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

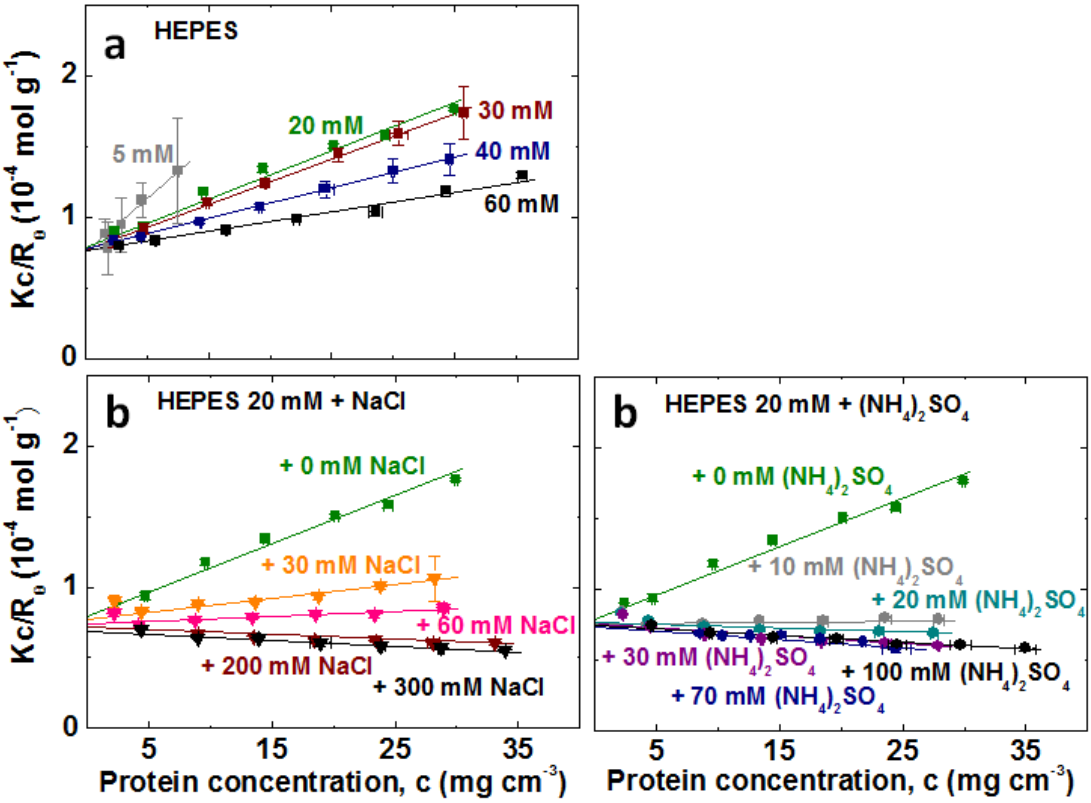

Figure S1. Characterization of the intermolecular interaction in dilute lysozyme solutions at increasing ionic strength. Debye plots Kc/R_θ (c) (K, instrument constant; c, lysozyme concentration; $R_\vartheta = I_\theta/I_0$, Rayleigh ratio of the intensity of light scattered at angle $\theta = 90^o$ 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.

