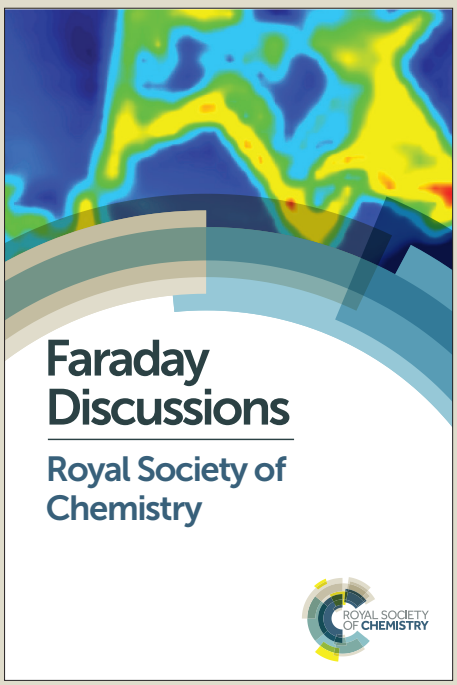
Faraday Discussions

Accepted Manuscript

This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting. All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: http://rsc.li/fd-upcoming-meetings



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

Recent advances in the understanding of two-step nucleation of protein crystals

Maria A. Vorontsova,[a] Dominique Maes,[b] and Peter G. Vekilov*[a,c]
DOI: 10.1039/b000000x

The two-step mechanism of nucleation of crystals in solutions posits that the formation of crystal nuclei occurs within structures of extended lifetimes, in which the nucleating solute is at high concentration. The validity of this mechanism has been demonstrated for proteins, small-molecule organic and inorganic materials, colloids, and polymers. Due to large molecule sizes, proteins are an ideal system to study the details of this nucleation pathway, in particular the formation mechanisms of the nucleation precursors and the associated physico-chemical rules. The precursors of protein crystal nuclei are protein-rich clusters of sizes ~100 nm that contain 10,000 – 100,000 molecules and occupy less than 10^{-3} of the total solution volume. Here we demonstrate, using oblique illumination microscopy, the liquid nature of the clusters of the protein lysozyme and reveal their inhomogeneous structure. We test a hypothesis put forth by theory that clusters primarily consist of transient protein oligomers. For this, we explore how varying the strength of Coulomb interaction affects the cluster characteristics. We find that the clusters size is insensitive to variations of pH and ionic strength. In contrast, the addition of urea, a chaotropic agent that leads to protein unfolding, strongly decreases the cluster size. Shear stress, a known protein denaturant, induced by bubbling of the solutions with an inert gas, elicits similar response. These observations support partial protein unfolding, followed by dimerization, as the mechanism of cluster formation. The amide hydrogen/deuterium exchange, monitored by nuclear magnetic resonance, highlights that lysozyme conformational flexibility is a condition for the formation of the protein-rich clusters and facilitates the nucleation of protein crystals.

Introduction

Protein structures are essential for understanding the fundamental processes in biology, are pivotal to the success of rational drug design, and enable numerous other laboratory and industry applications. Hence, structural genomics/proteomics projects around the world have invested considerable funds and effort in their

[a] *Department of Chemical and Biomolecular Engineering, University of Houston, Houston, Texas, 77204-4004, USA. E-mail: vekilov@uh.edu*
[b] *Structural Biology Brussels, Flanders Institute for Biotechnology (VIB), Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel, Belgium*
[c] *Department of Chemistry, University of Houston, Houston, Texas, 77204-5003, USA*

determination. The most widely used method of protein structure determination is X-ray crystallography[1] and obtaining high quality crystals is a crucial prerequisite for progress in this field.[2-5] The difficulty of obtaining crystals of such complex molecules limits the success of these projects and brings the need for novel approaches in protein crystallization, based on understanding of the mechanisms and its constituent processes.[6-8]

Besides structural biology, the formation of protein crystals and similar ordered arrays of folded proteins are of interest for several fields of science and technology. They underlie several human pathological conditions. An example is the crystallization of hemoglobin C and the polymerization of hemoglobin S that cause, respectively, the CC and sickle cell diseases.[9-12] Another area which relies on protein crystals is pharmacy: crystal dissolution at a controlled rate is used to achieve sustained release of medications, such as insulin, interferon-α, or the human growth hormone.[13-17] Work is underway on the crystallization of other therapeutically-active proteins, e.g., antibodies for foreign proteins, which can be dispensed as a microcrystalline preparation. If the administered dose consists of a few equidimensional crystallites, steady rates of medication release can be maintained for longer periods than for doses comprised of many smaller crystallites.

In all of these areas, control of crystallisation hinges on the ability to achieve and control nucleation, the first step of crystallization. Nucleation determines the main properties of the crystal population, including the crystal polymorph, the number of crystals, and their size and size distribution. The nucleation outcome favored in classical crystallography is to have a population consisting of one large crystal of a stable and robust polymorph, or, failing that, of several well-separated crystals of similar sizes and of a single polymorph. In the novel method of femtosecond X-ray protein nanocrystallography, which relies on crystals as small as 200 nm,[18] the need to grow the crystals after they have nucleated is nearly eliminated and nucleation emerges as the sole process to be controlled.

