Article

# Lack of Dependence of the Sizes of the Mesoscopic Protein Clusters on Electrostatics

Maria A. Vorontsova,[1] Ho Yin Chan,[2] Vassiliy Lubchenko,[2,3] and Peter G. Vekilov[1,3,*]
[1]Department of Chemical and Biomolecular Engineering, [2]Department of Physics, and [3]Department of Chemistry, University of Houston, Houston, Texas

ABSTRACT Protein-rich clusters of steady submicron size and narrow size distribution exist in protein solutions in apparent violation of the classical laws of phase equilibrium. Even though they contain a minor fraction of the total protein, evidence suggests that they may serve as essential precursors for the nucleation of ordered solids such as crystals, sickle-cell hemoglobin polymers, and amyloid fibrils. The cluster formation mechanism remains elusive. We use the highly basic protein lysozyme at nearly neutral and lower pH as a model and explore the response of the cluster population to the electrostatic forces, which govern numerous biophysical phenomena, including crystallization and fibrillization. We tune the strength of intermolecular electrostatic forces by varying the solution ionic strength *I* and pH and find that despite the weaker repulsion at higher *I* and pH, the cluster size remains constant. Cluster responses to the presence of urea and ethanol demonstrate that cluster formation is controlled by hydrophobic interactions between the peptide backbones, exposed to the solvent after partial protein unfolding that may lead to transient protein oligomers. These findings reveal that the mechanism of the mesoscopic clusters is fundamentally different from those underlying the two main classes of ordered protein solid phases, crystals and amyloid fibrils, and partial unfolding of the protein chain may play a significant role.

## INTRODUCTION

Protein solutions exhibit at least three distinct classes of compact aggregates that are often referred to as "clusters". Small clusters containing 2–10 molecules have been observed in solutions of lysozyme, insulin, and a monoclonal human antibody (1-4) and are likely present in solutions of other proteins under conditions conducive of mild intermolecular repulsion. These aggregates hold a significant fraction of the total soluble protein, and their average size strongly increases with the protein concentration (2-4). Importantly, these clusters do not represent permanent structures (5) and are often viewed as dynamic formations with intermediate-range order and a lifetime of ~25 ns (2,6).

Clusters of the second class contain ~1000 protein molecules (7-9). They only exist at conditions at which short-range attraction and long-range repulsion are delicately balanced: near the isoelectric point of the respective proteins and in the presence of a finely tuned concentration of a crowding agent (7), or in the vicinity of charge inversion induced by bound multivalent cations (8,9). Similarly to the small clusters, they hold a significant fraction of the total soluble protein and the cluster sizes increase with the protein concentration (6,10). Interestingly, while the small clusters strongly increase the solution viscosity (4,11), the formation of the larger clusters reduces viscosity (7). In view of their effects on the solution's rheology, clusters of these two classes are scrutinized with the goal of increasing the fluidity of concentrated solutions of proteins with medical applications, e.g., monoclonal antibodies (low viscosity is deemed essential for enhanced production, purification, and delivery of these drugs (4,7)).

Clusters of the third class have been called "mesoscopic" and demonstrated in solutions of numerous proteins at various pH values ionicities, temperatures, and compositions (12-16). They are distinct in several ways from the former two. Their diameters vary from ~100 nm for the relatively small lysozyme (14) to several hundred nanometers for larger proteins (15,17). These clusters are likely liquid (the liquid state of clusters has been evidenced in solutions of lysozyme (18), glucose isomerase (18), lumazine synthase (12), three human hemoglobin variants (19), and several other proteins (17)), and are stable for extended periods (14). Assuming that the protein concentration in the clusters is ~500 mg $mL^{-1}$, similar to that in the dense protein liquid existing at similar conditions, each of them contains $10^5$–$10^6$ protein molecules (13,15,17,19). This number is orders-of-magnitude greater than for clusters of classes I and II. The free energy cost of high protein concentration in the clusters was evaluated for lysozyme and hemoglobin as ~10 $k_BT$ ($k_B$, Boltzmann constant; $T$, temperature) by integrating the concentration dependence of the osmotic compressibility (determined by static light scattering) of homogeneous solutions with concentrations similar to that in the clusters (13,20). The Boltzmann relation predicts that owing to this free energy excess, the mesoscopic

*Submitted July 9, 2015, and accepted for publication September 21, 2015.*
*Correspondence: vekilov@uh.edu
Editor: Rohit Pappu.

0006-3495/15/11/1959/10
http://dx.doi.org/10.1016/j.bpj.2015.09.025

clusters would hold only $10^{-5}$–$10^{-3}$ of the total soluble protein. This prediction is borne out by experiment, demonstrating that the clusters are in approximate equilibrium with the solution (14,20); an early study had claimed that the clusters are nonequilibrium formations (13). Experimental determinations of the fraction of the solution volume occupied by the cluster population yield values in the range $10^{-7}$–$10^{-3}$ (13-15,17,19). The low fraction of protein held in the mesoscopic clusters is in stark contrast with the clusters of the other two classes. In further contrast, the sizes of the mesoscopic clusters do not depend on the protein concentration in the host solution (14,15,21).

Because of their low population volume, the mesoscopic clusters do not affect the bulk solution properties. They are of interest because they may hold the key to understanding and control of ordered protein aggregation: evidence suggests that in many cases the clusters present essential sites for the nucleation of ordered solids of both folded proteins, such as crystals (17,18,22,23) and sickle-cell hemoglobin polymers (24), and partially misfolded chains that form amyloid fibrils (25-27). Furthermore, the mesoscopic clusters may relate to the non-membrane-bound compartments (nucleoli, centrosomes, Cajal bodies, etc.) found to consist of dense protein/RNA liquid in several organisms (28-31).

The mechanisms of formation of the clusters of classes I and II appear to be well understood in the context of colloid clustering theories (32-34). These clusters represent a balance among short-range attractions (due to depletion agents (7,10), ion bridges (8), or shape complementarity (4)), which cause condensation, and long-range screened Coulomb repulsion, which constrains the size of the condensate domains (4,7,8,10,35). In this approach, the larger size of the class II clusters appears as the result of the nearly neutral charge of the protein molecules at the conditions of their synthesis. Furthermore, for clusters of both classes I and II, tuning the strength of Coulomb repulsion through the solution ionicity becomes an essential tool to control the average cluster size, degree of clustering, and the associated bulk solution properties (4,7,8,10).

There is significantly less clarity on the nature and mechanisms of the mesoscopic clusters. Their small population volume and stable mesoscopic size challenge our understanding of phase ordering. Although the clusters are likely liquid (12,17,19), their region of existence in the protein solution phase diagram is away from the conditions of liquid-liquid coexistence (12,13,19). These clusters are much larger than what the colloid scenario for clusters of classes I and II predicts (36). An entirely distinct approach proposes that the mesoscopic clusters consist of a concentrated mixture of transient protein oligomers and monomers (13). The clusters result from interplay of monomer influx, oligomer formation, and subsequent oligomer outflow and decay. Similar reaction-diffusion schemes are derived within the hydrodynamic description by connecting the rate of approach to equilibrium and the degree of deviation from equilibrium (37). By solving two coupled reaction-diffusion equations, in which both diffusivities and reaction constants are explicitly present, the cluster size $R_2$ emerges as a function of the oligomer diffusivity $D_{\text{oligomer}}$ and decay rate constant $k_{\text{oligomer}}$ (13):

$$R_2 = \left(D_{\text{oligomer}}/k_{\text{oligomer}}\right)^{1/2}. \quad (1)$$

While the oligomer mechanism appears to fit the available data on the mesoscopic clusters better than the colloid theory, the applicability of either mechanism to the latter cluster class has never been systematically tested. In light of the important role that Coulomb forces play for the clusters of classes I and II, here we probe their effects on the properties of the mesoscopic clusters as a test of the colloid clustering scenario.

From a fundamental perspective, Coulomb forces determine protein three-dimensional structure (38,39), substrate binding (40,41), enzyme activation (42,43), signal transduction (44), etc. Importantly, Coulomb forces govern two major classes of protein aggregation: amyloid fibrillation (45-47) and crystallization (48,49). Hence, understanding of their role in cluster formation will highlight the similarities and differences between the formation mechanisms of the mesoscopic clusters and those two major classes of protein aggregates. Furthermore, as electrostatic forces can bind protein oligomers (50), Coulomb interactions could contribute to the oligomer scenario. Thus, quantifying the effect of Coulomb forces on the mesoscopic clusters is crucial for establishing how the clusters form. In turn, these insights may suggest strategies to control cluster populations and, in view of the clusters' role in nucleation, the formation of ordered protein solids. Finally, the mesoscopic clusters are sometimes misidentified as clusters of class II (7-9). Thus, establishing the mechanism of formation of the mesoscopic clusters and the respective role of the Coulomb forces will allow clear distinctions between these three cluster classes and elucidate complex clustering behaviors in protein solutions.

## MATERIALS AND METHODS

To quantify the characteristics of the cluster populations, we employ dynamic light scattering (DLS) and Brownian microscopy (BM) (19,51). The application of DLS to characterization of the cluster populations is discussed in detail in Pan et al. (19) and Li et al. (51). BM, illustrated later in Fig. 1, *a–c*, is a relatively new method that serves to detect and monitor the motions of protein-rich clusters as in Fig. 1 *b*; see the Supporting Material for details. Fig. 1 *d* displays an example of a cluster size distribution, determined by BM. Both DLS and BM rely on light scattered at wavevectors of order $\mu m^{-1}$ and probe length scales in the range $10^{-3}$–10 $\mu$m. The Rayleigh law, according to which the scattered intensity scales as the sixth power of the scatterers' sizes, makes these two techniques particularly well suited to study the mesoscopic clusters, which are 50–100× larger than the monomers, but are present at very low concentration.

Small-angle x-ray and neutron scattering (SAXS and SANS, respectively) have been employed to characterize clusters of classes I and II (1,2,4,6,8,9). These methods record intensity scattered at wavevectors of order $Å^{-1}$ and probe length scales of the order of nanometers. They identify

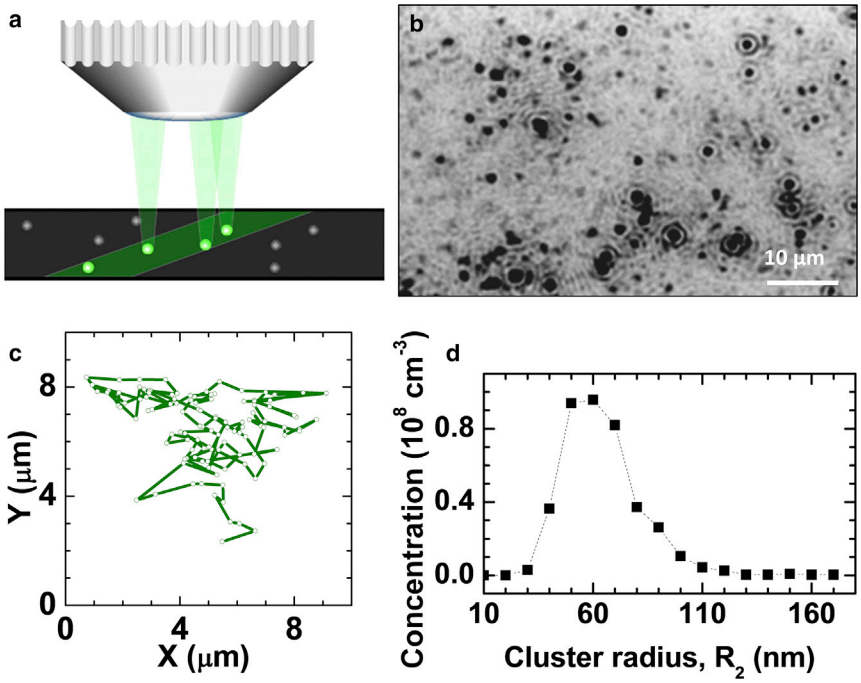


FIGURE 1 Cluster characterization by Brownian microscopy. (*a*) Schematic of the BM setup. A green laser illuminates a thin solution layer. The light scattered by particles in the solution is collected by a microscope lens. (*b*) A representative BM image shown as a negative. The observed volume is ~120 × 80 × 5 $\mu m^3$. The clusters are seen as black spots. (*c*) A typical cluster trajectory determined from a sequence of images. The cluster diffusivities and sizes are evaluated from such trajectories. (*d*) Distribution of cluster sizes, determined from trajectories such as the one in (*c*). Only clusters registered for longer than 1 s are considered. To see this figure in color, go online.

clusters owing to local ordering at the probed length scales within the clusters and only detect clusters that occupy a significant fraction of solution volume. This requirement also applies to nuclear spin echo, a method employed to probe the dynamic properties of small clusters (2,6). Thus, these three methods would be of limited utility in investigations of the mesoscopic clusters.