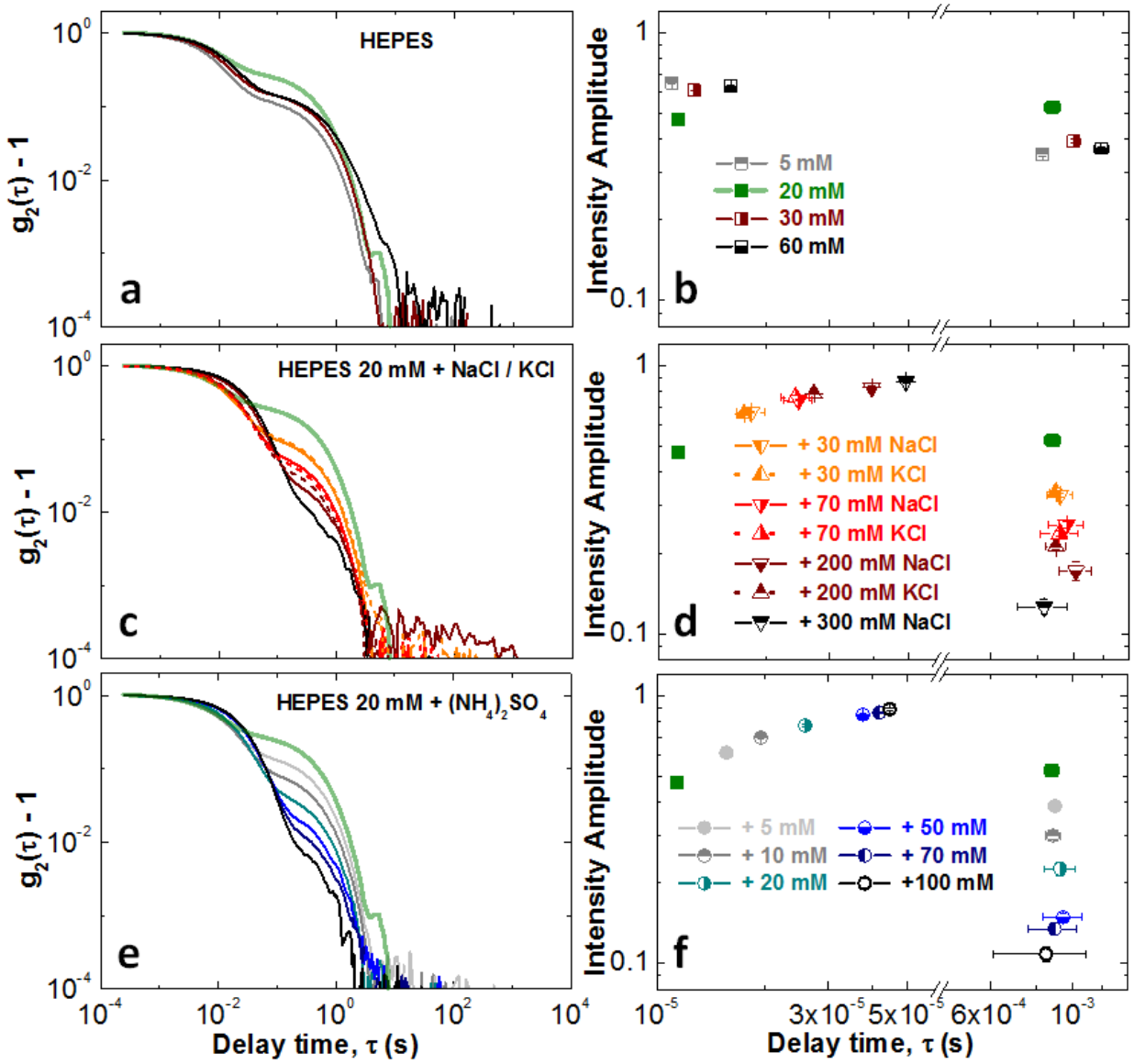

Figure S2. Monomers and clusters in lysozyme solutions at varying ionic strength and salt identity, characterized by dynamic light scattering (DLS). All solutions contain 100 mg ml^{-1} lysozyme in HEPES buffer at pH = 7.8. **a, c, e** Normalized autocorrelation functions $g_2 - 1$. The autocorrelation 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.

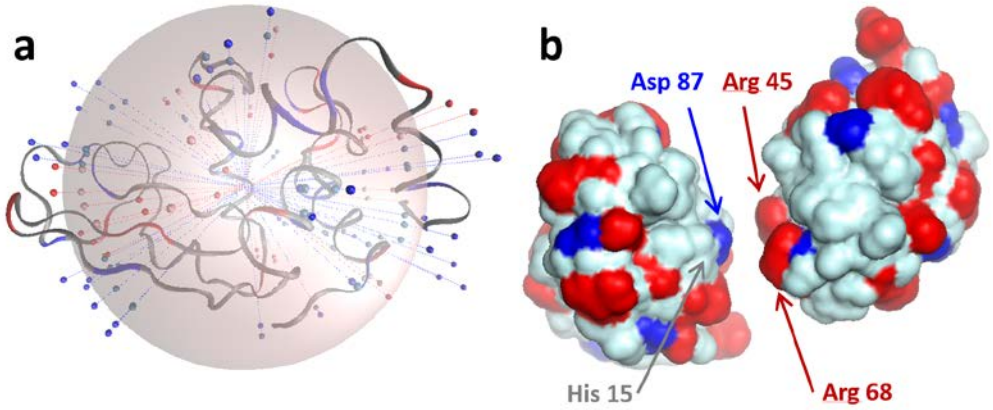

Figure S3. The charged groups on the surface of a lysozyme molecule. PBD structure file 2VB1 was used. The positive and negative surface charges are shown in red and blue, respectively. **a**, We represent a lysozyme molecule (whose peptide chain is shown here as a ribbon) as a sphere with radius 1.7 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.