A novel outlook on protein nucleation in solution suggests that the nucleation of crystals is preceded by the formation of clusters of mesoscopic size composed of dense protein liquid.[19-23] Crystal nuclei then form inside the clusters.[24-26] Evidence of the action of this two-step mechanism has been provided for protein crystals[27, 28] and sickle cell anemia fibers,[29] amyloid fibrils,[30, 31] crystals of small-molecule organics,[25, 32, 33,34] colloids,[35-37] biominerals,[38-40] polymers,[41] and other substances. Evidence from multiple studies has indicated that the two-step mechanism is broadly applicable to protein crystallization.[27]

The likely precursor to crystal nuclei are protein-rich clusters. They have been observed in solutions of several proteins: lysozyme,[42, 43] hemoglobin A and S,[44] glucose isomerase,[45] lumazine synthase,[46, 47] insulin. The crucial role of the protein-rich clusters as locations of protein crystal nucleation has been supported by experimental evidence with hemoglobin S,[21, 29, 48] glucose isomerase,[45] and lysozyme;[27, 49] crystal nuclei forming inside similar mesoscopic clusters have been directly observed for colloids,[36] which are larger and move slower than most molecules, and an ingeniously chosen organic system.[33] The size of the protein clusters varies from several tens to several hundreds of nanometers and their total volume fraction remains less than 10^{-3}. Their extended lifetimes indicate that the clusters are not concentration fluctuations. It was shown that these clusters are present even in homogeneous region of protein phase diagram[42, 43] and their properties, average size and total population volume, do not undergo any singularity

upon crossing the liquidus.[21],[20] It was demonstrated that the cluster population volume responds to varying protein concentration in exact corresponds with Boltzmann predictions based on the free energy cost of regions of concentration as high as that in the clusters.[23, 43, 48] This observation indicates that the clusters are in equilibrium with the solution and excludes the possibility that the observed clusters represent aggregates of irreversibly denatured protein. On the other hand, the cluster size is independent of protein concentration. The decoupled behaviors of the cluster size and population volume are in stark contrast with the classical phase transition trends and highlight the unusual nature of the clusters. Experimental data and theoretical analyses that address the mechanisms determining the cluster size concluded that the size is determined by the dynamics of formation, diffusion and decay of protein oligomers of a limited lifetime. Hence, such oligomers are required for clusters' existence.[43] The hypothesized oligomers are not expected to be bound by crystal contacts and, hence are not crystal embryos. We hypothesize that their existence enables the clusters, which, in turn, facilitate crystal nucleation by lowering the free surface energy that limits the growth of the nuclei.

Here we address several open questions related to the oligomer mechanism of cluster formation: about the roles of electrostatic and water-structuring interactions, the liquid nature of the clusters, first suggested in previous work,[19-21] the uniformity of the cluster structure throughout their volume, and the significance of partial protein unfolding for cluster formation. We find that the classical interplay between electrostatic and hydrophobic interactions is significantly shifted to the latter ones. We employ a novel method of oblique illumination scattering microscopy and demonstrate the liquid nature of the clusters and suggest that they have inhomogeneous structure with an inside core of high viscosity. We show that partial protein unfolding and hydrophobic stabilization appear to be the major factors contributing to oligomerization and cluster formation. The results presented here suggest that fine-tuning of water-structuring interactions promises to be a powerful tool to control clusters population and, in view of the cluster role in protein crystal nucleation, enhance or suppress the nucleation process.

For these tests we chose the protein lysozyme because of its high isoelectric point, pI = 11.35,[50] which results in a considerable molecular net positive charge in the range of tested pH (from $+15\bar{e}$ at pH 3.8 to $+8\bar{e}$ at pH 7.8). This high charge amplifies the significance of electrostatic forces in cluster formation. Furthermore, previous studies with lysozyme have revealed that lysozyme forms well detectable clusters in the homogeneous region of the lysozyme phase diagram[43] and this allows the study of cluster characteristics and behaviors before crystal nucleation occurs.

We used the second osmotic virial coefficient B_2 to characterize the response of lysozyme in the solution to modified electrostatic and water-structuring interactions. B_2 is an azimuthally and spatially averaged characteristic of the pairwise interaction potential, revealed the attenuation of intermolecular repulsion with increasing ionic strength. We used static light scattering to determine B_2.[43] The response of the cluster population was characterized in terms of average cluster size R_2 and cluster population volume, expresses as a fraction of the total solution volume ϕ_2; R_2 and ϕ_2 were determined by dynamic light scattering.[20, 21, 23, 42, 43] We employed Brownian and oblique illumination microscopies to characterize the cluster phase state.

Materials and methods

Solution preparation

We used chicken egg white lysozyme (Affymetrix); KCl (Fisher); $(NH_4)_2SO_4$ and HCl, CH_3COOH (Fisher Scientific); NaCl and KOH (Mallinckrodt Chemicals); D_2O (Sigma-Aldrich); HEPES (Fisher and Calbiochem); we observed no difference between solutions of HEPES from the two companies.

Unless specified otherwise, all experiments were carried out in 20 mM HEPES buffer (N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonate) at pH 7.8 (adjusted with KOH). Since the pK_a of the respective acid is 7.5, the ionic strength of the buffer is ~0.667× of the total HEPES concentration. We determined the lysozyme concentration with Beckman Coulter Du 800 Spectrophotometer using extinction coefficient $\varepsilon_{280\ nm} = 2.64$ ml mg^{-1} cm^{-1}.[51] Stock solutions of ~150 mg ml^{-1} lysozyme in HEPES buffer were dialyzed over two days (44 - 50 hours). After dialysis, the solutions were diluted to a desired concentration and, if necessary, the ionic strength was adjusted by adding salts: NaCl, KCl or $(NH_4)_2SO_4$. All experiments were carried out at 22°C. Prior to all measurements, the solutions were filtered through 0.22 μm PES syringe filters (Lightlabs).