## RESULTS AND DISCUSSION

### Model system

Our model protein, lysozyme, has an isoelectric point at pH = 11.35 (52), which is one of the highest documented for any protein (53). As a result, even at the highest pH = 7.8 probed here, protonation of basic and acidic surface amino-acid groups leads to a significant +8 net charge on the lysozyme monomer (54); 17 positive and nine negative groups were identified at this pH in Chan et al. (50). At the lowest pH = 3.8 tested here, the net charge increases to +15 (54). This high net positive charge, illustrated in Fig. 2, amplifies the significance of the Coulomb forces for aggregation behaviors of lysozyme, and makes this protein a suitable model system for this work.

### Characterization of the intermolecular Coulomb forces

We tune the strength of the Coulomb interactions between lysozyme molecules in two ways: by varying the solution ionic strength $I$, which directly controls the Debye screening (55), and the solution pH, which determines the protein's charge. We characterize the intermolecular interactions in terms of three parameters: the second osmotic virial coefficient, $B_2$, an azimuthally and spatially averaged characteristic of the pairwise interaction potential (56) that is obtained from the slopes of Debye plots, determined by static light scattering, and presented in Fig. S1 in the Supporting Material; the diffusivity of protein monomers in dilute solutions, $D_1^{dilute}$; and the diffusivity in concentrated solutions, $D_1^{conc}$. The values $D_1^{dilute}$ and $D_1^{conc}$ were determined from the faster shoulder in the autocorrelation function of the intensity of light scattered

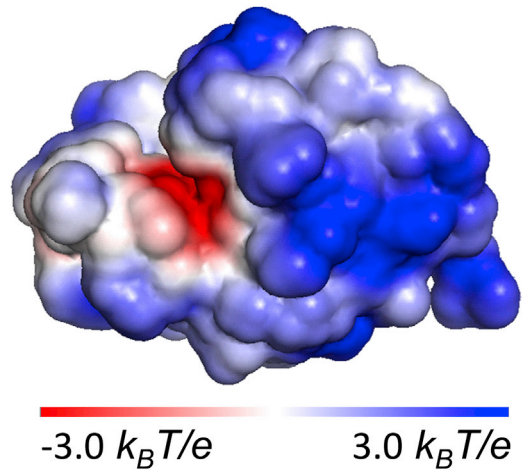


FIGURE 2 The distribution of electrostatic potential at the solvent-accessible surface of a lysozyme molecule at pH = 7.8; $k_B$, Boltzmann constant; $T$, temperature; and $e$, elementary charge. The Protein Data Bank (PBD) structure file PDB: 2VB1 was used. The protonation state of each acid or basic residue was evaluated at the chosen pH with PROPKA 3.0 (propka.org). The electrostatic map was computed online with the Adaptive Poisson-Boltzmann Solver and drawn using the software PyMOL (www.pymol.org). To see this figure in color, go online.

off the solution, seen in Fig. S2. The viscosities of all solvents used in this study were practically independent of the salt concentration and identity. Hence, $D_1^{\text{dilute}}$ and $D_1^{\text{conc}}$ trends in Fig. 3 *b* indicate the response of the protein dynamics to variations in intermolecular interactions. Note that while $B_2$ and $D_1^{\text{dilute}}$ account for the interactions at long intermolecular separations, $D_1^{\text{conc}}$ is weighted toward short separations.

The dependences of $B_2$, $D_1^{\text{dilute}}$, and $D_1^{\text{conc}}$ on the ionic strength $I$ are displayed in Fig. 3, *a* and *b*. We have varied $I$ from 3 to 333 mM by increasing the concentration of HEPES buffer or adding NaCl, KCl, or $(NH_4)_2SO_4$. The data sets corresponding to the four salts follow the same trend, implying that the intermolecular interactions depend on the solution electrostatics but not on the salt identity. The decreasing values $B_2$, $D_1^{\text{dilute}}$, and $D_1^{\text{conc}}$ at increasing $I$ are consistent with the expectation that the Debye screening due to free ions significantly weakens the Coulomb repulsion. At $I > 100$ mM, $B_2$ becomes lower than its value for hard spheres $B_2$(hs), while $D_1^{\text{dilute}}$ drops below the value of lysozyme's self-diffusivity $D_0$. Both observations indicate a switch to weak intermolecular attraction and imply that electrostatic repulsion is largely screened by the ions in the solution. The diffusivity $D_1^{\text{conc}}$ exhibits a stronger dependence on $I$ than $D_1^{\text{dilute}}$, indicating that Coulomb repulsion is more sensitive to electrolyte concentration at short than at long separations.

The dependences of $B_2$, $D_1^{\text{dilute}}$, and $D_1^{\text{conc}}$ on pH, displayed in Fig. 4, *a* and *b*, reveal that the values of the three parameters decrease as pH increases. This is expected: because higher pH values are closer to the isoelectric point, the protein net molecular charge should decrease, leading to weaker electrostatic repulsion. The decrease in $D_1^{\text{conc}}$ in Fig. 4 *b* is stronger than in $D_1^{\text{dilute}}$, again implying that Coulomb repulsion is more sensitive to the protein's charge at short than at long separations.

While the effects of pH on $B_2$, $D_1^{\text{dilute}}$, and $D_1^{\text{conc}}$ link directly to the decrease of the molecule's charge at higher pH, the correlation between $B_2$ and $I$ displayed in Fig. 3 *a* requires additional discussion. We employ a computational model following Chan et al. (50). We represent every protein molecule as a sphere with discrete charges as illustrated in Fig. S3. We consider interactions of pairs of molecules. Besides the Coulomb forces, we include an adjustable short-range contribution to the overall interaction that accounts for van der Waals attraction and steric repulsion; we assume that neither of the latter forces depends on $I$. Because of the molecules' net positive charge, the majority of pair configurations are repulsive; still, there are several attractive configurations, such as the one depicted in Fig. S3 *b*, in which a negative $Asp^{87}$ faces a positive $Arg^{45}$ while a neutral $His^{15}$ faces a positive $Arg^{68}$.

We sample all possible orientations of a pair of molecules using appropriate Boltzmann weights (50). The resulting angular-averaged potential of mean force (PMF) at $I = 13.3$ mM and pH = 7.8 is shown in Fig. 3 *c*. Three characteristics of this PMF relevant to the discussion of aggregation include the energy at contact, association barrier,

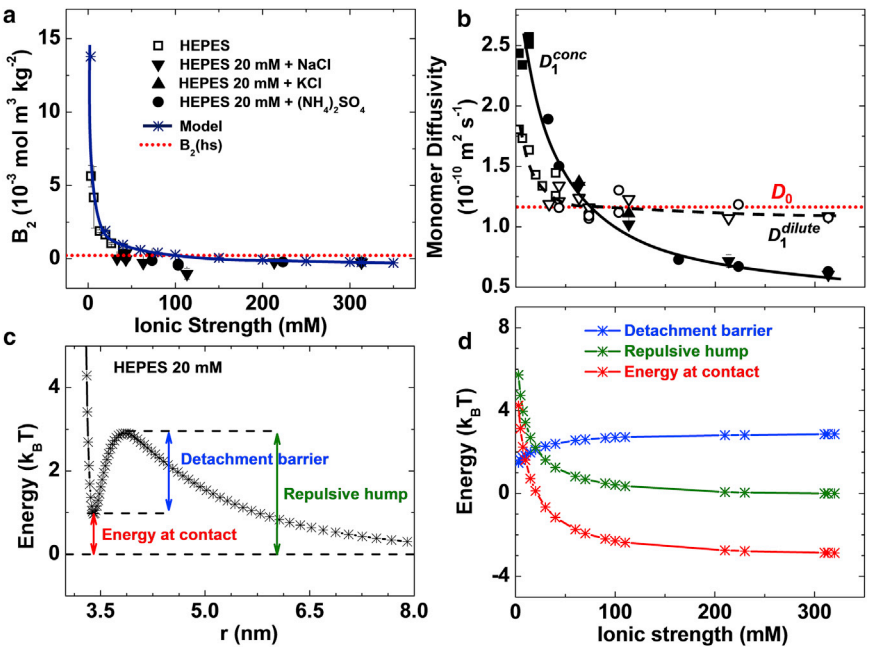


FIGURE 3 Characterization of the intermolecular interactions in solution at pH = 7.8. (*a*) The dependence of the second osmotic virial coefficient $B_2$ on the ionic strength $I$, varied through the concentration of HEPES buffer or by adding KCl, NaCl, or $(NH_4)_2SO_4$. The values of $B_2$ computed using the model represented in Fig. S3 are also shown. $B_2(\text{hs}) = 4\,V_M\,N_A\,M_w^{-2} = 2.35 \times 10^{-4}$ mol m$^3$ kg$^{-2}$ for hard spheres ($V_M = 2.0 \times 10^{-26}$ m$^3$, molecular volume; $N_A$, Avogadro's number; $M_w = 14.5$ kg mol$^{-1}$, lysozyme molecular weight) is shown for comparison. (*b*) The dependence of monomer diffusivity in dilute ($D_1^{\text{dilute}}$, measured at 9 mg mL$^{-1}$, *open symbols*) and concentrated ($D_1^{\text{conc}}$, measured at 100 mg mL$^{-1}$, *solid symbols*) solutions on the ionic strength, varied by the addition of the same electrolytes as in (*a*). The Stokes-Einstein diffusivity (or self-diffusivity) $D_0 = 1.20 \times 10^{-10}$ m$^2$ s$^{-1}$ of a sphere of radius 1.7 nm in a solution with viscosity 1.06 mPa s is shown. Solid and dashed lines are guides for the eye. (*c*) Potential of mean force (PMF) between a pair of molecules as a function of the distance between their centers of mass calculated using a numerical model illustrated in Fig. S3 at ionic strength $I = 13.3$ mM. (*d*) The dependences of the energy at contact, repulsive hump, and detachment barrier, defined in (*c*) on the ionic strength. To see this figure in color, go online.

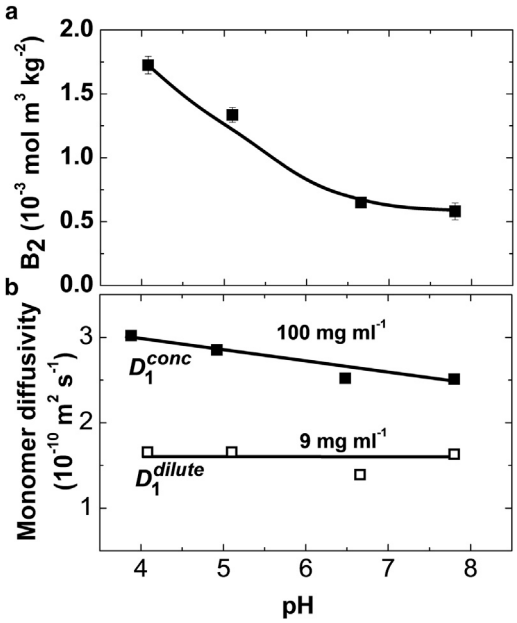


FIGURE 4 pH effects on the intermolecular interactions. (*a*) Dependence of the second osmotic virial coefficient $B_2$ on the solution pH. (*b*) Dependencies of the monomer diffusivities in dilute and concentrated solutions, $D_1^{dilute}$ and $D_1^{conc}$, respectively.

i.e., repulsive hump, and detachment barrier; their dependences on $I$ are displayed in Fig. 3 *d*.

The values of $B_2$ computed using these pairwise potentials are shown in Fig. 3 *a*. The predictions of the model agree well with the experimental data at high ionic strengths, and slightly underestimate the attraction at $I < 120$ mM. The latter discrepancy is likely due to solvent-structuring interactions that are not included in the model. The good overall agreement of the model with the $B_2$ data indicates that the Coulomb interactions adequately account for the observed response of the pairwise intermolecular interactions to increasing ionic strength.