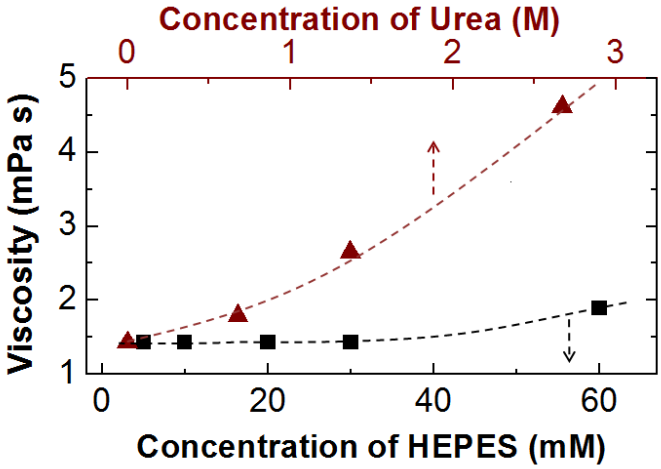

Figure S4. The viscosity of 100 mg ml^{-1} lysozyme solutions hosting the protein-rich clusters. Viscosity was determined by dynamic light scattering from the characteristic diffusion time using OptiLink carboxylate-modified polystyrene microparticles with diameter 0.424 μm suspended in 100 mg ml^{-1} 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.

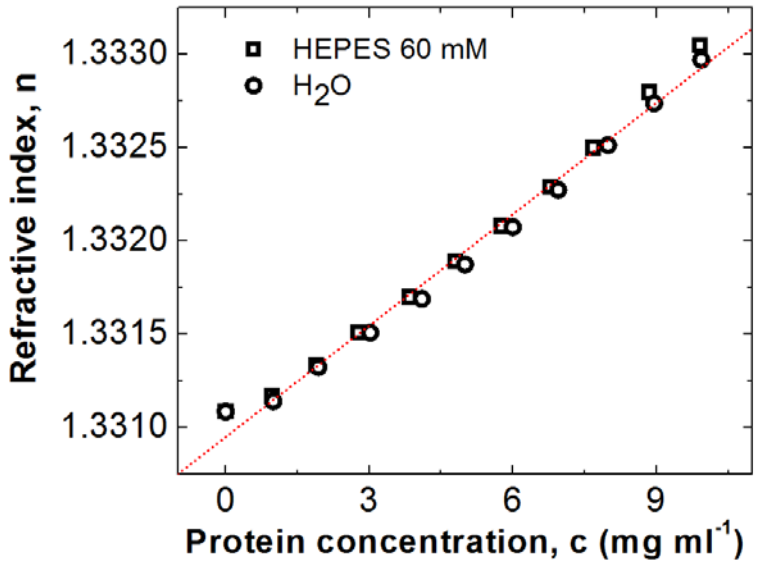

Figure S5. The refractive index increment *dn/dc* of lysozyme solutions. *dn/dc* was measured in two solvents, in water, where the ionic strength $I \cong 0$ mM, and in 60 mM HEPES at pH = 7.8, where I = 40 mM. Measurements did not reveal significant difference between the two solvents: both data sets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

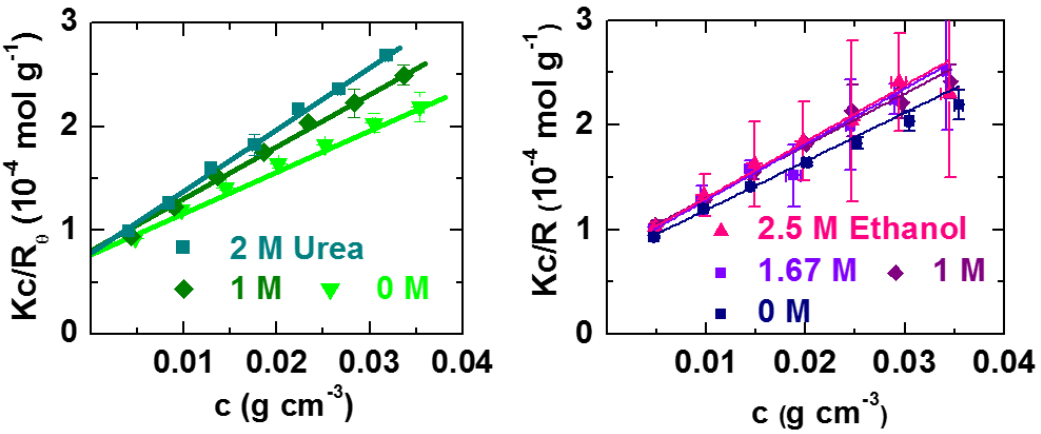

Figure S6. Characterization of the intermolecular interaction in dilute lysozyme solutions in the presence of urea and ethanol at concentrations indicated in the plots. Debye plots used to determine B_2. $R_\theta = \frac{I_\theta}{I_0}$ is the Rayleigh ratio of the scattered at angle θ = 90° to the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