Dynamic and static light scattering (DLS and SLS)

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. Data analysis was performed as in Ref. [44] The viscosity of protein solutions used to evaluate clusters' hydrodynamic radius R_2 was determined independently as described in Ref. [42] using OptiLink carboxylate-modified polystyrene microparticles ($d = 0.424\ \mu m$). The static light scattering data on monomer interactions were collected at 90°. If the correlation function taken in parallel indicated the presence of clusters, we subtracted from the total scattered intensity the scattered intensity due to the clusters. The second virial coefficient B_2 was determined from Debye plots. For the refractive index increment of the solutions we used the value $dn/dc = 0.199 \pm 0.003\ ml\ g^{-1}$ obtained by Brookhaven differential refractometer calibrated with KCl aqueous solutions at 25°C, $\lambda = 620\ nm$.

Brownian and oblique illumination microscopies

We employed Nanosight LM10-HS microscope (Nanosight Ltd) equipped with green laser with $\lambda = 532$ nm operating at temperature $T = 22 \pm 0.1^0$C to examine the Brownian motion of individual clusters in the tested solutions. The raw data of this method are movies of point spread functions of clusters undergoing Brownian motion. The rate of movie acquisition depends on the camera characteristics and settings, in our experiments it was ~ 25 fps. We determined the clusters' diffusivity by using a custom designed software package which analyzes individual cluster trajectories. The cluster radius R was calculated using the Stokes-Einstein equation accounting for previously determined viscosity. The number of clusters on each frame gave an estimate of the cluster concentration.

Oblique illumination microscopy (OIM) uses the same setup as Brownian

microscopy with a more powerful objective lens and a faster, sensitive camera (S-PRI, AOS Technologies AG). OIM records the point spread functions of individual clusters, whose diameter ($2R_2 \approx 150\,nm$) is comparable to the diffraction limit ($\lambda/2 \approx 250\,nm$). We used this method to monitor the time resolved behavior of individual clusters.

Results and discussion

The role of Coulomb interactions in cluster formation

Coulomb interactions are a core of biophysics. They determine protein 3D structure,[52-57] substrate binding,[58, 59] transitions between activated and deactivated states,[60, 61] signals transmission,[62] and others. Proteins structures were optimized by evolutionary design to prevent undesirable aggregation resulting from electrostatic interactions.[63, 64] Hence, Coulomb forces were tested for their role in the formation of the mesoscopic protein rich clusters.

The strength of the intermolecular Coulomb forces depends on the charge of protein monomers and the Debye screening from free ions in a solvent, i.e. the ionic strength I of a solution. We probe the influence of electrostatics on our system by varying the ionic strength from 3 to $313\,mM$ covering the physiological range of $100-200\,mM$. Low I values were obtained by varying the concentration of the HEPES buffer. For higher $I's$ we used NaCl, KCl and $(NH_4)_2SO_4$. Dimensionless values of this parameter ranged from $B_2 = 114$ at the lowest probed $I = 3\,mM$ to $B_2 = -20$ at $I = 313\,mM$, attaining the value of non-interacting hard spheres $B_2 = 4$ at ionic strength of $50-60\,mM$.

We display the response of the cluster size to modified electrostatic forces in Fig. 1. We observe that cluster radius R_2 is independent of the ionic strength of the solution and on the salt identity used to increase I. The cluster volume fraction φ_2 (data not shown) decreased five-fold as the ionic strength increased to $\sim 150\,mM$ and stayed constant at higher I values.

We modified the monomers charge by changing the solution pH from 3.8 to 7.8 and performed identical tests. The results (data not shown) in terms of that the second virial coefficient B_2 and diffusivity D reveal that intermolecular interactions are significantly affected by electrostatic forces. Similar to the behavior in response to increasing ionic strength, cluster volume fraction shows an apparent increase as the protein charge decreases at higher pH values, while the cluster size does not change within the experimental error.

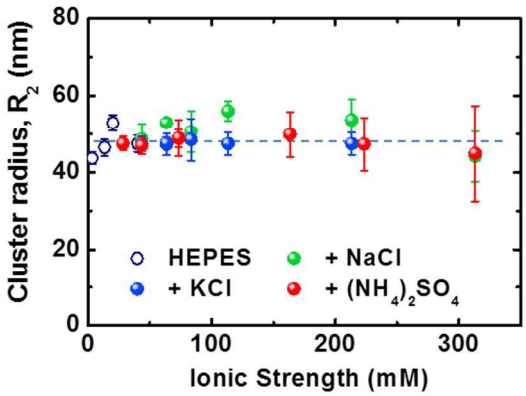


Fig. 1 The role of electrostatics in cluster formation: dependence of cluster radius R_2 on ionic strength. R_2 were evaluated from the cluster diffusion times, determined from the DLS correlation functions. The viscosity used in these evaluations was measured independently. The ionic strength of solutions was induced in four different ways: by changing concentration of HEPES buffer, by adding salts KCl, NaCl, $(NH_4)_2SO_4$.

The decoupled behaviors of R_2 and ϕ_2 at increasing I and pH indicate that these two characteristics of the cluster population 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 [23], while ϕ_2 reflects the high free energy cost of bringing together positively charged molecules [23, 43]. 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. These observations exclude a colloid aggregation scenario based on the balance of van der Waals attraction and Coulomb repulsion[65, 66] as the basis of the mesoscopic protein rich clusters.[66] Furthermore, these observations indicate that a mechanism relying on electrostatically bound oligomer could not underlie the mesoscopic clusters in lysozyme solutions.