## The effects of the Coulomb forces on the cluster population

The responses of the average cluster radius $R_2$ and the volume fraction occupied by the clusters $\varphi_2$ to changes in the solution ionic strength $I$ are displayed in Fig. 5, *a* and *b*. $R_2$ is independent of $I$, within the experimental error, while $\varphi_2$ decreases approximately fourfold as $I$ increases from 3 to ~100 mM and saturates at higher $I$ values. Fig. 5, *c* and *d*, reveals that the values of $R_2$ and $\varphi_2$ are relatively steady in time (the slow $R_2$ growth likely reflects an Ostwald-like ripening of the clusters (14)). The effects of pH on $R_2$ and $\varphi_2$ are displayed in Fig. 6, *a* and *b*, respectively. Similarly to the trend in Fig. 5 *a*, the cluster radius $R_2$ depends weakly on solution pH; several repetitions of this experiment revealed no pH dependence. The cluster volume fraction $\varphi_2$ increases by ~4× as pH increases from 3.8 to 7.8.

The apparent increase in the cluster volume fraction $\varphi_2$ at higher pH is expected: the protein charge should decrease with pH, thus reducing protein-protein repulsion; this reduction should be stronger when the molecules are closer, i.e., at higher concentrations. The decreasing trends of $D_1^{high}$ and $D_1^{low}$ with increasing pH in Fig. 3 *b* are consistent with these expectations. However, Figs. 5 and 6 also reveal at least three anomalous cluster behaviors. First, the behaviors of $R_2$ and $\varphi_2$ as functions of $I$ are decoupled; this observation is in contrast with conventional phase transformations—such as solidification or liquefaction—in which the domain size of the incipient phase increases concurrently with its overall volume. Second is the anomaly of the cluster size, which is independent of the solution's ionic strength or pH despite the decreasing intermolecular repulsion at higher values of the two parameters, evidenced by Figs. 3, *a* and *b*, and 4, *a* and *b*. Third, there is the puzzling behavior is the decreasing cluster volume fraction $\varphi_2$ at high ionic strength $I$ in Fig. 5 *b*. This contradicts the expectation that weaker repulsion in concentrated solutions, revealed by the $D_1^{high}$ trend in Fig. 3 *b*, should lead to cluster stabilization and, hence, to a higher cluster volume fraction, similarly to the observations at higher pH in Fig. 6 *b*. Note that the three anomalous behaviors contradict general rules of phase transformations and solution thermodynamics irrespective of a specific model of cluster formation.

The decoupled behaviors of $R_2$ and $\varphi_2$ at increasing $I$ indicate that $R_2$ and $\varphi_2$ are controlled by distinct mechanisms. This observation agrees with the oligomer mechanism of cluster formation by which $R_2$ is determined by the kinetics of decay of the oligomers accumulated in the clusters (13), while $\varphi_2$ reflects the high free energy cost of bringing together positively charged molecules (13,14). The second feature, the lack of correlation between the cluster size $R_2$ and the solution ionic strength and pH, indicates that cluster formation is not governed by Coulomb interactions. Thus, neither the colloid scenario of Coulomb-regulated cluster formation (32,36), discussed above, nor a mechanism relying on electrostatically bound oligomer, could underlie the mesoscopic clusters in lysozyme solutions. The third peculiarity of the above data is discussed in the next subsection.

Another example of Coulomb-independent behavior is presented by proteins of the γ-crystalline family (16). These crystallines form clusters that are clearly mesoscopic: each cluster contains a large number of monomers; the total cluster population occupies low volume. Similarly to the lysozyme clusters, the clusters of γ-crystalline retain a size of ~100 nm as pH is varied from 6.8 to 10 and the NaCl concentration, from 150 to 350 mM (16). (The cluster disaggregation at pH and NaCl concentration values outside these ranges (16) may be due to a protein-specific mechanism

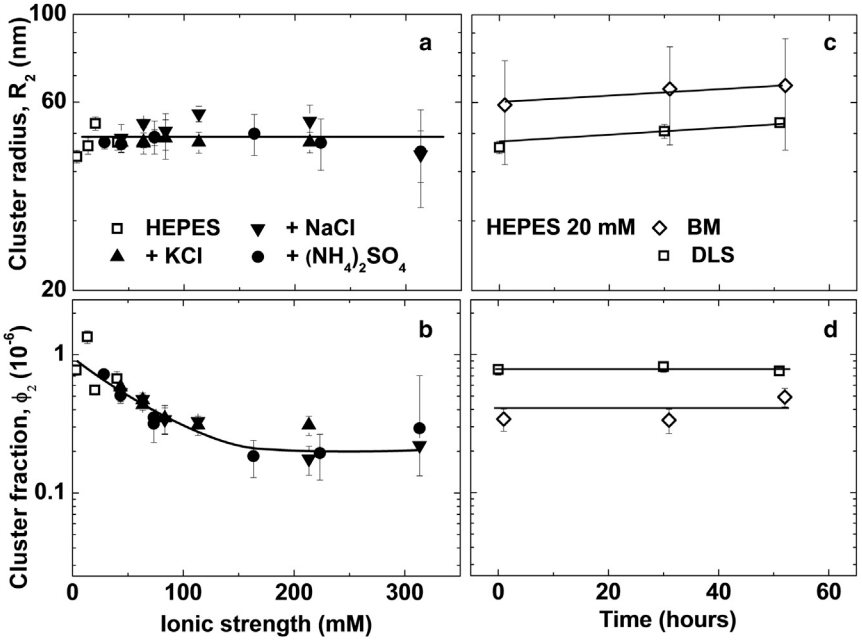


FIGURE 5 Populations of protein-rich clusters in 100 mg $mL^{-1}$ lysozyme solutions in HEPES buffer at pH = 7.8 characterized by DLS and BM. (*a*) Dependence of the average cluster radius $R_2$ on ionic strength. Solution ionic strength was varied through the concentration of four electrolytes, as indicated in the graph. (*b*) Cluster volume fraction $\varphi_2$, determined from the same DLS autocorrelation functions as $R_2$, as a function of the solution ionic strength. (*c* and *d*) The evolutions of cluster size $R_2$ in (*c*) and volume fraction $\varphi_2$ in (*d*), determined by DLS and BM. The results of the two methods are similar, within their errors; the inherent error of the $\varphi_2$ determination may be up to 50% (16).

that is beyond the assumptions of the clustering models discussed here.)

With many other studied proteins, clusters are observed at ionic strengths higher than 100 mM (12-14,16,17,24), at which the Debye length is shorter than the molecular size and, hence, the lifetimes of electrostatically bound oligomers would be insignificant. Hence, Coulomb-regulated colloid clustering and Coulomb-mediated oligomerization can be dismissed as formation mechanisms of the mesoscopic clusters in solutions of these proteins.

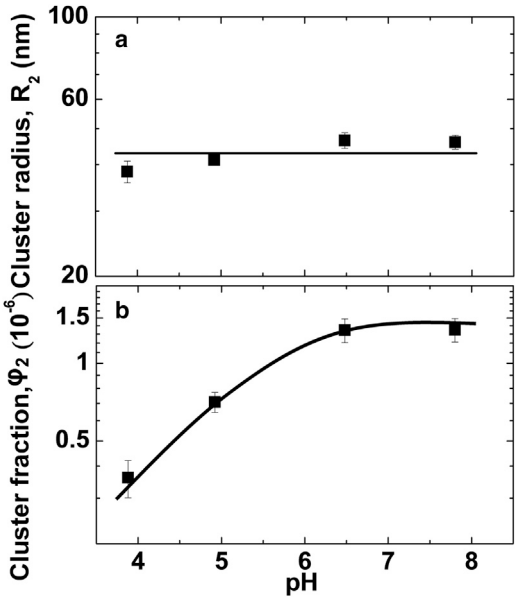


FIGURE 6 pH effects on the cluster characteristics, the cluster radius $R_2$ in (*a*), and the cluster volume fraction $\varphi_2$ in (*b*).

## Water-structuring interactions and partial protein unfolding

The anomalous decrease of $\varphi_2$ at high $I$ in Fig. 5 *b* is akin to salting-in, i.e., the increase of solubility of proteins and colloids at increasing ionicity. This decrease contradicts the trend of decreasing molecular repulsion at high ionic strength, revealed by Fig. 3, *a* and *b*, for protein concentrations up to 100 mg $mL^{-1}$, and suggests that forces other than Coulomb are at play. These hypothetical forces must then destabilize the dense liquid held in the clusters, in which the protein concentration is ~500 mg $mL^{-1}$ (13) and the intermolecular separation is shorter than 1 nm (13). Possible candidates are water-structuring forces that operate at similarly short separations (55). They are classified either as hydration, due to water structuring at polar surface patches and augmented by the presence of ions and other kosmotropes (55,57), or hydrophobic, due to water layering along nonpolar surface patches (55). Thus, increasing concentrations of kosmotropic ions could induce the buildup of hydration layers and hydration repulsion at short intermolecular separations that destabilize the cluster phase.

To test the role of hydration and hydrophobic forces in cluster formation, we added urea and ethanol to the probed solutions. Urea is known to destabilize the native structure of most proteins; addition of 8 M urea in aqueous solutions causes full protein unfolding (58-60). The contemporary consensus appears to be that urea is a universal denaturant because it interacts favorably with the peptide backbone (61). The amino-acid side chains assist the action of urea by additional preferential interaction with it and by diluting the effective concentration of the backbone amides (62-64). The interactions of urea with the backbone and side chains

involve intercalation and destruction of the water structures (chaotropic action) (65). Ethanol forms homogeneous solutions with water at concentrations <2.8 M (66). Similarly to urea, ethanol is a chaotropic agent; however, it accumulates in the vicinity of nonpolar amino-acid residues and disrupts adjacent water structures. It strips off as many as 16 bound water molecules from the lysozyme surface (67) and may form hydrogen bonds to its hydroxyl groups (68). In important contrast to urea, ethanol does not interact with the peptide backbone and, hence, it induces protein unfolding only at high concentrations: a recent study demonstrated that ethanol does not affect the conformations of lysozyme $\alpha$-helixes and $\beta$-sheets at concentrations as high as 2.5 M (68).

We have characterized the effects of urea and ethanol at concentrations up to ~2.5 M on the protein interactions in terms of the second osmotic virial coefficient $B_2$ (determined from plots in Fig. S6) and the product of monomer diffusivity and buffer viscosity $D_1{}^{conc}\eta_{buffer}$ (in contrast to the salts used to modify protein interactions in Figs. 3, 4, and 5, urea and ethanol significantly affect the buffer viscosity at the applied concentrations). The results in Fig. 7, *a* and *b*, reveal that the addition of urea or ethanol perceptibly enhances intermolecular repulsion, likely by weakening the hydrophobic attraction. The ethanol effects are consistent with disruption of the water structures, likely around the nonpolar surface amino-acid residues. In addition, urea likely acts also on the peptide backbone exposed to the solvent (backbone segments accessible to urea in the native confirmation are illustrated in the *inset* in Fig. 7 *a*; the solvent exposure of additional backbone segments due to partial protein unfolding is supported by evidence below). Thus, stronger urea-induced repulsion is consistent with weakening of the water structures around these backbone segments.

While the ethanol-induced increase in $B_2$ (~15%) is comparable to that of $D_1{}^{conc}\eta_{buffer}$, the increase in $B_2$ with adding urea (~40%) is stronger than that of $D_1{}^{conc}\eta_{buffer}$ (~20%). As discussed above, $D_1{}^{conc}\eta_{buffer}$ is weighted toward short intermolecular separations. Hence, this discrepancy indicates that urea boosts repulsion more efficiently at long distances. Because the mismatch is not observed with ethanol, we assign it to urea-enhanced partial protein unfolding. It exposes to the solvent nonpolar side chains that in the native structure are tucked inside. The resulting hydrophobic attraction acts at the short range, at which it mitigates the repulsion caused by urea coating the protein backbone.

