

**PHYSICOCHEMICAL ASPECTS OF THE TWO-STEP
MECHANISM OF NUCLEATION IN PROTEIN
SOLUTIONS**

A Dissertation

Presented to

the Faculty of the Department of Chemical and Biomolecular Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in Chemical Engineering

by

Maria A. Kaissaratos

May 2016

**PHYSICOCHEMICAL ASPECTS OF THE TWO-STEP
MECHANISM OF NUCLEATION IN PROTEIN
SOLUTIONS**

Maria Kaissaratos

Approved:

Chair of the Committee
Peter G. Vekilov, PhD
Francqui International Professor
John and Rebecca Moores Professor
Chemical and Biomolecular Engineering
Chemistry

Committee Members:

Vassiliy Lubchenko, PhD
Associate Professor
Chemistry

Jacinta C. Conrad, PhD
Assistant Professor
Chemical and Biomolecular Engineering

Patrick C. Cirino, PhD
Associate Professor
Chemical and Biomolecular Engineering

Anatoly B. Kolomeisky, PhD
Professor
Chemistry
Rice University

Suresh K. Khator, PhD
Associate Dean
Cullen College of Engineering

Michael P. Harold, PhD
Chair and M.D. Anderson Professor
Chemical and Biomolecular Engineering

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge various people without whom this thesis might not have been written. Firstly, I would like to thank my advisor Dr. Peter G. Vekilov whose help, support and guidance were so important in persuasion of my PhD degree. He continually encouraged me in regard to research and professional success and without his trust in his students and me, many of my accomplishments would not be feasible to achieve.

I wish to express my sincere thanks to Dr. Vassiliy Lubchenko for developing theoretical grounds of my projects together with his student Ho Yin Chan and for our numerous discussions and meeting resolving research obstacles and giving ideas to new experiments.

I also thank Dr. Jacinta C. Conrad for her desire to help and invaluable comments in light scattering research. I am extremely thankful for the opportunity she gave to work with her equipment which resulted in several chapters of my research.

I take this opportunity to than Dr. Robert Fox and his group who helped us with nuclear magnetic resonance studies.

I would like to thank the committee members of my dissertation: Dr. Patrick C. Cirino and Dr. Anatoly B. Kolomeisky for their time and kindest to be present at my defense.

I thank my group members who were always here to help and to support and with whom we had amazing times of getting together.

I acknowledge the grants for funding support which made my PhD possible, NSF (MCB-1244568, MCB-1518204), and NASA (NNX14AD68G and NNX14AE79G).

**PHYSICOCHEMICAL ASPECTS OF THE TWO-STEP
MECHANISM OF NUCLEATION IN PROTEIN
SOLUTIONS**

An Abstract

of a

Dissertation

Presented to

the Faculty of the Department of Chemical and Biomolecular Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in Chemical Engineering

by

Maria A. Kaissaratos

May 2016

ABSTRACT

Protein-rich liquid clusters exist in solutions of numerous proteins. They play the role of nucleation precursors of ordered solids of both folded proteins and partially misfolded chains. Examples include protein crystals, sickle-cell hemoglobin polymers, and amyloid fibrils. The clusters hold the key to the understanding and control of protein aggregation, and hence insights of their physical properties is needed for development of successful crystallization recipes.

We prove that protein clusters are not the nuclei of the dense liquid but rather represent a new phase which exists in homogeneous field of the protein phase diagram. With nuclear magnetic resonance method we find the regions of protein molecules flexibility, potentially participating in cluster formation. We prove that water structuring interactions and partial protein unfolding contribute to clustering. We show that common organic additives used in crystallization increase cluster volume fraction and surface area. The tests of insulin protein solutions explain why the two-step mechanism of nucleation is selected.