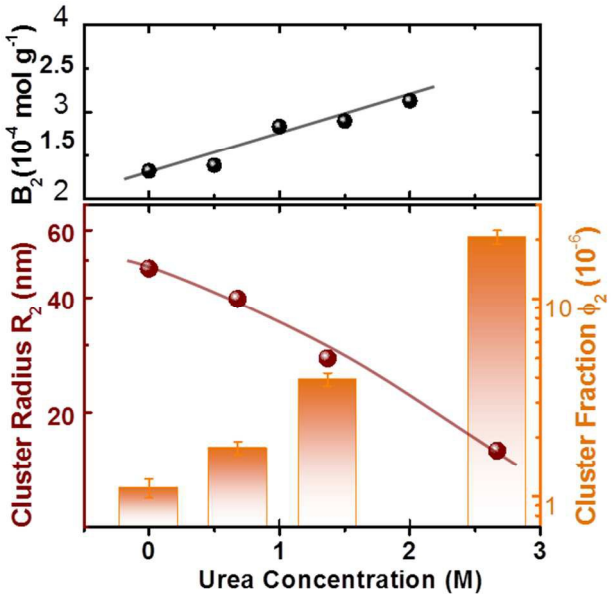
With many other studied proteins, clusters are observed at ionic strengths higher than 100 mM [19, 23, 29, 43, 67, 68], at which the Debye length is shorter than the molecular size and, hence, the lifetimes of electrostatically-bound oligomers would be insignificant. Hence, Coulomb-regulated colloid clustering and Coulomb-mediated oligomerization can be excluded as formation mechanisms of the mesoscopic clusters in solutions of these proteins.

Hydrophobic interactions in protein clusters formation

Water structuring and hydrophobic interactions are often a dominant force in protein aggregation.[69-72] These forces are short-range and may play an important role in the stabilization of the protein-rich clusters, in which the protein concentration is $\sim 450\ mg\ ml^{-1}$ and the intermolecular separations are shorter than 1 nm.[23]

Urea is a chaotropic agent, which accumulates at the protein peptide backbone and breaks the water structure.[73] This mechanism underlies urea's action as a universal protein denaturant that causes complete molecule unfolding at concentrations $7-8\ M$.[74-76] We employed urea concentrations up to $3\ M$,

significantly below the levels when complete protein unfolding occurs. The increase of the second virial coefficient B_2 from 38 to 52 upon increasing of urea concentration to 2 M, shown in Fig. 2 a, suggests that it causes an increase of intermolecular repulsion. This urea-induced increase of repulsion is consistent with breaking water structures around the protein solvent-exposed backbone segments and weakening the hydrophobic attraction at low protein concentrations.



X

Fig. 2 The role of hydrophobic interactions in cluster formation. (a) Variation of the pairwise intermolecular interactions in lysozyme solutions, characterized in terms of the second virial coefficient B_2 (left ordinate, closed symbols) as a function of urea concentration. (b) The response of the cluster radius R_2 (left ordinate, spheres) and volume fraction f_2 (right ordinate, columns) to urea concentration. The data reveal that addition of up to 2.7 M urea barely affects the pairwise intermolecular interactions, but significantly impacts clustering behaviors: R_2 decreases three-fold and f_2 increases 10-fold.

The cluster radius R_2 decreases three-fold at urea concentration 2.7 M while the clusters volume fraction decreases ten-fold. The decoupled behavior of these characteristics excludes the possibility of lysozyme denaturation and aggregation caused by urea and strongly supports the hypothesis of protein oligomerization contributing to cluster formation. Urea accumulates around solvent exposed segments of a protein backbone[77] shortening the lifetime of dynamic protein complexes. Following the theoretical model of cluster formation,[43] this leads to a smaller cluster radius R_2. At high protein concentration inside the clusters partial molecule unfolding caused by urea results in enhancement of attractive hydrophobic interactions. This stabilizes the concentrated protein phase and increases cluster volume fraction φ_2.

Bubbling of a lysozyme solution with an inert gas is known to decrease the protein activity by causing molecules unfolding at the hydrophobic solvent-gas interface[78] or because of the induce shear stress.[79] We performed experiments in which lysozyme solutions were bubbled with nitrogen at different flow rates ranging

from 0.2 to 0.5 L/min over the course of 2.5 hours. We did not register the change of pairwise intermolecular interactions, indicating that the majority of protein monomers was not affected by bubbling.

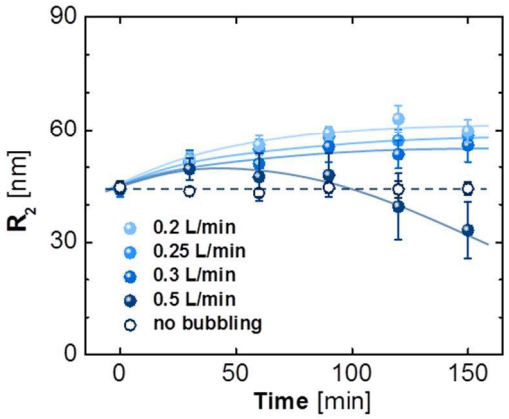


Fig. 3 The effects of shear flow on cluster formation. The response of the cluster radius R_2 to solution bubbling with N_2. The data shows that moderate bubbling rates (0.2 – 0.3 L/min) increase R_2; at 0.5 L/min shear flows have a negative effect on cluster formation and R_2 decreases with bubbling time.

We show the evolution of cluster radius R_2 in Fig. 3. The reference solution with no bubbling was analyzed to verify the steadiness of cluster radius throughout the experiment. We noticed the increase of R with increase of the gas flow rate from 0 to 0.3 L/min and increase of the bubbling time. This observation corresponds to a proposed cluster formation mechanism due to protein complex formation. Solution bubbling at moderate rates leads to protein unfolding, stabilization of protein oligomers, and increase of cluster radius. The decrease of cluster radius at the highest probed flow rate 0.5 L/min can result from the increase of shear rates in solutions which are not the subject of this discussion.