The data in Fig. 7 *c* demonstrate that the addition of urea reduces the cluster radius $R_2$ approximately threefold, while increasing the cluster population volume fraction $\varphi_2$ by an order of magnitude. The addition of ethanol does not affect $R_2$ and weakly lowers $\varphi_2$ (Fig. 7 *d*). The decoupled behaviors of $R_2$ and $\varphi_2$ in the presence of urea exclude protein denaturation and aggregation induced by this additive as the cause of the observed trends. We carried out two additional tests of the possibility of denaturation. First, we determined $R_2$ and $\varphi_2$ in a protein solution containing 1.25 M urea, prepared by mixing a solution with 2.5 M urea with an equal volume of a protein solution of the same concentration and no urea. The measured $R_2$ and $\varphi_2$ (Fig. S7) were practically identical to those in directly prepared 1.25 M urea, indicating that cluster formation and its constituent processes are reversible. Second, we monitored the evolution of the cluster population over 24 h. We found (Fig. S8) that $R_2$ and $\varphi_2$ did not change from the values established within 30 min after the addition of urea (Fig. 7 *c*). In combination with the conclusion of cluster reversibility, the latter observation implies that the cluster population is in equilibrium with the solution, similar to its behavior in the absence of urea (13).

The $R_2$ and $\varphi_2$ responses to urea are anomalous from a classical viewpoint: a significant increase in the cluster-phase volume is accompanied by a decrease in its characteristic dimension. On the other hand, they are compatible with the oligomer mechanism of cluster

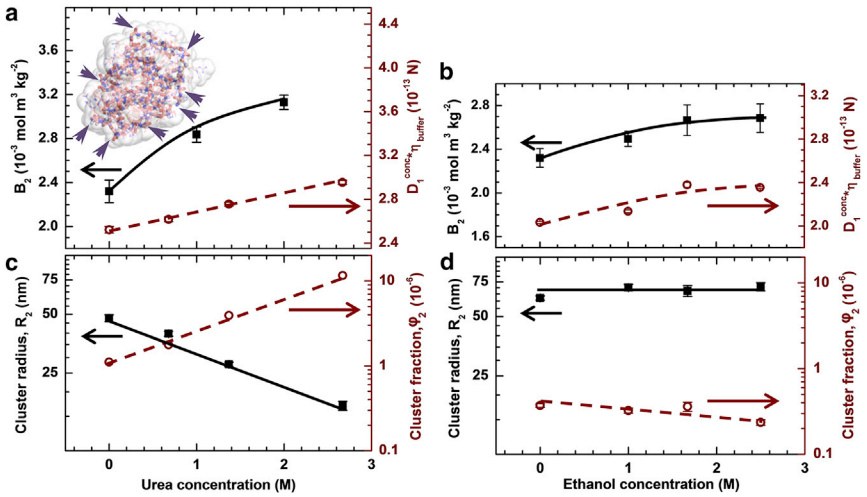


FIGURE 7 The role of hydrophobic interactions and partial unfolding in cluster formation. All data are for 100 mg mL$^{-1}$ lysozyme solutions in 20 mM HEPES at pH = 7.8, in which the ionic strength $I$ = 13.3 mM. (*a* and *b*) Variation of the second virial coefficient $B_2$ (*left ordinate*, *solid symbols*) and of the product of monomer diffusion coefficient $D_1{}^{conc}$ in 100 mg mL$^{-1}$ solutions and the buffer viscosity $\eta_{buffer}$ (*right ordinate*, *open symbols*) as functions of the concentration of urea and (*a*) and ethanol in (*b*). (*Inset* in *a*) The native structure of lysozyme and its solvent-accessible surface. (*Arrows*) Locations at which the peptide backbone is exposed to the solvent. (*c* and *d*) The response of the cluster radius $R_2$ (*left ordinate*, *solid symbols*) and volume fraction $\varphi_2$ (*right ordinate*, *open symbols*) to increasing concentrations of urea and (*c*) and ethanol in (*d*). To see this figure in color, go online.

formation, according to which $R_2$ and $\varphi_2$ are independently regulated. Furthermore, comparing the variations of $R_2$ due to the addition of urea and ethanol indicates that oligomers bound by backbone-to-backbone contacts are crucial for cluster formation. The accumulation of urea around the peptide backbone would accelerate the decay of such oligomers and increase the corresponding rate constant $k_{\mathrm{oligomer}}$ and lead, according to Eq. 1, to smaller clusters. Because ethanol does not interact with the backbone, it does not affect $R_2$. The responses of $\varphi_2$ to urea and ethanol highlight the role of partial protein unfolding in oligomer stabilization and cluster formation. Enhanced protein unfolding by urea (tentatively indicated by the discrepancy in the $B_2$ and ${D_1}^{\mathrm{conc}}\eta_{\mathrm{buffer}}$ trends in Fig. 7 *a*) exposes hidden nonpolar amino-acid residues. Because the attractive hydrophobic interactions between the residues are short-ranged, this stabilizes the cluster phase more than in the dilute solution and hence increases $\varphi_2$.

Looking back at the effects of electrolytes, we note that one of the used salts, $(NH_4)_2SO_4$, combines electrostatic with kosmotropic and chaotropic actions due to its two ions, i.e., $SO_4^{2-}$ stabilizes the water shells around proteins and the native protein conformation, while $NH_4^+$ destabilizes water structures and tends to denature proteins (69). It appears that at the highest concentration used here, 100 mM (higher concentrations lead to fast crystallization), the chaotropic action is not exhibited; the protein conformation is stable; and the action of $(NH_4)_2SO_4$ is fully accounted for by its contribution to the ionic strength $I$.

The responses of $R_2$ and $\varphi_2$ to the presence of urea and ethanol in Fig. 7 are not dramatic, implying that the cluster formation mechanism has not been modified by these two additives. These responses identify partial protein unfolding as the likely cause behind the existence of mesoscopic clusters in lysozyme solutions with widely ranging compositions. Note that only a small fraction, $10^{-6}$–$10^{-4}$, of the total soluble protein partially unfolds and is held in the clusters. The unfolding exposes to the solvent the peptide backbone and nonpolar amino-acid residues, hidden in the native conformation, enables hydrophobic bonds between backbone segments, and stabilizes the cluster phase through hydrophobic attraction between the exposed nonpolar amino-acid residues. We have demonstrated that the constituent steps in this scenario are reversible, which indicates that it is fully compatible with the oligomer mechanism of cluster formation (13), wherein backbone-to-backbone contacts support transient oligomers.

## CONCLUSIONS

These results demonstrate that the Coulomb forces that govern aggregation in biological systems and many other phenomena in nature do not affect the size of the mesoscopic clusters in lysozyme solutions. In addition to their large size, high amount of protein contained in each cluster, small fraction of total protein held in the clusters, and concentration independence of the size, the insensitivity of the cluster size to Coulomb forces distinguishes the mesoscopic clusters sharply from the two other classes of clusters observed in protein solutions. The mesoscopic clusters exhibits other behaviors that are in contrast with established laws of phase equilibrium: decoupled responses of cluster phase volume and cluster size to variations of the ionic strength, pH, and additive concentration; and decreased cluster phase volume upon stronger intermolecular attraction. These responses demonstrate that the mesoscopic clusters represent, to our knowledge, a novel class of protein condensate that forms by a fundamentally different mechanism from protein crystals and amyloid fibrils, and from the two other known types of protein clusters. Our observations indicate that the clusters form by a unique mechanism, i.e., by the accumulation of transient protein oligomers that are linked by hydrophobic bonds between the peptide backbones exposed to the solvent after partial protein unfolding. Because the mesoscopic clusters have been suggested in many cases as crucial precursors to the formation of the two main classes of protein aggregates, crystals and amyloid fibrils, our findings indicate that fine-tuning of the intra- and intermolecular water-structuring interactions may be an essential tool to control the cluster population and in this way enhance or suppress protein crystallization and fibrillization.

## SUPPORTING MATERIAL

Supporting Materials and Methods, eight figures, and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00996-0.

## AUTHOR CONTRIBUTIONS

M.A.V. performed the experiments, H.Y.C. designed the numerical model and carried out the model computations, V.L. designed the numerical model and edited the text, and P.G.V. conceived the study, designed the experiments and wrote the article with contributions from M.A.V. and V.L. All authors discussed the results and commented on the article.

## ACKNOWLEDGMENTS

We thank A. Kolomeisky for discussion of cluster properties, M. G. Murray for assistance in the tests with ethanol, and R. C. Willson for help with experimental procedures.

This work was supported by the National Science Foundation (grant Nos. MCB-1244568 and MCB-1518204), the *National Aeronautics and Space Administration* (grant Nos. NNX14AD68G and NNX14AE79G), and the Welch Foundation (grant No. E-1765).

## REFERENCES

1. Stradner, A., H. Sedgwick, …, P. Schurtenberger. 2004. Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature*. 432:492–495.

