. Landrum, and W.B. Russel. 2014. A micro-mechanical study of coarsening and rheology of colloidal gels: cage building, cage hopping, and Smoluchowski's ratchet. J. Rheol. 58: 1121–1157.

33. Nallet, F., R. Laversanne, and D. Roux. 1993. Modelling x-ray or neutron scattering spectra of lyotropic lamellar phases: interplay between form and structure factors. J. Phys. II. 3: 487–502.

34. Barker, J.G., and J.S. Pedersen. 1995. Instrumental smearing effects in radially symmetric small-angle neutron scattering by numerical and analytical methods. J. Appl. Crystallogr. 28: 105–114.

35. Asthagiri, D., A.M. Lenhoff, and D.T. Gallagher. 2000. Role of competitive interactions in growth rate trends of subtilisin s88 crystals. J. Cryst. Growth. 212: 543–554.

36. Timasheff, S.N., and T. Arakawa. 1988. Mechanism of protein precipitation and stabilization by co-solvents. J. Cryst. Growth. 90: 39–46.

37. Chiew, Y.C., D. Kuehner, H.W. Blanch, and J.M. Prausnitz. 1995. Molecular thermodynamics for salt-induced protein precipitation. AIChE J. 41: 2150–2159.

Table 1. Best-fit parameters from core-shell model fit to USANS data. Error bars are the standard deviations of the fitted parameters as reported by the IGOR fitting package.

Sample	Core radius (nm)	Shell thickness (nm)	Polydispersity
2.4 M	722 ± 184	231 ± 64	0.65 ± 0.15
2.5 M	1864 ± 28	750 ± 21	0.1 (fixed)
2.8 M	1834 ± 51	235 ± 43	0.1 (fixed)

Table 2. Best-fit parameters for mid-q data from lamellae, cylinder and triaxial ellipsoid models. Error bars are the standard deviation of the fitted and measured parameters.

Lamellae	Thickness (nm)		
	13.3 ± 0.6		
Cylinder	Diameter (nm)	Length (nm)	
	17.70 ± 0.07	29.39 ± 0.12	
Triaxial ellipsoid	Diameter A (nm)	Diameter B (nm)	Length (nm)
	17.12 ± 0.03	20.67± 0.10	47.83 ± 0.97

Figure captions

Figure 1. Ovalbumin phase behavior. a) Aggregation boundaries for ovalbumin in H_2O and D_2O. Lines are guides to the eye. The nomenclature matches that in Dumetz et al. (3). – ■ –: 1st aggregation boundary in H_2O; -▲-: 2nd aggregation boundary in H_2O; ·△·: 2nd aggregation boundary in D_2O. b) Gel beads formed between the 1st and 2nd aggregation boundaries. c) Dense phase formed by immediate demixing beyond the 2nd aggregation line.

Figure 2. TEM images of ovalbumin precipitates formed at 2.4 M ammonium sulfate in H_2O. a) Low-resolution image of particles, b) the core and shell regions of the particle, c) the shell region, d) surface rendering of a TEM tomogram of the shell region.

Figure 3. USANS intensities for ovalbumin samples prepared at 2.4, 2.5 and 2.8 M ammonium sulfate. The solid line is the best fit to a core-shell model for each case. Profiles have been shifted vertically for clarity.

Figure 4. SANS intensities after background subtraction for a) precipitates prepared in 2.5 M ammonium sulfate and aged between 20 minutes and 28 days, and b) samples prepared in 2.4-2.9 M ammonium sulfate and aged 2 days. In a) the inset shows the development of the high-q peak (arrow) from precipitates to crystals. In b) the labels indicate 1) low-q peaks, 2) mid-q upturn, 3) high-q peak.

Figure 5. Fits of structural models to scattering data, at 2.7 M ammonium sulfate and 2 days of aging, in the high-q region. a) Fit of crystalline cluster model (Eq. 1, red line). b) fit of random cluster model (Eq. 2, gold line).