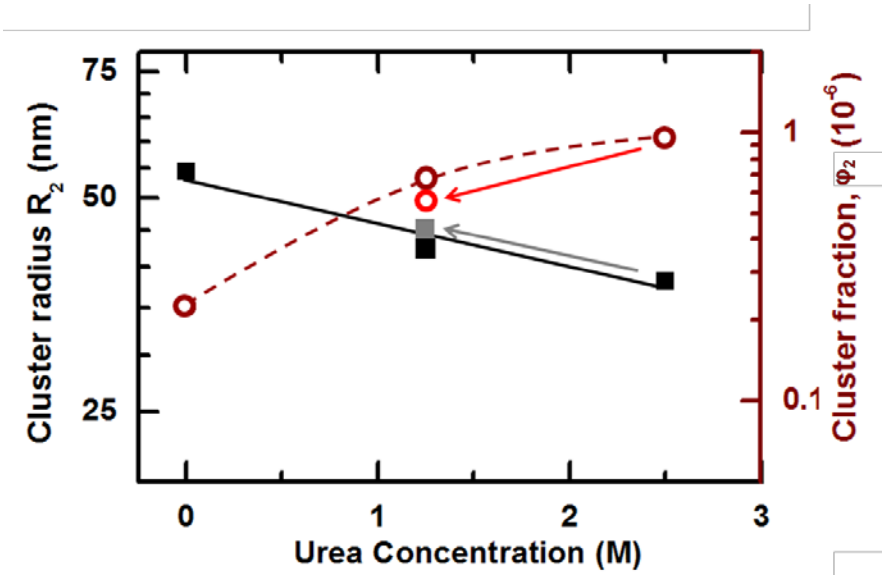

Figure S7. The response of the cluster radius R_2 (left ordinate, closed symbols) and volume fraction φ_2 (right ordinate, open symbols) to increasing or decreasing concentrations of urea in a 100 mg ml^{-1} lysozyme solutions in 20 mM HEPES at pH = 7.8, in which the ionic strength *I* = 13.3 mM. The observed trends are identical to those in Fig. 6 c; differences in values of R_2 and φ_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 φ_2 are very close to the other data pair for the same urea concentration, indicating that cluster formation is reversible.

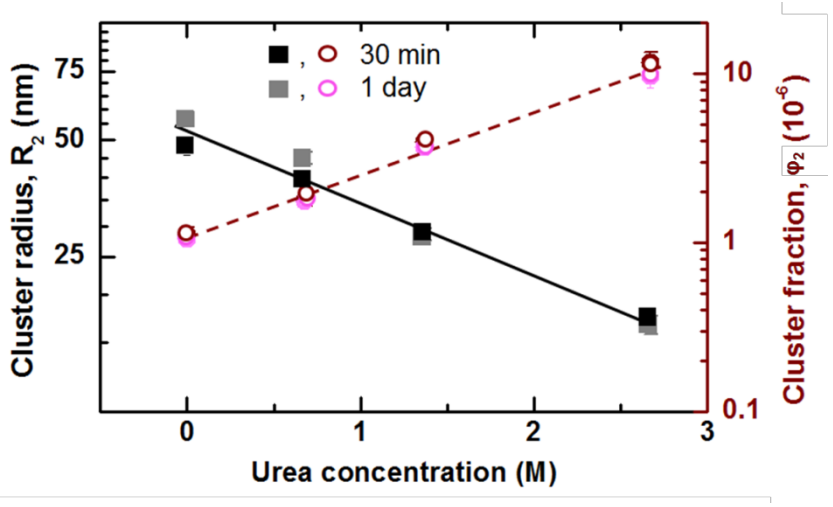

Figure S8. The consistency of the radius R_2 and volume fraction φ_2 of the clusters in the presence of urea. Solutions were characterized 30 min and 1 day after preparation. The data for 30 min are from Fig. 6 c.

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 ϕ_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 ϕ_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 ϕ_2 and this increase partially masks the response of D_1^{conc} and ϕ_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 φ_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, τ_1 and τ_2, and amplitudes, A_1 and A_2. From the time τ_1 we determined the protein diffusivity used to characterize the intermolecular interactions. From τ_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 φ_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_θ is the Rayleigh ratio of the intensity of the light scattered at angle θ 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 φ_2 is of order unity. The quantities ρ_1 and ρ_2 are the protein densities in the monomers and in the clusters, respectively; we use ρ_1 = 1.18 g cm^{-3} Refs. (9, 10). Since the clusters contain dense protein liquid, we assume ρ_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 φ_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 τ_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 φ_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_{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
α	12	c_1	16.32
σ, 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.