We develop a new spatial cross-correlation tracking method suitable for large ($> \lambda/2$) clusters. Monitoring of shape variations of intensity patterns of a single cluster indicates that protein clusters are liquid. We employ depolarized oblique illumination microscopy to study the nucleation process and we show that crystals of lysozyme and glucose isomerase proteins indeed nucleate within protein-rich liquid clusters. These are the first experiments of a direct observation of a two-step mechanism of nucleation in protein solutions.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	xi
LIST OF TABLES	xvi
CHAPTER 1. INTRODUCTION	1
CHAPTER 2. PRINCIPAL EXPERIMENTAL TECHNIQUES	17
2-1. Dynamic light scattering	18
2-2. Static light scattering	20
2-3. Brownian microscopy	22
CHAPTER 3. MOLECULAR LEVEL VIEW OF OLIGOMERIZATION IN PROTEIN SOLUTIONS	25
3-1. Materials and methods	28
3-2. Monomer flexibility as revealed by nuclear magnetic resonance	28
3-3. Conclusions	30
CHAPTER 4. TERMINAL SIZE IN CLUSTER EVOLUTION	31
4-1. Materials and methods	32
4-2. Results	32
4-3. Conclusions	34
CHAPTER 5. LACK OF DEPENDENCE OF THE SIZES OF THE MESOSCOPIC PROTEIN CLUSTERS ON ELECTROSTATICS	36

5-1. Materials and methods.....	37
5-2. Model system.....	43
5-3. Characterization of the intermolecular Coulomb forces.....	44
5-4. The effects of the Coulomb forces on the cluster population.....	51
5-5. Water-structuring interaction and partial protein unfolding.....	55
5-6. Purging of lysozyme solution with gases induces shear stress, crucial for lysozyme cluster formation.....	62
5-7. Conclusions	67
CHAPTER 6. ALTERNATIVE MECHANISMS OF CLUSTER FORMATION ON THE EXAMPLE OF PROTEIN GLUCOSE ISOMERASE.....	68
6-1. Materials and methods.....	69
6-2. Pairwise monomers interactions and solution viscosity.....	72
6-3. The role of partial protein unfolding and shear-stress influence on cluster formation in glucose isomerase solutions.....	73
6-4. Conclusions	76
CHAPTER 7. ORGANIC ADDITIVES IN CLUSTERS FORMATION.....	79
7-1. Materials and methods.....	84
7-2. Results and discussion.....	85
7-3. Conclusions	94
CHAPTER 8. TRACKING COMPLEX OBJECTS – PROTEIN CLUSTERS WITH TIME-DEPENDENT ASYMMETRIC INTENSITY PATTERNS	95
8-1. Materials and methods.....	97
8-2. Challenges of tracking micron-size objects.....	99

8-3. The spatial cross-correlation (SCC) method	103
8-4. The accuracy of the spatial cross-correlation algorithm.....	106
8-5. Comparison with other particle tracking methods.....	112
8-6. The frame rate and the length of the movie.....	114
8-7. Conclusions	115
CHAPTER 9. THE LIQUID NATURE OF PROTEIN-RICH CLUSTERS.....	118
9-1. Materials and methods.....	119
9-2. Intensity patterns of protein clusters are indicators of their liquid nature	120
9-3. Lysozyme liquid clusters can devour small-size particles and reveal their highly viscous core	124
9-4. Conclusions	125
CHAPTER 10. THE TWO-STEP MECHANISM OF NUCLEATION IN PORCINE INSULIN SOLUTIONS	127
10-1. Materials and methods.....	128
10-2. Insulin-rich liquid clusters help to explain the increase of nucleation rate	129
10-3. Conclusions	131
CHAPTER 11. THE TWO-STEP MECHANISM OF NUCLEATION – DIRECT OBSERVATION	132
11-1. Materials and methods.....	132
11-2. Nucleation of glucose isomerase and lysozyme crystals inside protein-rich clusters in a real time experiment	133
11-3. Conclusions	136