Supporting Information

Local crystalline structure in an amorphous protein dense phase

D. G. Greene, S. Modla, N. J. Wagner, S. I. Sandler, A. M. Lenhoff

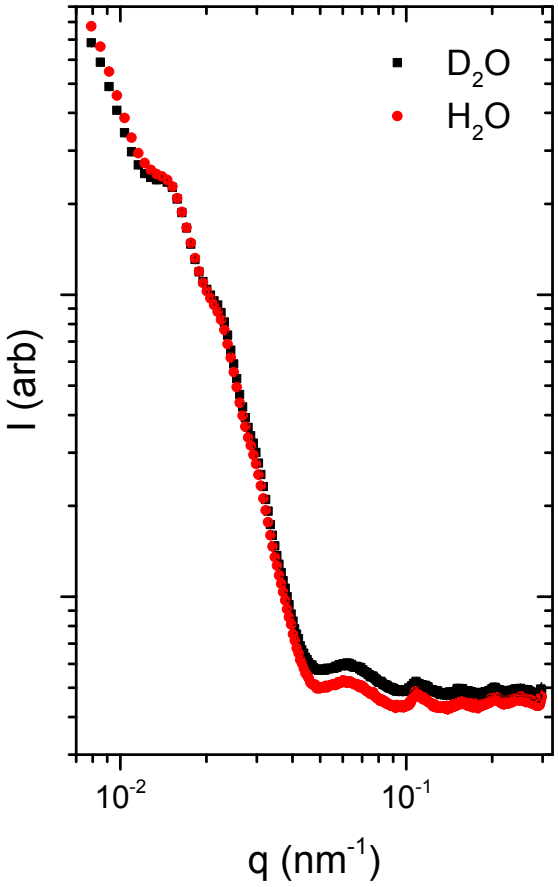


Figure S1. SAXS from ovalbumin precipitates prepared at 2.4 M ammonium sulfate in D_2O and 2.0 M ammonium sulfate in H_2O.