2. Porcar, L., P. Falus, …, Y. Liu. 2009. Formation of the dynamic clusters in concentrated lysozyme protein solutions. *J. Phys. Chem. Lett.* 1:126–129.
3. Erlkamp, M., S. Grobelny, …, R. Winter. 2014. Solvent effects on the dynamics of amyloidogenic insulin revealed by neutron spin echo spectroscopy. *J. Phys. Chem. B.* 118:3310–3316.
4. Yearley, E. J., P. D. Godfrin, …, Y. Liu. 2014. Observation of small cluster formation in concentrated monoclonal antibody solutions and its implications to solution viscosity. *Biophys. J.* 106:1763–1770.
5. Shukla, A., E. Mylonas, …, D. I. Svergun. 2008. Absence of equilibrium cluster phase in concentrated lysozyme solutions. *Proc. Natl. Acad. Sci. USA.* 105:5075–5080.
6. Liu, Y., L. Porcar, …, P. Baglioni. 2011. Lysozyme protein solution with an intermediate range order structure. *J. Phys. Chem. B.* 115:7238–7247.
7. Johnston, K. P., J. A. Maynard, …, K. J. Kaczorowski. 2012. Concentrated dispersions of equilibrium protein nanoclusters that reversibly dissociate into active monomers. *ACS Nano.* 6:1357–1369.
8. Soraruf, D., F. Roosen-Runge, …, F. Schreiber. 2014. Protein cluster formation in aqueous solution in the presence of multivalent metal ions—a light scattering study. *Soft Matter.* 10:894–902.
9. Zhang, F., F. Roosen-Runge, …, F. Schreiber. 2012. The role of cluster formation and metastable liquid-liquid phase separation in protein crystallization. *Faraday Discuss.* 159:313–325.
10. Borwankar, A. U., A. K. Dinin, …, K. P. Johnston. 2013. Tunable equilibrium nanocluster dispersions at high protein concentrations. *Soft Matter.* 9:1766–1771.
11. Cardinaux, F., E. Zaccarelli, …, P. Schurtenberger. 2011. Cluster-driven dynamical arrest in concentrated lysozyme solutions. *J. Phys. Chem. B.* 115:7227–7237.
12. Gliko, O., N. Neumaier, …, P. G. Vekilov. 2005. A metastable prerequisite for the growth of lumazine synthase crystals. *J. Am. Chem. Soc.* 127:3433–3438.
13. Pan, W., P. G. Vekilov, and V. Lubchenko. 2010. Origin of anomalous mesoscopic phases in protein solutions. *J. Phys. Chem. B.* 114:7620–7630.
14. Li, Y., V. Lubchenko, …, P. G. Vekilov. 2012. Ostwald-like ripening of the anomalous mesoscopic clusters in protein solutions. *J. Phys. Chem. B.* 116:10657–10664.
15. Gliko, O., W. Pan, …, P. G. Vekilov. 2007. Metastable liquid clusters in super- and undersaturated protein solutions. *J. Phys. Chem. B.* 111:3106–3114.
16. Mohr, B. G., C. M. Dobson, …, M. Muthukumar. 2013. Electrostatic origin of in vitro aggregation of human $\gamma$-crystallin. *J. Chem. Phys.* 139:121914.
17. Sleutel, M., and A. E. van Driessche. 2014. Role of clusters in nonclassical nucleation and growth of protein crystals. *Proc. Natl. Acad. Sci. USA.* 111:E546–E553.
18. Maes, D., M. A. Vorontsova, …, P. G. Vekilov. 2015. Do protein crystals nucleate within dense liquid clusters? *Acta Crystallogr. F Struct. Biol. Commun.* 71:815–822.
19. Pan, W., O. Galkin, …, P. G. Vekilov. 2007. Metastable mesoscopic clusters in solutions of sickle-cell hemoglobin. *Biophys. J.* 92:267–277.
20. Uzunova, V., W. Pan, …, P. G. Vekilov. 2012. Control of the nucleation of sickle cell hemoglobin polymers by free hematin. *Faraday Discuss.* 159:87–104.
21. Uzunova, V. V., W. Pan, …, P. G. Vekilov. 2010. Free heme and the polymerization of sickle cell hemoglobin. *Biophys. J.* 99:1976–1985.
22. ten Wolde, P. R., and D. Frenkel. 1997. Enhancement of protein crystal nucleation by critical density fluctuations. *Science.* 277:1975–1978.
23. Giegé, R. 2013. A historical perspective on protein crystallization from 1840 to the present day. *FEBS J.* 280:6456–6497.
24. Galkin, O., W. Pan, …, P. G. Vekilov. 2007. Two-step mechanism of homogeneous nucleation of sickle cell hemoglobin polymers. *Biophys. J.* 93:902–913.
25. Krishnan, R., and S. L. Lindquist. 2005. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature.* 435:765–772.
26. Luiken, J. A., and P. G. Bolhuis. 2015. Prediction of a stable associated liquid of short amyloidogenic peptides. *Phys. Chem. Chem. Phys.* 17:10556–10567.
27. Ruff, K. M., S. J. Khan, and R. V. Pappu. 2014. A coarse-grained model for polyglutamine aggregation modulated by amphipathic flanking sequences. *Biophys. J.* 107:1226–1235.
28. Li, P., S. Banjade, …, M. K. Rosen. 2012. Phase transitions in the assembly of multivalent signalling proteins. *Nature.* 483:336–340.
29. Brangwynne, C. P. 2013. Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.* 203:875–881.
30. Weber, S. C., and C. P. Brangwynne. 2015. Inverse size scaling of the nucleolus by a concentration-dependent phase transition. *Curr. Biol.* 25:641–646.
31. Hyman, A. A., C. A. Weber, and F. Jülicher. 2014. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* 30:39–58.
32. Sciortino, F., S. Mossa, …, P. Tartaglia. 2004. Equilibrium cluster phases and low-density arrested disordered states: the role of short-range attraction and long-range repulsion. *Phys. Rev. Lett.* 93:055701.
33. Mani, E., W. Lechner, …, P. G. Bolhuis. 2014. Equilibrium and non-equilibrium cluster phases in colloids with competing interactions. *Soft Matter.* 10:4479–4486.
34. Groenewold, J., and W. K. Kegel. 2001. Anomalously large equilibrium clusters of colloids. *J. Phys. Chem. B.* 105:11702–11709.
35. Cardinaux, F., A. Stradner, …, E. Zaccarelli. 2007. Modeling equilibrium clusters in lysozyme solutions. *EuroPhys. Lett.* 77:48004.
36. Hutchens, S. B., and Z.-G. Wang. 2007. Metastable cluster intermediates in the condensation of charged macromolecule solutions. *J. Chem. Phys.* 127:084912.
37. Eyink, G., J. Lebowitz, and H. Spohn. 1996. Hydrodynamics and fluctuations outside of local equilibrium: driven diffusive systems. *J. Stat. Phys.* 83:385–472.
38. Goto, Y., L. J. Calciano, and A. L. Fink. 1990. Acid-induced folding of proteins. *Proc. Natl. Acad. Sci. USA.* 87:573–577.
39. Thornton, J. M. 1982. Electrostatic interactions in proteins. *Nature.* 295:13–14.
40. Meltzer, R. H., E. Thompson, …, S. E. Pedersen. 2006. Electrostatic steering at acetylcholine binding sites. *Biophys. J.* 91:1302–1314.
41. Anderson, C. F., and M. T. Record. 1982. Polyelectrolyte theories and their applications to DNA. *Annu. Rev. Phys. Chem.* 33:191–222.
42. Ozkirimli, E., S. S. Yadav, …, C. B. Post. 2008. An electrostatic network and long-range regulation of Src kinases. *Protein Sci.* 17:1871–1880.
43. Dean, A. M., and D. E. Koshland, Jr. 1990. Electrostatic and steric contributions to regulation at the active site of isocitrate dehydrogenase. *Science.* 249:1044–1046.
44. Hemsath, L., R. Dvorsky, …, M. R. Ahmadian. 2005. An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. *Mol. Cell.* 20:313–324.
45. Buell, A. K., P. Hung, …, T. P. Knowles. 2013. Electrostatic effects in filamentous protein aggregation. *Biophys. J.* 104:1116–1126.
46. Chiti, F., M. Calamai, …, C. M. Dobson. 2002. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc. Natl. Acad. Sci. USA.* 99 (*Suppl 4*):16419–16426.
47. Marshall, K. E., K. L. Morris, …, L. C. Serpell. 2011. Hydrophobic, aromatic, and electrostatic interactions play a central role in amyloid fibril formation and stability. *Biochemistry.* 50:2061–2071.
48. McPherson, A. 2009. Introduction to Macromolecular Crystallography. John Wiley, Hoboken, New Jersey.
49. Chayen, N. E., J. R. Helliwell, and E. H. Snell. 2010. Macromolecular crystallization and crystal perfection. *In* IUCr Monographs on Crystallography, *Vol. 24*. Oxford University Press, Oxford, New York.

50. Chan, H. Y., V. Lankevich, …, V. Lubchenko. 2012. Anisotropy of the Coulomb interaction between folded proteins: consequences for mesoscopic aggregation of lysozyme. *Biophys. J.* 102:1934–1943.

51. Li, Y., V. Lubchenko, and P. G. Vekilov. 2011. The use of dynamic light scattering and Brownian microscopy to characterize protein aggregation. *Rev. Sci. Instrum.* 82:053106.

52. Wetter, L. R., and H. F. Deutsch. 1951. Immunological studies on egg white proteins. IV. Immunochemical and physical studies of lysozyme. *J. Biol. Chem.* 192:237–242.

53. Richetti, P. G., and G. Tudor. 1981. Isoelectric points and molecular weights of proteins a new table. *J. Chromatogr. A.* 220:115–194.

54. Roxby, R., and C. Tanford. 1971. Hydrogen ion titration curve of lysozyme in 6 M guanidine hydrochloride. *Biochemistry.* 10:3348–3352.

55. Leckband, D., and J. Israelachvili. 2001. Intermolecular forces in biology. *Q. Rev. Biophys.* 34:105–267.

56. George, A., and W. W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr. D Biol. Crystallogr.* 50:361–365.

57. Petsev, D. N., and P. G. Vekilov. 2000. Evidence for non-DLVO hydration interactions in solutions of the protein apoferritin. *Phys. Rev. Lett.* 84:1339–1342.

58. Das, A., and C. Mukhopadhyay. 2008. Atomistic mechanism of protein denaturation by urea. *J. Phys. Chem. B.* 112:7903–7908.

59. Caballero-Herrera, A., K. Nordstrand, …, L. Nilsson. 2005. Effect of urea on peptide conformation in water: molecular dynamics and experimental characterization. *Biophys. J.* 89:842–857.

60. Hua, L., R. Zhou, …, B. J. Berne. 2008. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc. Natl. Acad. Sci. USA.* 105:16928–16933.

61. Auton, M., L. M. F. Holthauzen, and D. W. Bolen. 2007. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. USA.* 104:15317–15322.

62. Courtenay, E. S., M. W. Capp, and M. T. Record, Jr. 2001. Thermodynamics of interactions of urea and guanidinium salts with protein surface: relationship between solute effects on protein processes and changes in water-accessible surface area. *Protein Sci.* 10:2485–2497.

63. Diehl, R. C., E. J. Guinn, …, M. T. Record, Jr. 2013. Quantifying additive interactions of the osmolyte proline with individual functional groups of proteins: comparisons with urea and glycine betaine, interpretation of *M*-values. *Biochemistry.* 52:5997–6010.

64. Holehouse, A. S., K. Garai, …, R. V. Pappu. 2015. Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. *J. Am. Chem. Soc.* 137:2984–2995.

65. Sagle, L. B., Y. Zhang, …, P. S. Cremer. 2009. Investigating the hydrogen-bonding model of urea denaturation. *J. Am. Chem. Soc.* 131:9304–9310.

66. Banerjee, S., R. Ghosh, and B. Bagchi. 2012. Structural transformations, composition anomalies and a dramatic collapse of linear polymer chains in dilute ethanol-water mixtures. *J. Phys. Chem. B.* 116:3713–3722.

67. Deshpande, A., S. Nimsadkar, and S. C. Mande. 2005. Effect of alcohols on protein hydration: crystallographic analysis of hen egg-white lysozyme in the presence of alcohols. *Acta Crystallogr. D Biol. Crystallogr.* 61:1005–1008.

68. Chattoraj, S., A. K. Mandal, and K. Bhattacharyya. 2014. Effect of ethanol-water mixture on the structure and dynamics of lysozyme: a fluorescence correlation spectroscopy study. *J. Chem. Phys.* 140: 115105.

69. Collins, K. D. 2004. Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods.* 34:300–311.

# Supporting Material for
# Lack of Dependence of the Sizes of the Mesoscopic Protein Clusters on Electrostatics

M.A. Vorontsova, H.Y. Chan, V. Lubchenko, P.G. Vekilov
[1]Department of Chemical and Biomolecular Engineering; [2]Department of Physics; [3]Department of Chemistry; University of Houston, Houston, Texas

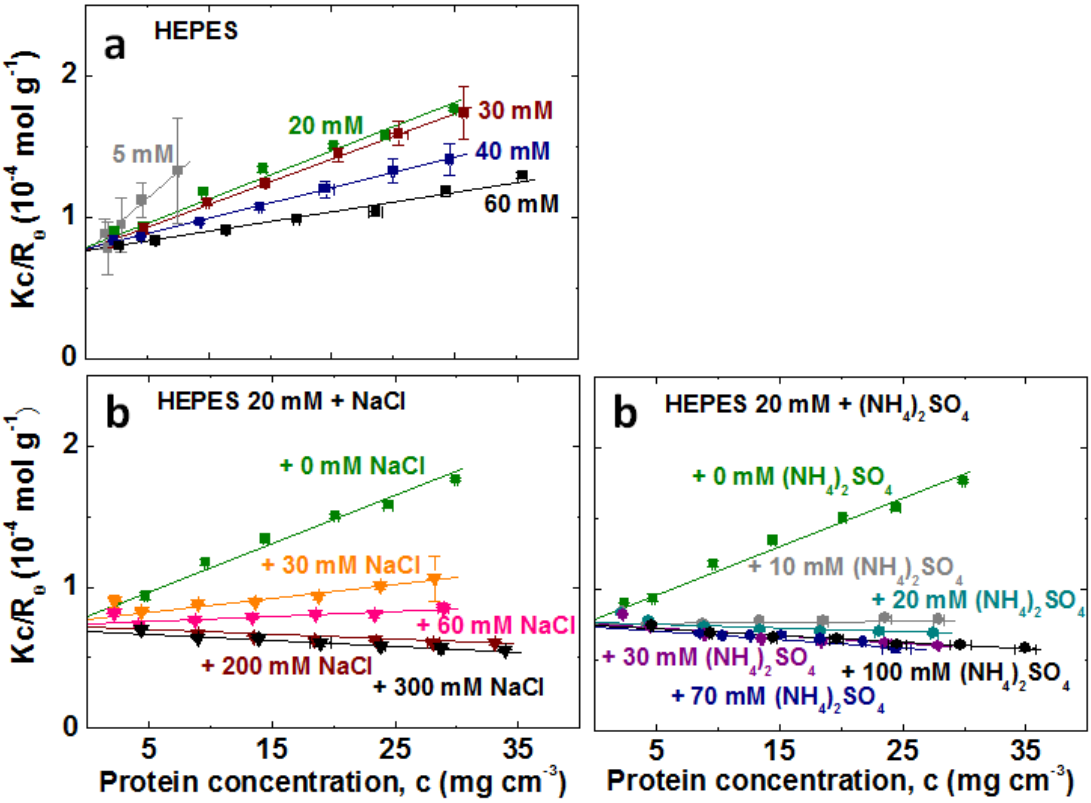


**Figure S1.** Characterization of the intermolecular interaction in dilute lysozyme solutions at increasing ionic strength. Debye plots $Kc/R_\theta$ ($c$) ($K$, instrument constant; $c$, lysozyme concentration; $R_\vartheta = I_\theta/I_0$, Rayleigh ratio of the intensity of light scattered at angle $\theta = 90^o$ to that of incident light) of lysozyme solutions. **a**, in HEPES buffer of concentrations shown in the plot; **b**, in 20 mM HEPES with added NaCl, a 1:1 electrolyte; and **c**, in 20 mM HEPES with added $(NH_4)_2SO_4$, a 1:2 electrolyte. The ionic strength in these solutions increases from 3.3 mM to 40 mM in **a**; and from 13.3 mM to 313 mM in **b** and **c**. At low c, $Kc/R_\theta = M_w^{-1} + 2B_2c$, where $B_2$ is the second osmotic virial coefficient and $M_w$ = 14,300 g mol$^{-1}$ is the molecular weight of lysozyme. In agreement with this relation, the intercept of all plots is numerically close to $M_w^{-1}$. With increasing ionic strength, the slope of the Debye plots, $2B_2$, which is an indicator of pairwise intermolecular interactions, decreases and becomes negative. This indicates that the Coulomb-driven repulsion between lysozyme molecules switches to slight attraction. At ionic strength > 100 mM the effect of electrolyte concentration on $B_2$ and the intermolecular interactions becomes weak.