CHAPTER 12. CONSLUSIONS.....	137
REFERENCES	143

LIST OF FIGURES

Figure 1-1. The change of the system free energy due to a formation of crystal nucleus.	8
Figure 2-1. Example of the normalized correlation function of the scattered light $g_2(t) - 1$ and the intensity distribution function $G(t)$ of a lysozyme solution	19
Figure 2-2. The refractive index increment dn/dc of lysozyme solutions	21
Figure 2-3. Cluster characterization by Brownian microscopy. a, Schematic of the BM setup. b, A typical BM image shown as a negative. c, An example of cluster trajectory. d, Distribution of cluster sizes.....	23
Figure 3-1. Formation of protein oligomers. a, Protein lysozyme structure (RCSB PDB: 2VB1). b, Schematic representation of the domain swapping mechanism.....	27
Figure 3-2. Schematic of NMR experiment. Two samples were prepared with initiated 1H - 2D initiated. After 1 or 72 hours from preparation 1H - 2D was stopped. Both solutions were tested for remaining 1H - ^{15}N signal by NMR.	29
Figure 3-3. 1H - ^{15}N HSQC spectra of a dilute (in red) and concentrated (black) lysozyme solution at 1 (left) and 72 hours (right) after solution preparation.	30
Figure 4-1. Intensity correlation functions from: a, a lysozyme solution with shown concentration at shown time after solution preparation, b, an identical solution containing latex spheres with $R_{spheres} = 200 \mu m$	33
Figure 5-1. The viscosity of 100 mg ml^{-1} lysozyme solutions hosting the protein-rich clusters.	40
Figure 5-2. The distribution of electrostatic potential of the solvent-accessible surface of a lysozyme molecule at $pH = 7.8$	44

Figure 5-3. Characterization of the intermolecular interaction in lysozyme solutions at increasing ionic strength. Debye plots $Kc/R\theta$ (c) of lysozyme solutions. a, in HEPES buffer; b, in 20 mM HEPES with added electrolytes.	45
Figure 5-4. Monomers and clusters in lysozyme solutions at varying ionic strength and salt identity, characterized by DLS. a, c, e Normalized autocorrelation functions $g_2 - 1$. b, d, f, The amplitudes of monomers and clusters..	47
Figure 5-5. Characterization of the intermolecular interactions in solution at pH = 7.8.	48
Figure 5-6. 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.....	49
Figure 5-7. The charged groups on the surface of a lysozyme molecule. PBD structure file 2VB1 was used.	50
Figure 5-8. Populations of protein-rich clusters in 100 mg ml ⁻¹ lysozyme solutions in HEPES buffer at pH = 7.8 characterized by dynamic light scattering (DLS) and Brownian microscopy (BM).	53
Figure 5-9. pH effects on the cluster characteristics, the cluster radius R_2 in a, and the cluster volume fraction ϕ_2 in b.....	54
Figure 5-10. The role of hydrophobic interactions and partial unfolding in cluster formation.....	58
Figure 5-11. 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.	61

Figure 5-12. 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. 5-10 c.	62
Figure 5-13. Characterization of intermolecular interaction in lysozyme solutions during bubbling test.....	64
Figure 5-14. Evolution of cluster characteristics during bubbling of N_2 and, O_2 , at 0.3 L/min flow rate and preservation of them within 24 hours. a, cluster radius R_2 , b, volume fraction ϕ_2 , c, concentration n_2	65
Figure 5-15. The effects of shear flow on cluster formation. The response of the cluster radius R_2 to solution bubbling with N_2	66
Figure 6-1. The structure of tetrameric glucose isomerase molecule drawn with Pymol. Different units are colored correspondingly.	69
Figure 6-2. The pairwise interactions in glucose isomerase solutions. a, Debye plots, b, the second virial coefficient B_2 , obtained from a.....	72
Figure 6-3. Viscosity of glucose isomerase solutions without and in the presence of urea.	73
Figure 6-4. Clusters characteristics in glucose isomerase solutions in the presence of urea, a, clusters radius R_2 , b, clusters volume fraction ϕ_2 , c, clusters concentration n_2	75
Figure 6-5. Distributions of glucose isomerase clusters population before and after bubbling with N_2 at 0.3 L min^{-1} in terms of concentration a, and volume fraction b.....	76

Figure 6-6. Evolution of glucose isomerase clusters characteristics after bubbling: a, cluster radius R_2 , b, cluster volume fraction ϕ_2 , c, cluster concentration n_2	77
Figure 7-1. Lysozyme crystals of tetragonal group of symmetry grown in 20 mM HEPES pH 7.8 at 250 mg ml ⁻¹ concentration.	83
Figure 7-2. Ball and stick representation of organic additives used in this work and their chemical formulae. First row: ethanol, acetone; second row: glycerol, MPD (from the left to the right).	84
Figure 7-3. Dependence of lysozyme monomer diffusivity on the concentration of organic additives used as cosolvents. The lines shown here are guides for an eye.	86
Figure 7-4. Static light scattering characterization of intermolecular interactions in protein solutions.....	88
Figure 7-5. The increase of chemical potential μ_2 of lysozyme with its concentration in the presence of organic additives at different volume fractions: a – ethanol, b – acetone, c – glycerol, and d – MPD.....	91
Figure 7-6. Lysozyme response to the presence of organic cosolvents.....	93
Figure 8-1. Imaging of liquid objects with oblique illumination microscopy. a. Images of protein-rich liquid clusters in a lysozyme solution and a glucose isomerase solution. b. The trajectories of the clusters imaged in a.....	102
Figure 8-2. Tracking of latex spheres and gold nanorods by currently existing methods. a. Gray-scale images of the particles. b. Two-dimensional trajectories of the particles traced by the five methods.....	103