Molecular level view of oligomerization

The lysozyme folded chain, illustrated in Fig. 4 a, contains two stabilized regions, called domains, linked by a relatively short and flexible sequence, called hinge.[80] Domain swapping, schematically illustrated in Fig. 4 b, is a mechanism of oligomerization observed for numerous proteins.[81] If the attraction between the domains weakens, a molecule can partially unfold, exposing its hydrophobic interfaces between the domains. The new conformation of unfolded monomers is energetically unstable. If two (or more) partially unfolded monomers come to a close proximity, a domain from one molecule may bind to a reciprocal domain from the other partially unfolded molecule. As a result, a dimer (or higher order oligomer) may form. Each step in this process is reversible. The free energy of the oligomer is lower than for its constituent unfolded monomers but higher than for their native conformation.

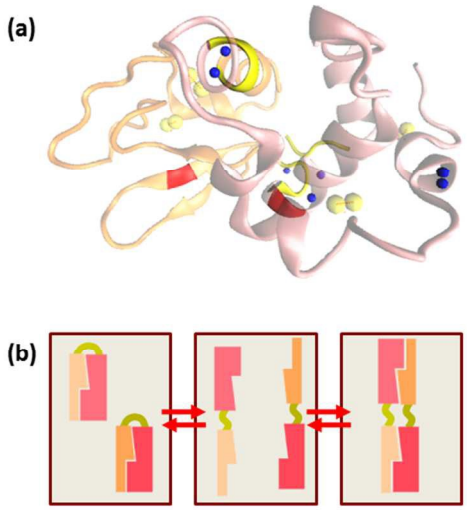


Fig. 4 Formation of protein oligomers. (a) Protein lysozyme structure (RCSB PDB: 2VB1). Two domains of the molecule are shown in pink and orange. The active center is highlighted in red, and the hinge in yellow. Yellow spheres represent the di-sulfide bridges. Blue spheres indicate the amide hydrogens of the aminoacid residues, revealed by NMR, potentially involved in oligomer formation. (b) Schematic representation of the domain swapping mechanism.

Even though there are no reports of domain swapped lysozyme oligomers, theoretical investigations reveal an internal molecular vibrational motion,[82, 83] which occurs between the two domains and involves the displacement of the hinge region. This process is crucial for lysozyme activity and may lead to partial molecule unfolding. Nuclear magnetic resonance (NMR) studies of lysozyme have indicated different hydrogen-deuterium exchange rates in native vs denatured states of the molecule [84] in D_2O-based solutions. These studies reveal that some hydrogens are more protected than others and the protection is strictly depends on the molecular conformation. We use this concept to test the involvement of partial protein unfolding in oligomerization leading to cluster formation.

Two protein solutions of high and low clusters volume fractions were prepared using D_2O-based HEPES buffer, $20\,mM, pH$ 7.8. The control sample was at concentration $25\,mg\,ml^{-1}$ at which cluster volume fraction is low ($< 10^{-7}$). To achieve a higher volume of the cluster population, we prepared a second similar solution with protein concentration of $200\,mg\,ml^{-1}$.

In both solutions, hydrogen-deuterium exchange was allowed to proceed for 1 and 72 hours before it was stopped by quenching to pH 3.8 with a solution of acetic acid in D_2O.[85]. Quenching diluted the concentration of the protein 8-fold. In the control solution, the concentration was brought back to $25\,mg\,ml^{-1}$ by centrifugation. The protein concentration in the tested solution was equal to that and it was not centrifuged.

The amino acid residues with unexchanged hydrogens yield signal in ^{1}H-^{15}N HSQC 2D spectra, based on the natural abundance of ^{15}N in the protein. The results on the rates of amide hydrogen exchange can be summarized as following: all non-amide and most amide hydrogens exchanged rapidly at both concentrations; several residues equally preserved their amide hydrogens in both control and cluster

containing solutions; no residue exhibited faster exchange rate in the control solution; several residues showed faster exchange rate in cluster-containing solution; these are $F34$, $N39$, $K97$, $Q121$, $A122$ at 1 hour, and $K33$, $I98$, $V92$ at 72 hour. We highlight these amide hydrogens in Fig. 4 a. The faster exchange in the high concentration solution is counterintuitive for a homogenously distributed protein and suggests that the protein in the clusters is partially unfolded; tests to eliminate the possibility that high concentration and not clusters contribute to the faster exchange are planned. Importantly, Fig. 4 a shows that the H-atoms exposed in the high concentration solution are near the hinges between the two lysozyme structural domains.[80] These results support the notion that partial unfolding of the domains followed by dimerization, as schematically depicted in Fig. 4 b, may be a part of this protein's cluster mechanism.

The liquid nature of protein-rich clusters

We employ oblique illumination microscopy (OIM) to monitor the dynamics of individual clusters.

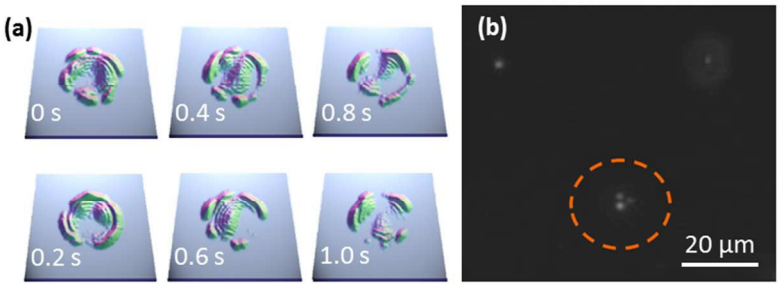


Fig. 5 The liquid clusters nature. (a) Intensity snapshots of a protein cluster revealed by oblique illumination microscopy at times indicated in the images. (b) Fluorescent microspheres observed through a bandpass filter, which allows the transmittance of only fluorescent light. A dashed circle indicates three particles entrapped inside a cluster.