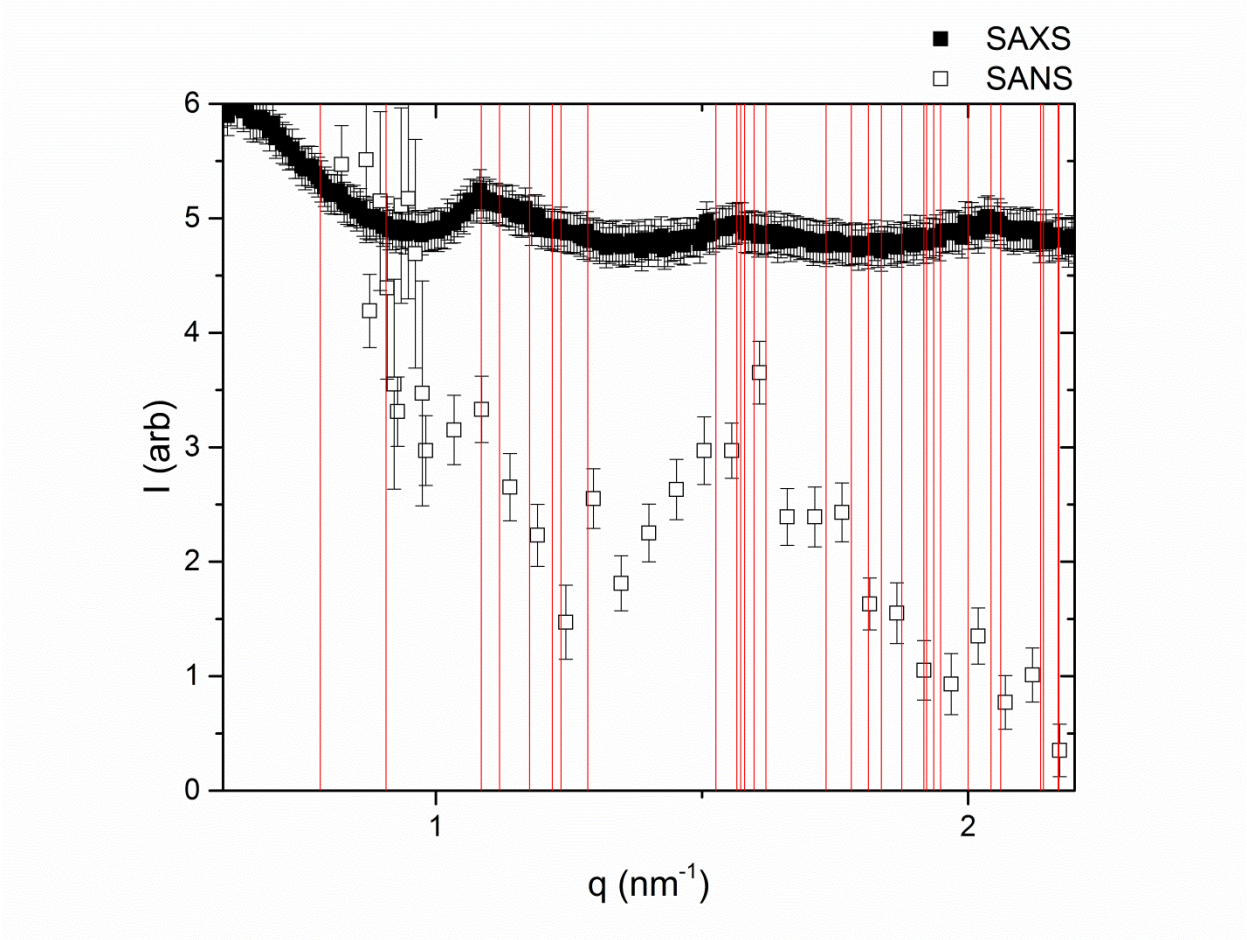


Figure S2. Comparison between SAXS (filled) and SANS (open) from ovalbumin precipitates in the high-q region. The SAXS data are those from Figure S1 and the SANS data those from Figure 5; as shown in Figure 4b and Figure S1, there is little variation of the spectra with the solution conditions used, so the comparison of these two data sets is reasonable. The red lines indicate the locations of the allowed Bragg reflections from the ovalbumin unit cell.

Table S1. Investigated dimensions of crystalline clusters. A, B, and C are the principal crystallographic axes.

Cluster #	# Unit cells in A	# Unit cells in B	# Unit cells in C	Approx. dimension in A (nm)	Approx. dimension in B (nm)	Approx. dimension in C (nm)
1	10	2	2	62.9	16.9	14.3
2	2	2	10	12.6	16.9	71.5
3	2	2	4	12.6	16.9	28.6
4	2	2	5	12.6	16.9	35.8
5	2	2	6	12.6	16.9	42.9
6	2	2	9	12.6	16.9	64.4
7	3	4	2	18.9	33.9	14.3
8	3	5	2	18.9	42.4	14.3
9	3	6	2	18.9	50.8	14.3
10	3	8	2	18.9	67.8	14.3
11	3	9	2	18.9	76.2	14.3
12	4	2	2	25.2	16.9	14.3
13	4	4	2	25.2	33.9	14.3
14	4	4	3	25.2	33.9	21.5
15	4	9	1	25.2	76.2	7.2
16	5	2	2	31.5	16.9	14.3
17	5	5	2	31.5	42.4	14.3
18	6	2	2	37.7	16.9	14.3
19	7	2	2	44.0	16.9	14.3
20	8	2	2	50.3	16.9	14.3
21	9	2	2	56.6	16.9	14.3

Expressions for particle form factors

Core-shell spheres:

$$P(q) = \frac{\alpha}{V_c + V_s}\left(\frac{3V_c[\rho_c - \rho_s]j_1(qr_c)}{qr_c} + \frac{3[V_c + V_s][\rho_s - \rho_{solv}]j_1(qr_{c+}qt)}{qr_{c+}qt}\right)^2 \quad \text{(S1)}$$

$j_1(x)$ is the first spherical Bessel function, the subscript s refers to the entire particle, the subscripts c, s, and solv refer to the core, shell and solvent respectively, ρ_i is the scattering length density, r_c is the core radius, t is the shell thickness, α is a scale factor, and V is the volume.

Cylinders:

$$P(q) = \frac{\alpha}{V_{cyl}}\int_0^{\pi/2}\left(2\left[\rho_{cyl} - \rho_{solv}\right]V_{cyl}j_0(qH\cos\alpha)\frac{j_1(qr\sin\alpha)}{qr\sin\alpha}\right)^2 \sin\alpha\, d\alpha \quad \text{(S2)}$$

V_{cyl} is the cylinder volume, ρ_{cyl} and ρ_{solv} are the SLD's of the cylinder and solvent respectively, H is the cylinder height, r is the cylinder radius, $j_i(x)$ is the first spherical Bessel function and α is a scale factor.