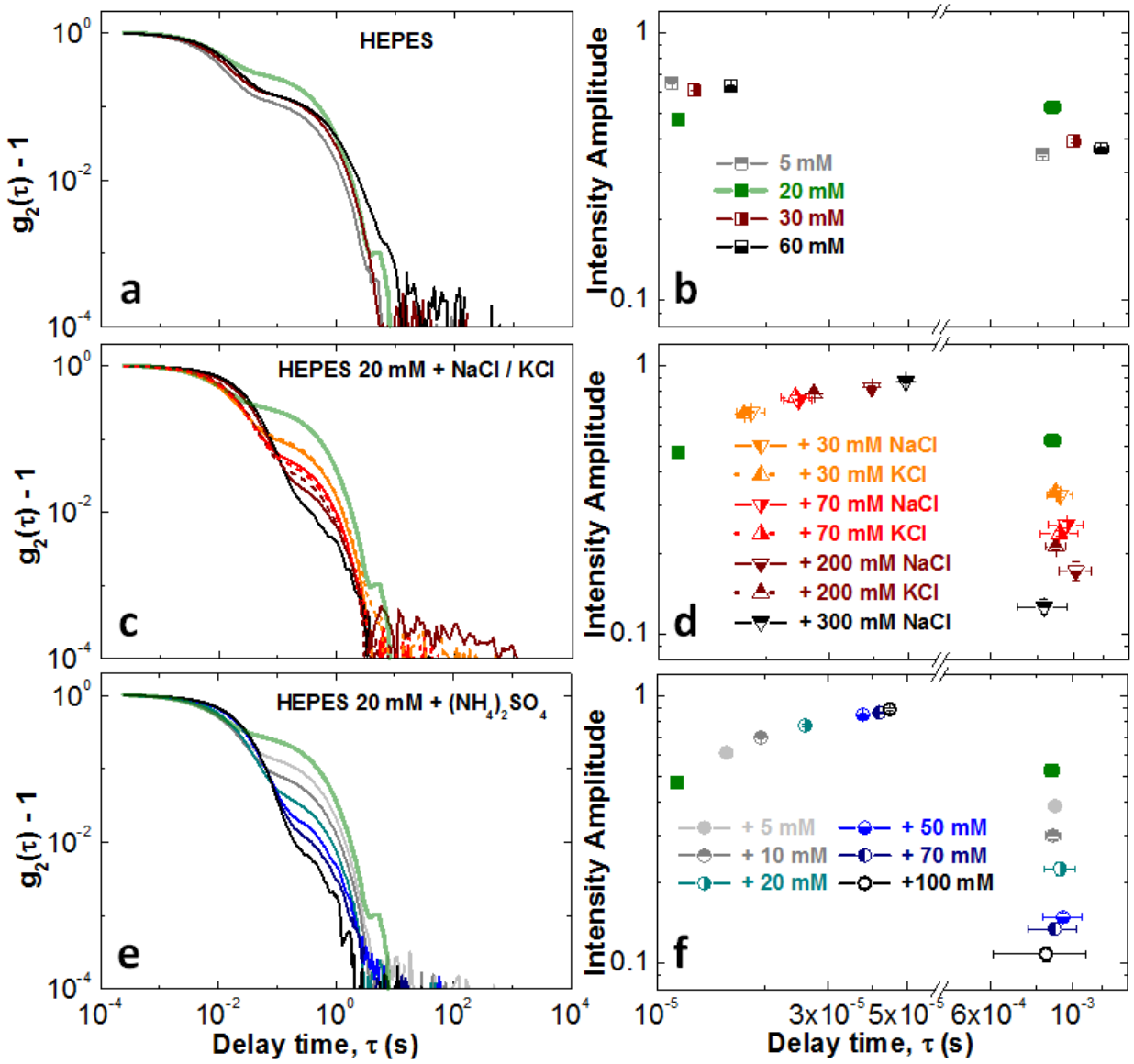


**Figure S2.** Monomers and clusters in lysozyme solutions at varying ionic strength and salt identity, characterized by dynamic light scattering (DLS). All solutions contain 100 mg ml$^{-1}$ lysozyme in HEPES buffer at pH = 7.8. **a, c, e** Normalized autocorrelation functions $g_2 - 1$. The autocorrelation function suggests that there are two populations of scatterers in the analyzed solutions. The shorter delay time corresponds to protein monomer diffusion, while the second decay corresponds to diffusion of protein-rich clusters. **b, d, f,** The amplitudes of the two peaks of the intensity distribution function computed from $g_2$ as discussed in Ref. (1), corresponding to monomers and clusters, respectively. **a, b**, At increasing HEPES concentration, shown in **b**. This increase in HEPES concentration augments the ionic strength from 3 to 40 mM. The amplitudes and characteristic times of the monomer and cluster peaks undergo little change. **c**, **d**, At 20 mM HEPES and increasing NaCl or KCl concentrations, shown in the **d**; **e, f,** At 20 mM HEPES and increasing $(NH_4)_2SO_4$ concentrations, shown in **f**. **c**, **d**, **e**, and **f**: The ionic strength increases from 13.3 mM to 313 mM. The monomer characteristic diffusion time strongly increases, while respective cluster time remains unchanged.

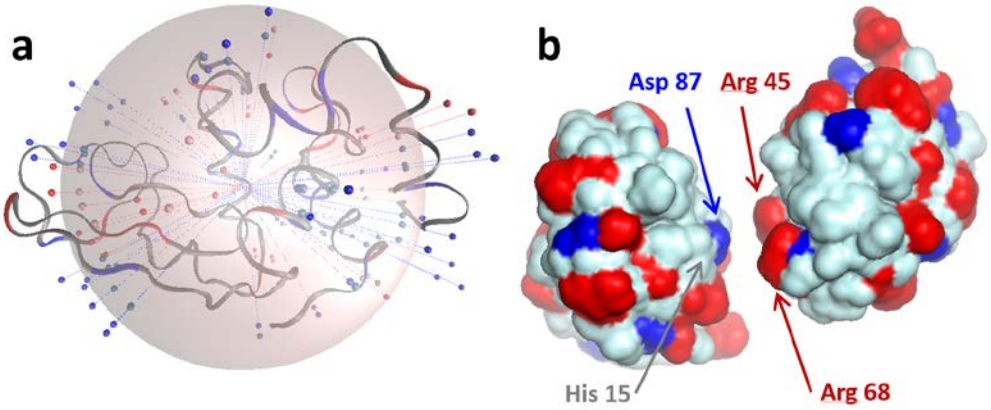


**Figure S3.** The charged groups on the surface of a lysozyme molecule. PBD structure file 2VB1 was used. The positive and negative surface charges are shown in red and blue, respectively. **a**, We represent a lysozyme molecule (whose peptide chain is shown here as a ribbon) as a sphere with radius 1.7 nm and position the positive and negative charges at a depth 0.15 nm beneath the sphere surface at the longitude and latitude equal to those in the molecule (2). **b**, A schematic of the least repulsive mutual orientation of two lysozyme molecules. Residues facing each other in this orientation are marked. Each molecule is represented with its solvent-accessible surface and drawn using PyMOL (www.pymol.org). The models in this figure are for pH = 7.8.

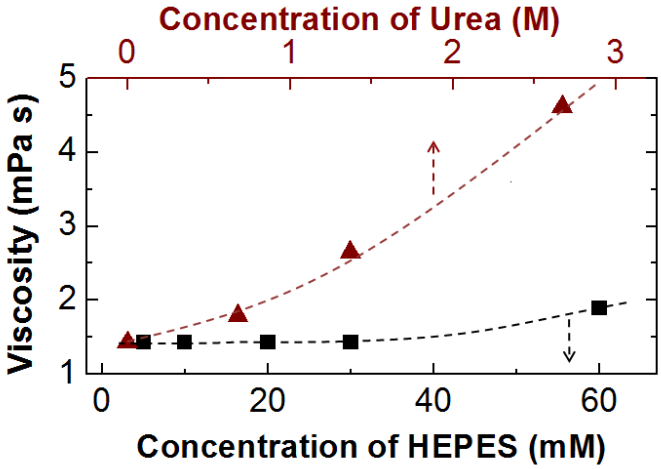


**Figure S4.** The viscosity of 100 mg $ml^{-1}$ lysozyme solutions hosting the protein-rich clusters. Viscosity was determined by dynamic light scattering from the characteristic diffusion time using OptiLink carboxylate-modified polystyrene microparticles with diameter 0.424 μm suspended in 100 mg $ml^{-1}$ solution of lysozyme in HEPES buffer at pH 7.8, for details, see Pan et al., (3) and Li et al., (1). Lower curve: the dependence of viscosity on the concentration of HEPES; viscosity is not affected by the addition of NaCl, KCl, and $(NH_4)_2SO_4$, used to adjust the ionic strength, or KOH and HCl, used to adjust pH. Upper curve: the dependence of viscosity on the concentration of urea in 20 mM HEPES at pH = 7.8. Lines are just guides for the eye.

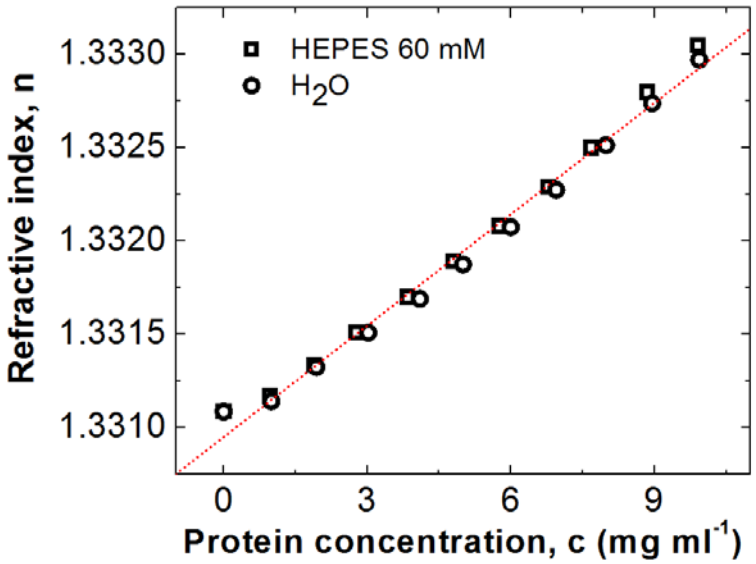


**Figure S5**. The refractive index increment *dn/dc* of lysozyme solutions. *dn/dc* was measured in two solvents, in water, where the ionic strength $I \cong 0$ mM, and in 60 mM HEPES at pH = 7.8, where $I$ = 40 mM. Measurements did not reveal significant difference between the two solvents: both data sets yield *dn/dc* = 0.199 ml/g.