Observations of individual clusters revealed that the point spread function of many clusters fluctuates, as shown in Fig. 5 a. Latex particles of a similar size did not exhibit the same behavior. Since particles with spherical symmetry do not show intensity fluctuations, we conclude that the protein clusters have complex asymmetrical shape. The cluster intensity pattern is a sum of point spread functions of its parts, which act as independent scatterers. The intensity fluctuations of the point spread functions indicate that the clusters are characterized by a complex non-spherical shape which may also be fluctuating. We determined the characteristic time of the fluctuations, ~ 60 ms. This time is much faster than the cluster rotation time, suggesting that the fluctuations reflect variations in cluster shape, i.e., the clusters are liquid. The shape variations are likely due to Brownian collisions with solvent molecules.

In another experiment we introduced fluorescent particles in a protein solution and placed a filter between the cuvette and the objective lens which obstructed the scattered light and only allowed fluorescent intensity to register. The majority of fluorescent particles diffused independently of each other. Several groups of

particles moved collectively. In Fig. 5 b we show three particles which exhibited this behavior. The separation between them was ~ 2 μm indicating that the particles were not bound to one another. After removing the fluorescence filter, we observed at their location a fluctuating intensity pattern, typical of a cluster. We conclude that the particles were trapped inside the cluster, confirming the liquid state of the clusters. We analyzed the evolution of the distances between the three fluorescent particles in the cluster and found that they do not change during the time of observation. This shows that the particles are embedded in a medium of high viscosity. Lysozyme solutions at high concentrations tend to exhibit gelation [24, 86] arresting the dynamics of the solution and increasing the viscosity. Thus, the observations in Fig. 5b suggest the presence of a viscous core inside the clusters.

Summary and conclusions

This investigation of the characteristics of the proteins-rich clusters in lysozyme solutions revealed several novel behaviors. We found that the Coulomb interactions do not significantly affect cluster populations: the cluster radius R_2 remains constant over the whole range of tested ionic strengths and pH, while the cluster volume fraction φ_2 slightly decreases. Experiments in which urea was added to the solutions revealed the importance of hydrophobic interactions for cluster formation. Bubbling experiments suggested the importance of partial protein unfolding in the clustering mechanism. Nuclear magnetic resonance tests located the regions of the lysozyme molecule that are potentially involved in partial unfolding during cluster formation, the interdomain hinges, and suggest that domain-swapped dimers may be the species that underlies cluster formation. Experiments with oblique illumination microscopy revealed the liquid nature of protein clusters and, for the first time, suggest inhomogeneous structure, with a viscous core, of clusters.

Acknowledgements

This work was supported by NSF (Grant MCB-1244568), and NASA (NNX14AD68G and NNX14AE79G).

References

1. K. Wütrich, *Nature Struct. Biol.*, 1998, **5**, 492-495.
2. M. F. Perutz and F. S. Mathews, *J Mol Biol*, 1966, **21**, 199-202.
3. S. R. Simon, W. H. Konigsberg, W. Bolton and M. F. Perutz, *J Mol Biol*, 1967, **28**, 451-454.
4. A. McPherson, *J. Crystal Growth*, 1991, **110**, 1-10.
5. Z. Derewenda, *Structure (Camb).* 2004, **12**, 529-535.
6. S.-H. Kim, D. H. Shin, J. Liu, V. Oganesyan, S. Chen, Q. S. Xu, J.-S. Kim, D. Das, U. Schulze-Gahmen, S. R. Holbrook, E. L. Holbrook, B. A. Martinez, N. Oganesyan, A. DeGiovanni, Y. Lou, M. Henriquez, C. Huang, J. Jancarik, R. Pufan, n.-G. Choi, J.-M. Chandonia, J. Hou2, B. Gold, H. Yokota, S. E. Brenner, P. D. Adams and R. Kim, *Journal of Structural and Functional Genomics*, 2005, **6**, 63–70.
7. R. C. Stevens and I. A. Wilson, *Science*, 2001, **293**, 519-520.
8. S. K. Burley, S. C. Almo, J. B. Bonanno, M. Capel, M. R. Chance, T. Gaasterland, D. Lin, A. Sali, F. W. Studier and S. Swaminathan, *Nature Genetics*, 1999, **23**, 151-157.