Triaxial ellipsoids:

$$P(q) = \frac{\alpha}{V_{el}}\int_{0,0}^{1,1} f^2\left(q[a^2\cos^2\left(\frac{\pi x}{2}\right) + b\sin^2\left(\frac{\pi x}{2}\right)(1-y^2) + c^2y^2]^{1/2}\right)dxdy \quad \text{(S3a)}$$

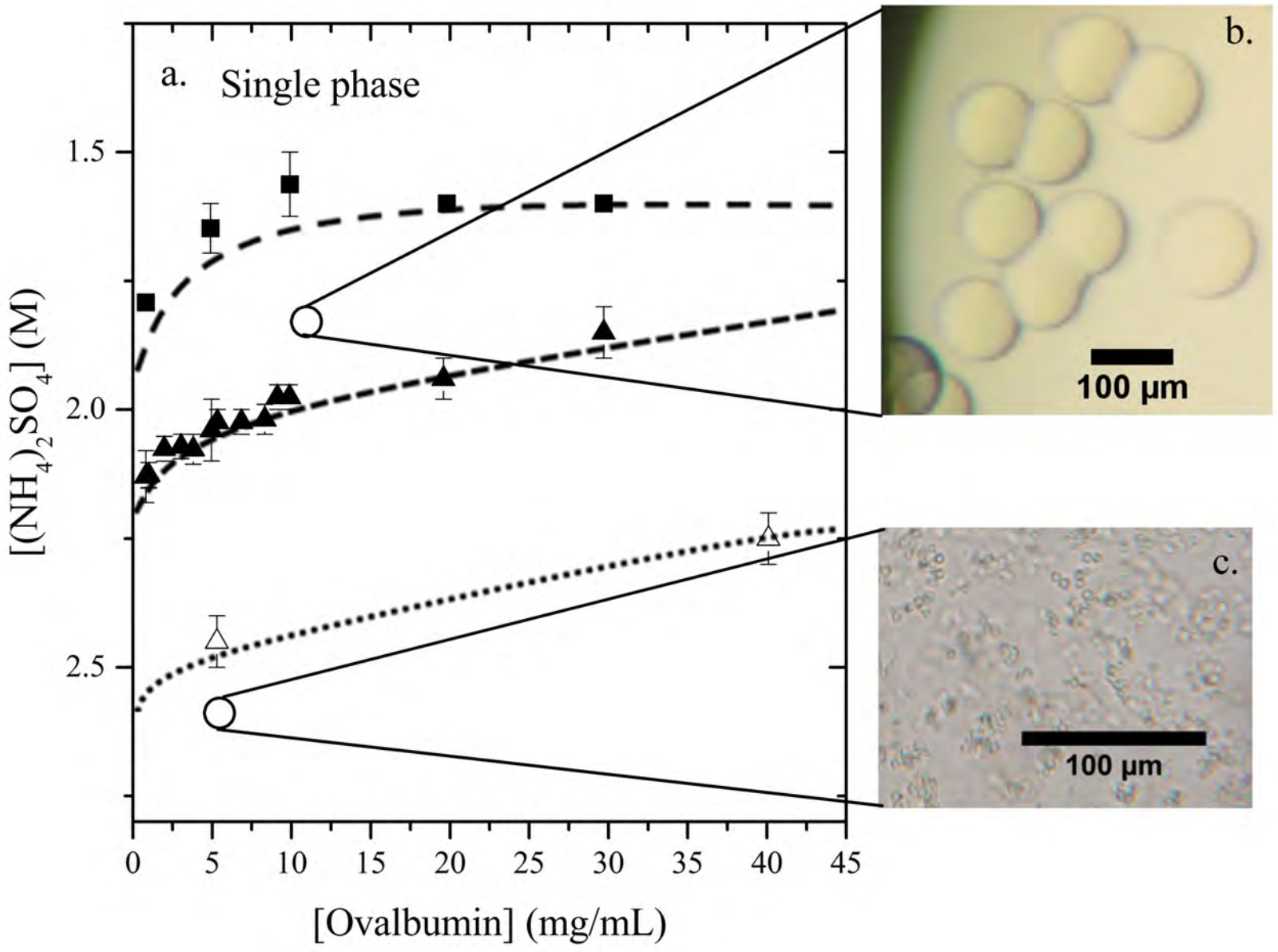
$$f^2(u) = 9\left(\frac{\sin u - u\cos u}{u^3}\right)^2 \quad \text{(S3b)}$$