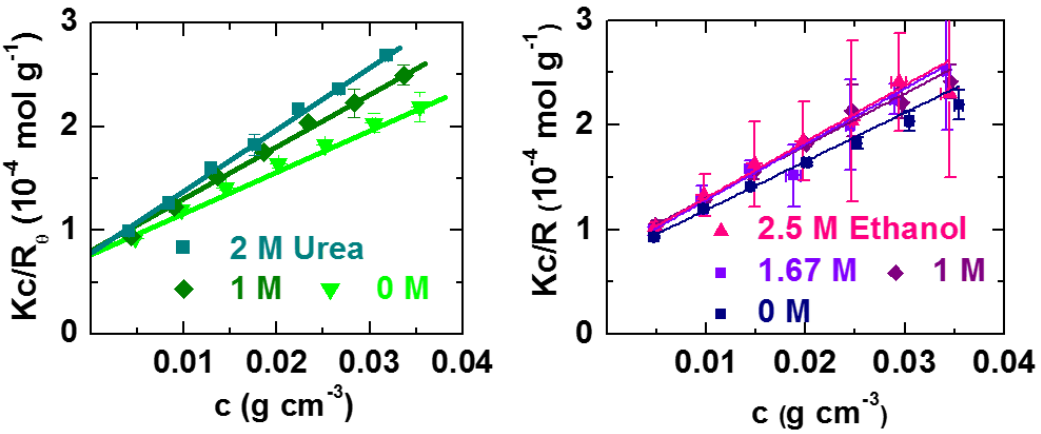


**Figure S6.** Characterization of the intermolecular interaction in dilute lysozyme solutions in the presence of urea and ethanol at concentrations indicated in the plots. Debye plots used to determine $B_2$. $R_\theta = \frac{I_\theta}{I_0}$ is the Rayleigh ratio of the scattered at angle $\theta$ = 90° to the incident light intensity, *c* is the protein concentration, *K* is an optical constant.

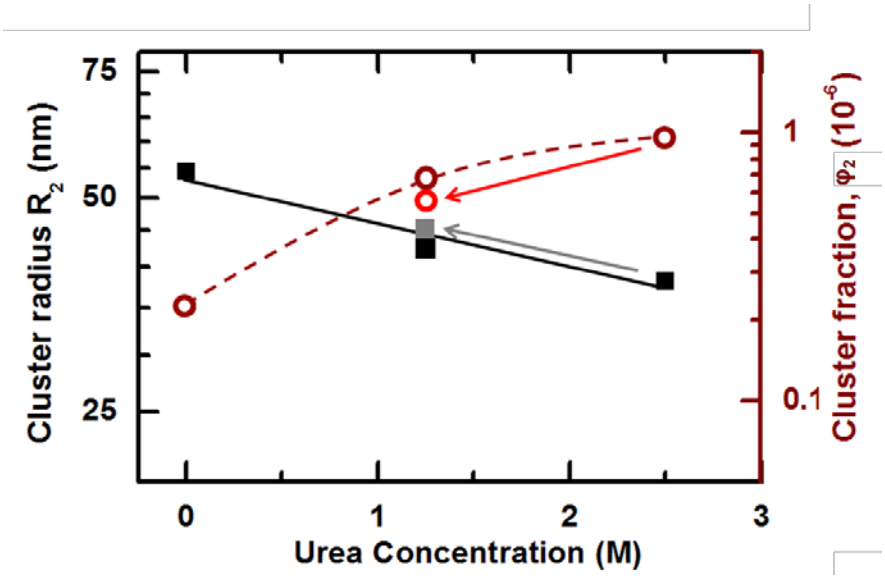


**Figure S7.** The response of the cluster radius $R_2$ (left ordinate, closed symbols) and volume fraction $\varphi_2$ (right ordinate, open symbols) to increasing or decreasing concentrations of urea in a 100 mg ml$^{-1}$ lysozyme solutions in 20 mM HEPES at pH = 7.8, in which the ionic strength *I* = 13.3 mM. The observed trends are identical to those in Fig. 6 c; differences in values of $R_2$ and $\varphi_2$ are due to a different protein batch. Solid black squares and open brown circles denote solutions prepared by the addition of respective urea amounts to lysozymes solutions. For the solutions denoted with grey solid squares and red circle, equal volumes of 0 and 2.5 M urea solutions were mixed, which brings the urea concentration to 1.25 M. The resulting $R_2$ and $\varphi_2$ are very close to the other data pair for the same urea concentration, indicating that cluster formation is reversible.

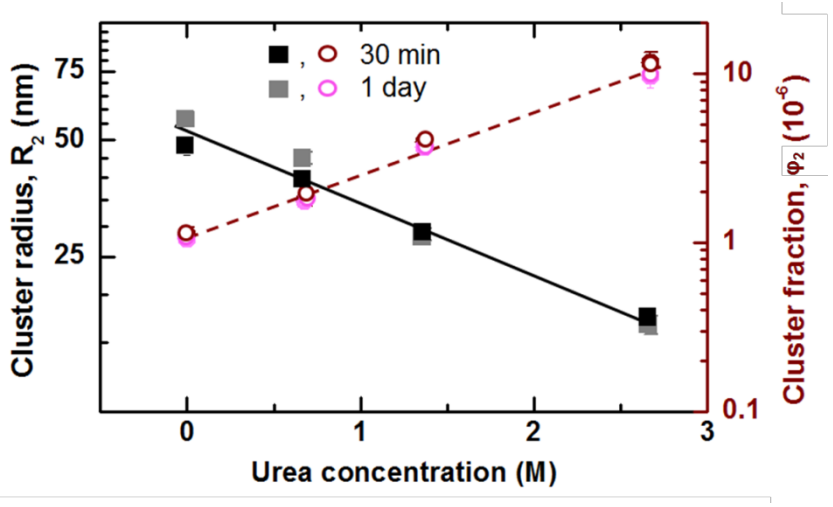


**Figure S8.** The consistency of the radius $R_2$ and volume fraction $\varphi_2$ of the clusters in the presence of urea. Solutions were characterized 30 min and 1 day after preparation. The data for 30 min are from Fig. 6 c.

## Materials and Methods

**Reagents and solutions.** We purchased lyophilized lysozyme from Affymetrix. We also used KCl (Fisher), $(NH_4)_2SO_4$ (Fisher), and NaCl (Mallinckrodt Chemicals). We used HEPES from Fisher and Calbiochem and observed no difference between HEPES from the two sources.

Lysozyme powder was dissolved in K-HEPES (potassium N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonate) buffer and dialyzed against this buffer with pH = 7.8 for two days. We determined the protein concentration using a Beckman Coulter DU 800 Spectrophotometer and extinction coefficient ε=2.64 ml $mg^{-1}$ $cm^{-1}$ at 280 nm (4). We prepared a stock solution of ~150 mg $ml^{-1}$ lysozyme in HEPES buffer of chosen concentration and dialyzed it against the same buffer overnight (14 - 17 hours) to remove low molecular weight acids acquired during production and purification. After dialysis we adjusted the concentration to 100 mg $ml^{-1}$ for dynamic light scattering and Brownian microscopy measurements and brought the ionic strength to the desired value by adding NaCl, KCl or $(NH_4)_2SO_4$. For static light scattering measurements the dialyzed solutions were diluted to an initial concentration ~35-40 mg $ml^{-1}$. All experiments were done at 22°C. Prior to all measurements, the solutions were filtered through 0.22 μm polyether sulfonate (PES) syringe filters (Lightlabs).

Solutions with pH < 7.8 were prepared by dialyzing lysozyme against 60 mM HEPES (chosen to increase the stability of lower pH values and below the threshold of 100 mM, above which the electrostatic interactions are fully screened, Fig. 1 c) at pH = 7.8 in, and titrating this solution to the desired pH with 0.10 M HCl.

**Estimation of solution ionic strength.** HEPES is a monobasic acid with pKa = 7.5, hence at pH = 7.8, about one third of the HEPES molecules are present in protonated form and two thirds, in deprotonated. The concentration of potassium ions is equal to that of deprotonated HEPES. Under these conditions, the ionic strength of the buffer is ca. 0.667× of the total HEPES concentration in the solution.

To estimate the ionic strength of solutions with lower pH, we note that the ionic strength of the starting solution in 60 mM HEPES with pH = 7.8 is *I* = 40 mM. Addition of HCl to a buffer does not alter *I*: the added $Cl^-$ ions compensate the neutralized acid anions. The protonation of the protein to increase its net charge from +8 at pH = 7.8 to +15 at pH = 3.8 requires additional 7 moles $H^+$/mole protein that are accompanied by an equal amount of $Cl^-$. During the determinations of $D_1^{conc}$, $R_2$, and $\phi_2$, the protein concentration is 100 mg $ml^{-1}$ = 6.8 mM. Hence, the ionic strength is 64 mM at pH = 3.8. Accounting for the lower protein charge (9.5 at pH = 5.0 and 8.5 at pH = 6.5 (5)), *I* = 45 and 42 mM at pH = 5.0 and 6.5, respectively.

During the determinations of $D_1^{conc}$, $R_2$, and $\phi_2$ as a function of pH, the ionic strength *I* = 64, 45, and 42 mM at pH = 3.8, 5.0 and 6.5, respectively. According to Figs. 2b and 5b, lower *I* leads to higher $D_1^{conc}$ and $\phi_2$ and this increase partially masks the response of $D_1^{conc}$ and $\phi_2$ to pH. During determinations of $B_2$, the starting ionic strength is 48, 42, 41 mM at pH values of 3.8, 5.0, and 6.5, respectively. These solutions are diluted with buffer solutions at *I* = 40 mM to a final concentration of about 4 mg $ml^{-1}$, in which *I* is between 41 and 40 mM. This decrease of *I* does not affect the slopes of the Debye plots, from which $B_2$ is determined. The higher ionic strength at low pH partially masks the response of $B_2$ to pH. In solutions with concentration 9 mg $ml^{-1}$, as during determinations of $D_1^{dilute}$, the ionic strength at pH values 3.8, 5.0, and 6.5 is, respectively, 42.2, 41, and 40.3 mM. This variation in *I* does not have significant effects on the response of $D_1^{dilute}$ to pH.

**Dynamic light scattering (DLS).** The DLS data were collected by ALV light scattering device equipped with He-Ne laser (λ=632.8 nm, 35 mW) and ALV-5000/EPP Multiple tau Digital Correlator (ALV-Gmbh, Langen, Germany). The autocorrelation functions were acquired at 90° for 60 s. For each sample we collected 10 autocorrelation functions. To allow convection to dampen, data collections started 20 min after the solution was introduced to the cuvette. From each autocorrelation function we determined

the average values of the cluster radius $R_2$ and cluster volume fraction $\varphi_2$. For this, we computed the intensity distribution function corresponding to each correlation function employing both the CONTIN inverse Laplace transform algorithm (6) and a modified cumulant method introduced in Li et al. (1). The intensity distribution functions contained two sharp peaks, for the protein monomers and clusters, respectively, each characterized with a delay time, $\tau_1$ and $\tau_2$, and amplitudes, $A_1$ and $A_2$. From the time $\tau_1$ we determined the protein diffusivity used to characterize the intermolecular interactions. From $\tau_2$ we determined the effective cluster radius $R_2$ employing the Stokes-Einstein equation (1, 3). The error bars shown on plots represent the standard deviations of these values. The viscosity of protein solutions used to evaluate $R_2$ was determined independently as described in Ref. (1) using OptiLink carboxylate-modified polystyrene particles with diameter 0.424 μm; the data are shown in Fig. S4.

To estimate the fraction $\varphi_2$ of the solution volume occupied by the cluster population, we use the amplitudes $A_1$ and $A_2$ (3)

$$\varphi_2 = \frac{A_2}{A_1}\frac{1}{P(qR_2)f(C_1)}\frac{(\partial n/\partial C_2)_T}{(\partial n/\partial C_1)_T}\left(\frac{\rho_1}{\rho_2}\right)^2\left(\frac{R_1}{R_2}\right)^3\varphi_1.$$

The shape factor of the clusters, assuming spherical shape, is

$$P(qR_2) = \left[\frac{3}{(qR_2)^3}\left(\sin(qR_2) - qR_2\cos(qR_2)\right)\right]^2.$$

The dimensionless quantity $f(C_1) = KC_1M_w/R_\theta$ ($K$ is an instrument constant, $M_w$ is the protein molecular mass, and $R_\theta$ is the Rayleigh ratio of the intensity of the light scattered at angle $\theta$ to the incident light intensity) accounts for intermolecular interactions between protein molecules and is determined by static light scattering as discussed below and in Ref. (7); the interactions between clusters are neglected because of their low concentration. The derivatives $(\partial n/\partial C_i)_{T,\mu}$ are the increments of the refractive index $n$ with the mass concentrations of the monomers $C_1$ and clusters $C_2$ ($C_2 = m_{cluster}n_2$, where $m_{cluster}$ is the average cluster mass and $n_2$ is the cluster number concentration); since most measurements of $\partial n/\partial C_1$ for different proteins fall in the range 0.1 – 0.2 (8), it is safe to assume that the ratio of the increments in the expression for $\varphi_2$ is of order unity. The quantities $\rho_1$ and $\rho_2$ are the protein densities in the monomers and in the clusters, respectively; we use $\rho_1$ = 1.18 g cm$^{-3}$ Refs. (9, 10). Since the clusters contain dense protein liquid, we assume $\rho_2$ = 0.500 g cm$^{-3}$ (11). Because only a small fraction of the protein transfers to the clusters, the monomer volume fraction $\varphi_1 = C_1/\rho_1$.