9. S. Charache, C. L. Conley, D. F. Waugh, R. J. Ugoretz and J. R. Spurrell, *J Clin Invest*, 1967, **46**, 1795-1811.
10. R. E. Hirsch, C. Raventos-Suarez, J. A. Olson and R. L. Nagel, *Blood*, 1985, **66**, 775-777.
11. W. A. Eaton and J. Hofrichter, in *Advances in protein chemistry*, eds. C. B. Anfinsen, J. T. Edsal, F. M. Richards and D. S. Eisenberg, Academic Press, San Diego, 1990, vol. 40, pp. 63-279.
12. P. Vekilov, *Brit. J. Haematol.*, 2007, **139**, 173-184.
13. J. Brange, *Galenics of Insulin*, Springer, Berlin, 1987.
14. M. L. Long, J. B. Bishop, T. L. Nagabhushan, P. Reichert, G. D. Smith and L. J. DeLucas, *J. Crystal Growth*, 1996, **168**, 233-243.
15. S. Matsuda, T. Senda, S. Itoh, G. Kawano, H. Mizuno and Y. Mitsui, *J. Biol. Chem*, 1989, **264**, 13381-13382.
16. S. Peseta, J. A. Langer, K. C. Zoon and C. E. Samuel, in *Annual Review of Biochemistry*, eds. C. C. Richardson, P. D. Boyer, I. B. Dawid and A. Meister, Annual Reviews, Palo Alto, 1989, vol. 56, pp. 727-778.
17. *US Patent Pat.,* 5,441,734., 1995.
18. H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Homke, C. Reich, D. Pietschner, L. Struder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kuhnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jonsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.-D. Schroter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson and J. C. H. Spence, *Nature*, 2011, **470**, 73-77.
19. O. Gliko, N. Neumaier, W. Pan, I. Haase, M. Fischer, A. Bacher, S. Weinkauf and P. G. Vekilov, *J. Amer. Chem. Soc.*, 2005, **127**, 3433-3438.
20. O. Gliko, W. Pan, P. Katsonis, N. Neumaier, O. Galkin, S. Weinkauf and P. G. Vekilov, *J. Phys. Chem. B*, 2007, **111**, 3106-3114.
21. W. Pan, O. Galkin, L. Filobelo, R. L. Nagel and P. G. Vekilov, *Biophys. J.*, 2007, **92**, 267-277.
22. P. G. Vekilov, W. Pan, O. Gliko, P. Katsonis and O. Galkin, in *Lecture Notes in Physics, vol. 752: Aspects of Physical Biology: Biological Water, Protein Solutions, Transport and Replication*, eds. G. Franzese and M. Rubi, Springer, Heidelberg, 2008, pp. 65-95.
23. W. Pan, P. G. Vekilov and V. Lubchenko, *J. Phys. Chem. B*, 2010, **114** 7620-7630.
24. O. Galkin and P. G. Vekilov, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 6277-6281.
25. B. Garetz, J. Matic and A. Myerson, *Phys. Rev. Lett.*, 2002, **89**, 175501.
26. P. G. Vekilov, *Crystal Growth and Design*, 2004, **4**, 671-685.
27. P. G. Vekilov, *Crystal Growth & Design*, 2010, **10**, 5007-5019.
28. Y. G. Kuznetsov, A. J. Malkin and A. McPherson, *J. Crystal Growth*, 2001, **232**, 30-39.
29. O. Galkin, W. Pan, L. Filobelo, R. E. Hirsch, R. L. Nagel and P. G. Vekilov, *Biophys. J.*, 2007, **93**, 902-913.

30. A. Lomakin, D. S. Chung, G. B. Benedek, D. A. Kirschner and D. B. Teplow, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 1125-1129.
31. R. Krishnan and S. L. Lindquist, *Nature*, 2005, **435**, 765-772.
32. J. E. Aber, S. Arnold and B. A. Garetz, *Phys. Rev. Lett.*, 2005, **94**, 145503.
33. K. Harano, T. Homma, Y. Niimi, M. Koshino, K. Suenaga, L. Leibler and E. Nakamura, *Nature Materials*, 2012, **11**, 877-881.
34. D. Erdemir, A. Y. Lee and A. S. Myerson, *Accounts Chem. Res.*, 2009, **42**, 621-629.
35. M. E. Leunissen, C. G. Christova, A.-P. Hynninen, C. P. Royall, A. I. Campbell, A. Imhof, M. Dijkstra, R. van Roij and A. van Blaaderen, *Nature*, 2005, **437**, 235-240.
36. J. R. Savage and A. D. Dinsmore, *Physical Review Letters*, 2009, **102**, 198302.
37. T. H. Zhang and X. Y. Liu, *Journal of Physical Chemistry B*, 2007, **111**, 14001-14005.
38. E. M. Pouget, P. H. H. Bomans, J. A. C. M. Goos, P. M. Frederik, G. de With and N. A. J. M. Sommerdijk, *Science*, 2009, **323**, 1455-1458.
39. D. Gebauer, A. Volkel and H. Colfen, *Science*, 2008, **322**, 1819-1822.
40. L. B. Gower, *Chemical Reviews*, 2008, **108**, 4551-4627.
41. J. F. Wang, M. Muller and Z. G. Wang, *Journal of Chemical Physics*, 2009, **130**.
42. Y. Li, V. Lubchenko and P. G. Vekilov, *Rev. Sci. Instr.*, 2011, **82**, 053106
43. Y. Li, V. Lubchenko, M. A. Vorontsova, L. Filobelo and P. G. Vekilov, *The Journal of Physical Chemistry B*, 2012, **116**, 10657-10664.
44. W. Pan, O. Galkin, L. Filobelo, R. L. Nagel and P. G. Vekilov, *Biophys J*, 2007, **92**, 267-277.
45. M. Sleutel and A. E. S. Van Driessche, *Proceedings of the National Academy of Sciences*, 2014, **111**, E546-E553.
46. O. Gliko, N. Neumaier, W. Pan, I. Haase, M. Fischer, A. Bacher, S. Weinkauf and P. G. Vekilov, *J Am Chem Soc*, 2005, **127**, 3433-3438.
47. O. Gliko, W. Pan, P. Katsonis, N. Neumaier, O. Galkin, S. Weinkauf and P. G. Vekilov, *J Phys Chem B*, 2007, **111**, 3106-3114.
48. V. Uzunova, W. Pan, V. Lubchenko and P. G. Vekilov, *Faraday Disc.*, 2012, **159** 87 - 104.
49. W. Pan, A. B. Kolomeisky and P. G. Vekilov, *J. Chem. Phys.*, 2005, **122**, 174905.
50. L. R. Wetter and H. F. Deutsch, *Journal of Biological Chemistry*, 1951, **192**, 237-242.
51. K. C. Aune and C. Tanford, *Biochemistry*, 1969, **8**, 4579-4585.
52. Y. Goto, L. J. Calciano and A. L. Fink, *Proc Natl Acad Sci U S A*, 1990, **87**, 573-577.
53. R. Fairman, K. R. Shoemaker, E. J. York, J. M. Stewart and R. L. Baldwin, *Proteins*, 1989, **5**, 1-7.
54. W. G. J. Hol, L. M. Halie and C. Sander, *Nature*, 1981, **294**, 532-536.
55. R. Fairman, K. R. Shoemaker, E. J. York, J. M. Stewart and R. L. Baldwin, *Biophys Chem*, 1990, **37**, 107-119.
56. L. Xiao and B. Honig, *Journal of Molecular Biology*, 1999, **289**, 1435-1444.
57. J. M. Thornton, *Nature*, 1982, **295**, 13-14.
58. R. H. Meltzer, E. Thompson, K. V. Soman, X. Z. Song, J. O. Ebalunode, T. G. Wensel, J. M. Briggs and S. E. Pedersen, *Biophys J*, 2006, **91**, 1302-1314.
59. C. F. Anderson and M. T. Record, *Annual Review of Physical Chemistry*, 1982, **33**, 191-222.
60. E. Ozkirimli, S. S. Yadav, W. T. Miller and C. B. Post, *Protein Sci*, 2008, **17**, 1871-1880.
61. A. M. Dean and D. E. Koshland, Jr., *Science*, 1990, **249**, 1044-1046.
62. L. Hemsath, R. Dvorsky, D. Fiegen, M. F. Carlier and M. R. Ahmadian, *Mol Cell*, 2005, **20**, 313-324.
63. J. P. K. Doye, A. A. Louis and M. Vendruscolo, *Physical Biology*, 2004, **1**, p9-p13.