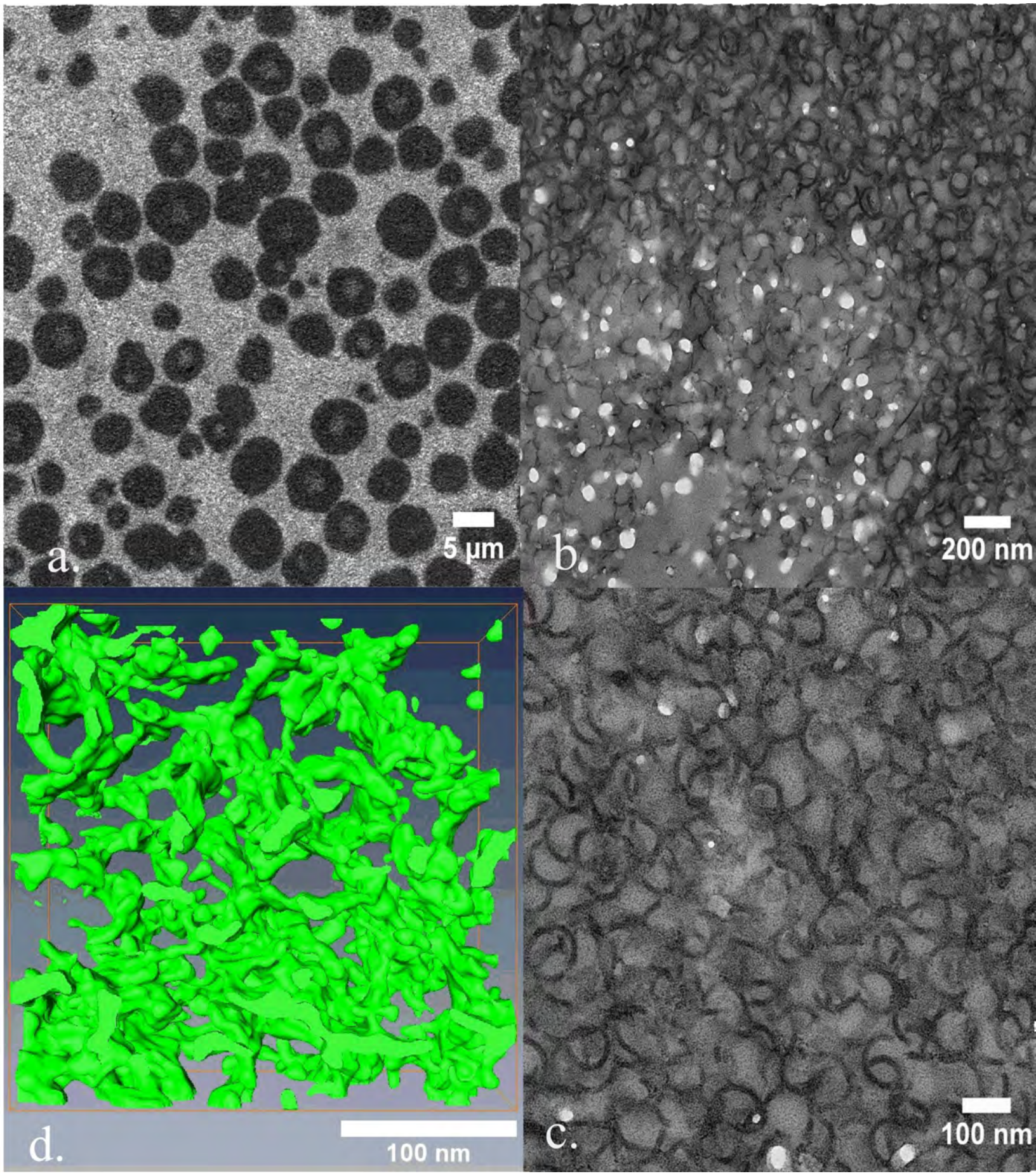
V_{el} is the ellipsoid volume, α is a scale factor, and a, b and c are the three semi-axes of the ellipsoid.