The uncertainties in the determinations of $R_2$ and $\varphi_2$ are mostly due to the noise inherent in the DLS correlation functions. The effects of the noise on the accuracy of the method was analyzed in ref. (3). It was found that if the ratio $A_2/A_1 > 0.10$, the error in determination of $\tau_2$, and, correspondingly, of $R_2$, is < 10 % and that of $A_2/A_1$ is < 20 %. The $A_2/A_1$ ratios in Fig. S2 are in the range 0.1 – 1. Thus the greatest uncertainty in the determination of $R_2$ is 10 % and of $\varphi_2$ (in which the uncertainty in $R_2^{-3}$ is added to that of $A_2/A_1$), 50 %.

**Static light scattering (SLS).** To characterize the pairwise interactions in protein solutions we performed static light scattering measurements on the same device used for DLS. The scattered intensity was collected at 90° and the results are shown as Debye plots. For molecules in the dilute solution regime the simplified scattering equation is

$$\frac{Kc}{R_\theta} = \frac{1}{M_w} + 2B_2c,$$

where $R_\theta = \frac{I_\theta}{I_0}$ is a Rayleigh ratio of the scattered to the incident light intensity, $c$ is the protein concentration, $K = \frac{1}{N_A}\left(\frac{2\pi n_0}{\lambda^2}\right)^2\left(\frac{dn}{dc}\right)^2$ is an optical constant, $N_A$ is the Avogadro number, $n_0 = 1.331$ is the refractive index of the solvent at the wavelength of the laser beam, assumed to be equal to that of

water, $\frac{dn}{dc} = 0.199 \pm 0.003\ ml\ g^{-1}$ is the refractive index increment of the solutions, Fig. S6. This parameter was determined for each solution composition, using a Brookhaven differential refractometer operating at $\lambda = 620$ nm and calibrated with KCl solutions in water 25°C.

Each data point in the Debye plot is an average value of six measurements at identical conditions. The vertical and horizontal error bars represent the standard deviations of the intensity and protein concentration measurements, respectively. If the correlation function taken in parallel indicated the presence of clusters, we subtracted the intensity scattered by the clusters from the total and used the difference to evaluate the Rayleigh ratio.

**Brownian microscopy.** We use Nanosight LM10-HS microscope (Nanosight Ltd) to examine the Brownian motion of individual clusters in the tested solutions. We loaded a solution sample in a thermostatically controlled cuvette of volume ~ 0.3 ml and depth 0.5 mm. A green laser beam with wavelength 532 nm passes through the solution. All species in the solution scatter the incident light. The intensity scattered by a cluster is $(\frac{R_2}{R_1})^6 \approx 30^6 \approx 7.3 \times 10^6$-fold greater than that scattered by a monomer ($R_1$ is the monomer radius), so the clusters are well seen on the background of monomers. A 20× lens transfers the entire picture to a sensitive CMOS camera that records a movie of clusters undergoing Brownian motion. The rate of movie acquisition depends on the camera settings; in our experiments it was about 25 fps. Each frame of the movie is an image of clusters as bright white spots on a dark background. The accompanying software package determines the center of these spots in each frame of the movie and builds contiguous cluster trajectories. The cluster diffusivity is obtained from the slope of the dependence of the mean squared displacement on lag time. The cluster radius $R_2$ is evaluated from the Stokes-Einstein equation using viscosity values determined as discussed above. The number of cluster spots in a frame (using the focal depth of 5 μm) yields the cluster concentration.

We carefully matched the movies recorded by the Nanosight device with the data file that it outputs. We found that objects recorded for times shorter than 1 s are interference spots from two or more clusters tracked for significantly longer times. This observation is supported by the estimate that a cluster with diffusivity $D_2 \approx 10^{-12}$ $m^2s^{-1}$ would be detectable in a focal plane with depth 5 μm for about 25 s. We did not consider them as parts of the cluster population in the determination of the cluster parameters.

**Numerical modeling.** The total free energy of the protein-protein interactions consists of three distinct contributions: the Coulomb interaction (subject to the Debye screening by the mobile ions in the solution), short-range attraction (due to dispersion and, possibly, other interactions), and steric repulsion. The protein-protein interaction is assumed to be fully pairwise. To facilitate sampling of mutual orientations of two molecules, we model a protein molecule as a dielectric sphere, as in the Kirkwood-Tanford model (12). The sphere radius is chosen at 1.7 nm so that its volume matches that of an actual lysozyme molecule. A charged residue is represented by 1, 2, or 3 point charges depending on the number of distinct charged atoms in the residue. For instance, an $(NH_3)^+$ group is represented as three point charges of +1/3. The charges are located at a depth 0.15 nm beneath the surface. The latitude and longitude of each charge, with respect to the center of mass, are set equal to those in the actual protein molecule.

The Coulomb contribution to the overall protein-protein interaction represents the totality of the electrostatic interactions between the charges on the protein molecules. The latter interactions are estimated using the Debye-Hückel approximation additionally modified to account for the effects of the dielectric discontinuity at the protein-solvent interface: we adopt ε = 2 inside and ε = 78 outside the protein molecule, as in water. In addition, we partially account for the possibility that the $pK_a$ value of a surface residue is affected by the proximity of charges on the other protein molecule. For the four residues closest to the midpoint between the molecules, two on each molecule, the charges on the

residues are determined self-consistently, so as to include, for instance, the possibility of deprotonation of a positively charged residue facing another positively charged residue. The charges on the other residues are assumed to be equal to those on an isolated protein molecule. The pH of the solvent is set at 7.8. The temperature of the solution is 22°C.

We model the effective potential stemming from the non-Coulomb interactions, $E_{mol}$, by a functional form that smoothly interpolates between the known value of the van der Waals attraction for two polarizable spheres, at larger distances, and a short-range interaction between surface residues modeled here by a modified Lennard-Jones type interaction with adjustable parameters.

$$E_{mol} = \begin{cases} E_{>}, r_s > r_2 \\ E_{<}, r_s < r_1 \end{cases},$$

$$E_{>} = -\frac{A_H}{12}\left(\frac{1}{(x+1)^2} + \frac{1}{(x^2+2x)} + 2ln\left(\frac{x^2+2x}{(x+1)^2}\right)\right),$$

$$E_{<} = 4\varepsilon\left[\left(\frac{\sigma}{r_s+\delta}\right)^{2\alpha} - \left(\frac{\sigma}{r_s+\delta}\right)^{\alpha}\right],$$

where $r_s$ the distance between protein surfaces and $x = r_s/2R_p$ is that same distance divided by the sphere diameter. $A_H$ is the Hamaker constant, whose numerical value for lysozyme has been estimated at $3.1\ k_BT$ (13, 14). The quantities $\varepsilon, \sigma, \delta, r_1$ and $r_2$ are adjustable parameters. At $r_1 < r_s < r_2$, a fifth-degree polynomial is used to smoothly patch the long-range and short-range portions of $E_{mol}$—$E_{>}$ and $E_{<}$, respectively, so that the derivatives of order two and lower are continuous.

The parameters in the equation above are fixed by (a) stipulating that the curvature at the minimum of the binding potential matches its typical value for two solvated residues, and (b) tuning the depth of the potential so as to match the resulting second virial coefficient to its experimental value measured at one specific value of the ionic strength, specifically 313.13 mM in this work. The resulting values of the parameters for the molecular energy are given in Supplementary Table S1, which corresponds to the table in the Supplementary Information in Ref. (2).

The potential of mean force $E_{PMF}$ between two protein molecules is computed as the sum of the full Coulomb interaction and $E_{\mathrm{mol}}$.

The osmotic second viral coefficient, $B_{22}$, is computed as (15, 16):

$$B_{22} = -\frac{2\pi N_A}{M^2}\langle\int_0^{\infty}(e^{-E_{\mathrm{PMF}}/k_BT} - 1){r_c}^2 dr_c\rangle$$

where $M$ is the protein mass, and $r_c \equiv r_s + 2R_p$ is the distance between the proteins' centers of mass.

**Supplementary Table S1.** Values of the parameters of the intermolecular interactions.

| Variable | Value | Variable | Value |
|---|---|---|---|
| $\alpha$ | 12 | $c_1$ | 16.32 |
| $\sigma$, nm | 1.64 | $c_2$ | 90.04 |
| $r_1$, nm | 0.33 | $c_3$ | -186.64 |
| $r_2$, nm | 0.73 | $c_4$ | 173.09 |
| $c_0$ | -0.25 | $c_5$ | -60.31 |

## References

1. Li, Y., V. Lubchenko, and P. G. Vekilov. 2011. The Use of Dynamic Light Scattering and Brownian Microscopy to Characterize Protein Aggregation. Rev. Sci. Instr. 82:053106
2. Chan, Ho Y., V. Lankevich, Peter G. Vekilov, and V. Lubchenko. 2012. Anisotropy of the Coulomb Interaction between Folded Proteins: Consequences for Mesoscopic Aggregation of Lysozyme. Biophysical Journal 102:1934-1943.
3. Pan, W., O. Galkin, L. Filobelo, R. L. Nagel, and P. G. Vekilov. 2007. Metastable mesoscopic clusters in solutions of sickle cell hemoglobin. Biophys. J. 92:267-277.
4. Aune, K. C., and C. Tanford. 1969. Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. I. Dependence on pH at 25°. Biochemistry 8:4579-4585.
5. Roxby, R., and C. Tanford. 1971. Hydrogen ion titration curve of lysozyme in 6 M guanidine hydrochloride. Biochemistry 10:3348-3352.
6. Provencher, S. W. 1982. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. Comp. Phys. Communications 27:229-242.
7. Pan, W., P. G. Vekilov, and V. Lubchenko. 2010. The origin of anomalous mesoscopic phases in protein solutions. J. Phys. Chem. B 114 7620-7630.
8. Huglin, M. B. 1962. Specific refractive index increments. In Light scattering from protein solutions. M. B. Huglin, editor. Academic Press. 165-331.
9. Steinrauf, L. K. 1959. Preliminary x-ray data for some crystaline forms of β-lactoglobulin and hen egg-white lysozyme. Acta Crystallogr. 12:77-78.
10. Piazza, R., V. Peyre, and V. Degiorgio. 1998. Sticky hard spheres model of proteins near crystallization: a test based on the osmotic compressibility of lysozyme solutions. Phys. Rev. E 58:R2733-R2736.
11. Petsev, D. N., X. Wu, O. Galkin, and P. G. Vekilov. 2003. Thermodynamic functions of concentrated protein solutions from phase equilibria. J. Phys. Chem. B 107:3921-3926.
12. Tanford, C., and J. G. Kirkwood. 1957. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. Journal of the American Chemical Society 79:5333-5339.
13. Beretta, S., G. Chirico, and G. Baldini. 2000. Short-Range Interactions of Globular Proteins at High Ionic Strengths. Macromolecules 33:8663-8670.
14. Nir, S. 1977. Van der Waals interactions between surfaces of biological interest. Progress in Surface Science 8:1-58.
15. Allahyarov, E., H. Löwen, J. Hansen, and A. Louis. 2003. Nonmonotonic variation with salt concentration of the second virial coefficient in protein solutions. Physical Review E 67.
16. Neal, B. L., D. Asthagiri, O. D. Velev, A. M. Lenhoff, and E. W. Kaler. 1999. Why is the osmotic second virial coefficient related to protein crystallization? J. Cryst. Growth 196:377-387.