64. J. P. K. Doye and W. C. K. Poon, *Current Opinion in Colloid & Interface Science*, 2006, **11**, 40-46.
65. F. Sciortino, S. Mossa, E. Zaccarelli and P. Tartaglia, *Phys. Rev. Lett.*, 2004, **93**, 055701.
66. S. B. Hutchens and Z.-G. Wang, *J. Chem. Phys.* , 2007 **127**, 084912
67. B. G. Mohr, C. M. Dobson, S. C. Garman and M. Muthukumar, *The Journal of Chemical Physics*, 2013, **139**, 121914.
68. M. Sleutel and A. E. Van Driessche, *Proceedings of the National Academy of Sciences of the United States of America*, 2014, **111**, E546-553.
69. R. D. Hills Jr and C. L. Brooks Iii, *Journal of Molecular Biology*, 2007, **368**, 894-901.
70. G. Favrin, A. Irbäck and S. Mohanty, *Biophysical Journal*, 2004, **87**, 3657-3664.
71. B. Raman, E. Chatani, M. Kihara, T. Ban, M. Sakai, K. Hasegawa, H. Naiki, M. Rao Ch and Y. Goto, *Biochemistry*, 2005, **44**, 1288-1299.
72. L. A. Munishkina, J. Henriques, V. N. Uversky and A. L. Fink, *Biochemistry*, 2004, **43**, 3289-3300.
73. M. Auton, L. M. F. Holthauzen and D. W. Bolen, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 15317-15322.
74. A. Das and C. Mukhopadhyay, *J Phys Chem B*, 2008, **112**, 7903-7908.
75. A. Caballero-Herrera, K. Nordstrand, K. D. Berndt and L. Nilsson, *Biophysical Journal*, 2005, **89**, 842-857.
76. L. Hua, R. Zhou, D. Thirumalai and B. J. Berne, *Proceedings of the National Academy of Sciences*, 2008, **105**, 16928-16933.
77. M. Auton, L. M. F. Holthauzen and D. W. Bolen, *Proceedings of the National Academy of Sciences*, 2007, **104**, 15317-15322.
78. M. Caussette, A. Gaunand, H. Planche, S. Colombié, P. Monsan and B. Lindet, *Enzyme and Microbial Technology*, 1999, **24**, 412-418.
79. I. B. Bekard, P. Asimakis, J. Bertolini and D. E. Dunstan, *Biopolymers*, 2011, **95**, 733-745.
80. J. A. McCammon, B. R. Gelin, M. Karplus and P. G. Wolynes, *Nature* 1976, **262**, 325-326.
81. M. P. Schlunegger, M. J. Bennett and D. Eisenberg, *Adv Protein Chem*, 1997, **50**, 61-122.
82. J. A. McCammon, B. R. Gelin, M. Karplus and P. G. Wolynes, *Nature*, 1976, **262**, 325-326.
83. A. De Simone, R. W. Montalvao, C. M. Dobson and M. Vendruscolo, *Biochemistry*, 2013, **52**, 6480-6486.
84. S. E. Radford, M. Buck, K. D. Topping, C. M. Dobson and P. A. Evans, *Proteins*, 1992, **14**, 237-248.
85. Y. Bai, J. S. Milne, L. Mayne and S. W. Englander, *Proteins: Structure, Function, and Bioinformatics*, 1993, **17**, 75-86.
86. M. Muschol and F. Rosenberger, *J. Chem. Phys.*, 1997, **107**, 1953-1962.