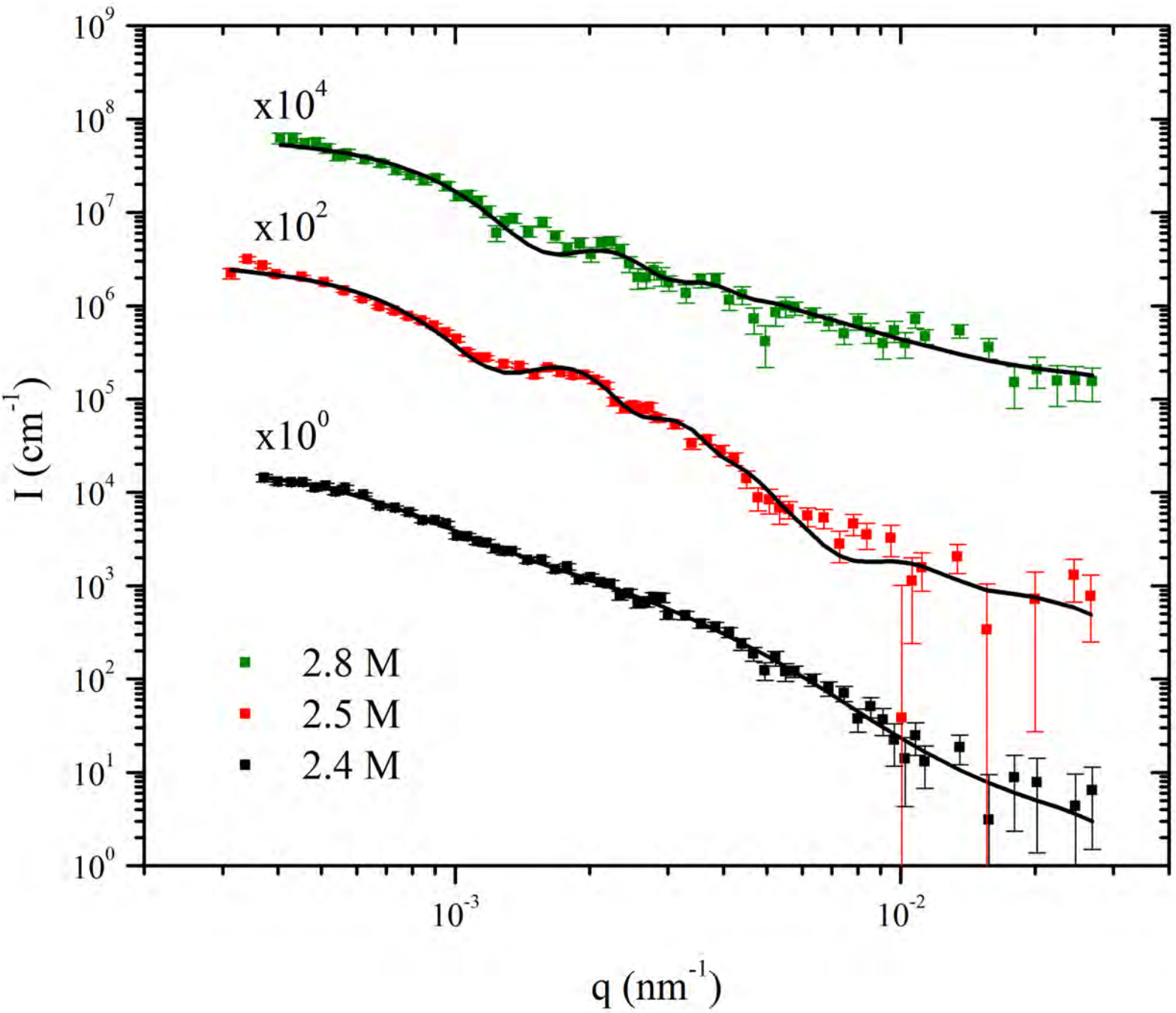
Lamellar sheets:

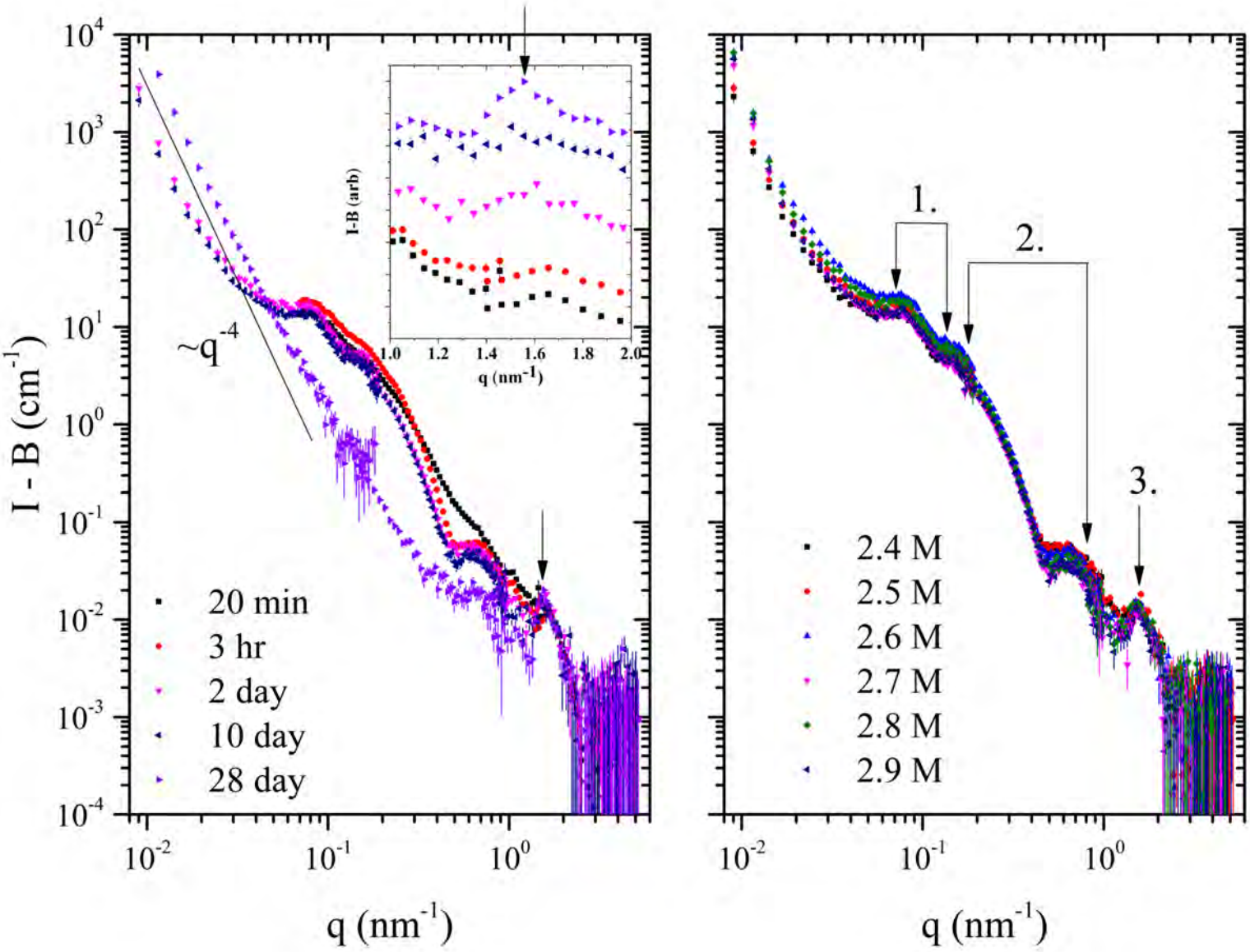
$$P(q) = \alpha\frac{2(\rho_{lam} - \rho_{solv})^2}{q^2t}\left[1 - \cos(qt)e^{-q^2\sigma^2/2}\right] \quad \text{(S4)}$$

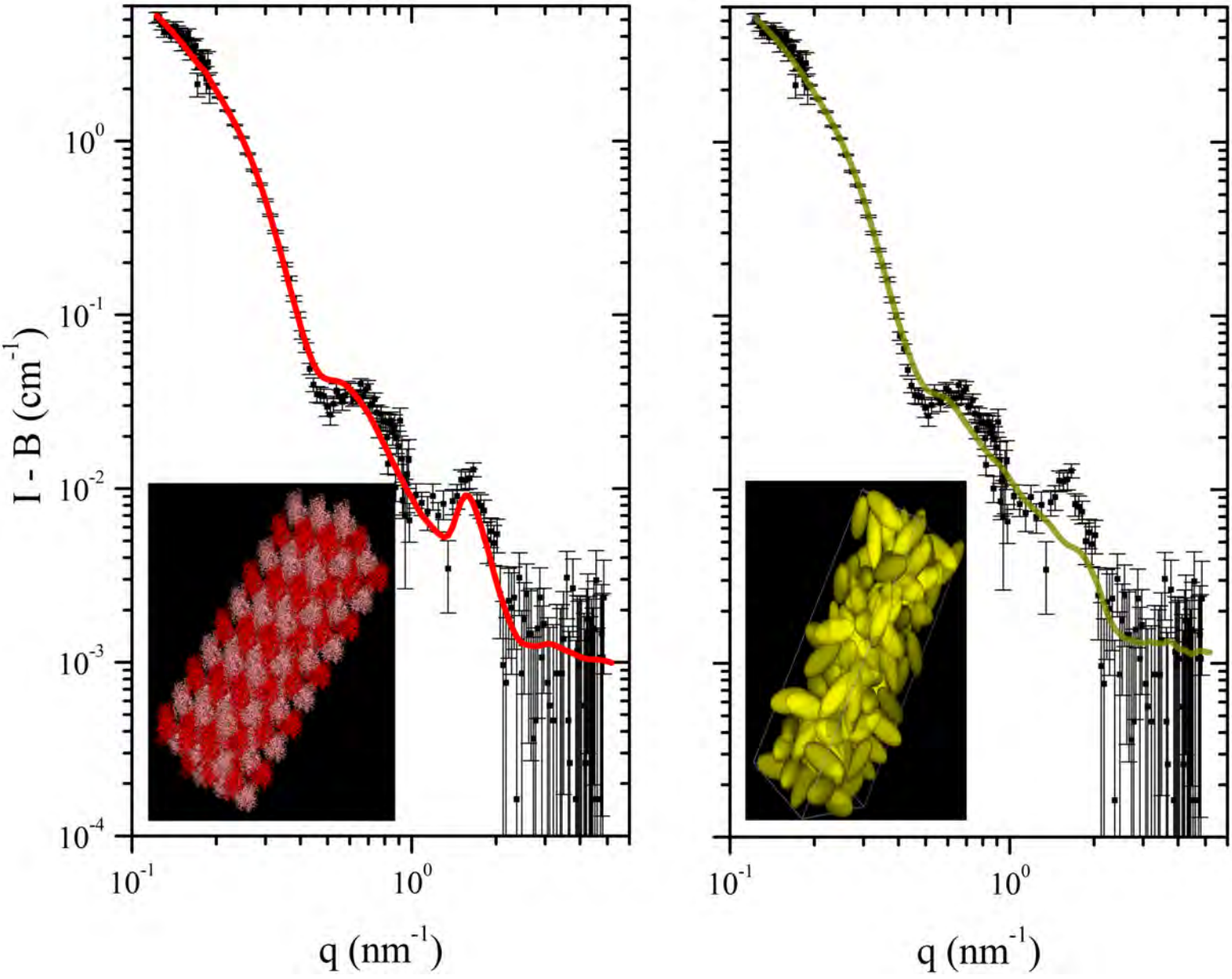
α is a scale factor, t is the lamellar thickness, ρ_{lam} and ρ_{solv} are the SLDs of the lamellar sheet and solvent respectively, σ is the standard deviation in the lamellar thickness distribution.


a.
Single phase
b.
100 µm
c.
100 µm
1.5
2.0
2.5
0
5
10
15
20
25
30
35
40
45
[(NH4)2SO4] (M)
[Ovalbumin] (mg/mL)


a.
5 µm
b.
200 nm
d.
100 nm
c.
100 nm


I (cm-1)
q (nm-1)
x10^4
x10^2
x10^0
2.8 M
2.5 M
2.4 M
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
10^-3
10^-2


I - B (cm^{-1})
q (nm^{-1})
~q^{-4}
20 min
3 hr
2 day
10 day
28 day
I-B (arb)
q (nm^{-1})
1.
2.
3.
2.4 M
2.5 M
2.6 M
2.7 M
2.8 M
2.9 M


I - B (cm^{-1})
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-1}
10^{0}
q (nm^{-1})
10^{-1}
10^{0}
q (nm^{-1})