Figure 8-3. Illustration of the spatial cross-correlation (SCC) method of particle tracking.	105
Figure 8-4. Evaluation of the accuracy of the SCC method. a, particle trajectory, b, MSD computed from a, c, diffusivity calculated for each lag time from b, d, distribution of particles diffusivities calculated as in c.....	110
Figure 8-5. Comparisons of particle tracking and MSD evaluation methods.....	113
Figure 8-6. The effects of movie length and frame rate.	115
Figure 9-1. Intensity patterns of microspheres (a) and lysozyme clusters (b) observed by oblique illumination dark field microscopy. The intensity spots signify unique properties of protein clusters shape oscillating in time.....	121
Figure 9-2. Time decay of the spatial cross-correlation (SCC) functions maximum of intensity profile fluctuations of a particle (a) and a cluster (b).....	122
Figure 9-3. Clusters a liquid with a highly viscous core inside. Fluorescent microspheres are observed through a bandpass filter. A dashed circle indicates three particles entrapped inside a cluster.....	125
Figure 10-1. Response of porcine insulin solutions to the presence of acetone. Tested protein concentrations: 2 mg/ml (a, c) and 4 mg/ml (b, d). The results are shown in terms of clusters concentration (a, b) and volume fraction (c, d).....	130
Figure 11-1. Snapshots from DOIDM showing three types of scatterers in a lysozyme solution under crystallization conditions: clusters, crystals, and clusters containing crystals. Orientation of polarizers is indicated in the figure.	135

LIST OF TABLES

Table 5-1. Values of the parameters of the intermolecular interactions.....	43
Table 7-1. The virial coefficients calculated using Kirkwood-Goldberg theory of scattering from multicomponent systems.	90

CHAPTER 1

INTRODUCTION

Protein crystallization is one of dominant areas in protein science. The necessity of proteins to be crystallized is brought by a desire to solve protein structures feasible nowadays by a significantly developed method of X-ray diffraction. This technique requires protein crystals to be grown with minimum defects to a sufficient size of 50 – 500 μm for conventional X-ray data collection¹. The challenges of protein crystallization are widely discussed in the literature and are particularly noticed for membrane proteins²⁻⁴. Hence, protein crystallization is a limiting step in characterization of protein single molecules crucial for elucidating of the mechanism of action and their role in living processes.

Giegé and Ducruix in introduction to “Crystallization of Nucleic Acids and Proteins.”⁵ notice that the interest for protein crystallization started even prior arrival of X-ray diffraction method. The early protein crystallization dates to 1840 when Hünefeld crystallized earthworm hemoglobin⁶. The crystals were obtained by empirical mistake when the blood was stored over a prolonged period of time between two microscope sides, so it slowly dried out. After that, hemoglobin variants from various invertebrates and vertebrates were crystallized as well. In 1851 Funke described the growth of hemoglobin crystals in solutions containing alcohol and ether^{7,8}. These discoveries initiated Felix Hoppe-Seyler studies on hemoglobin function which shed the light on oxygen binding to erythrocytes, a major hemoglobin function⁹.

Protein crystallization those days was mostly fascinating for bioscientists rather than had a research interest. Yet, it brought significant discoveries in understanding of protein nature and helped to distinguish proteins from enzymes, small macromolecular catalysts. The profound historical background on protein crystallization and understanding their particular nature is described the work of the French professor Claude Debru, “L'esprit des Protéines : Histoire et Philosophies Biochimiques”¹⁰.

Since the first experiments with hemoglobin, crystallization of various proteins became more and more extensive. In 1859 the first plant protein excelsin, the reserve protein of the Brazil nut, was crystallized by Maschke¹¹. In 1890 – 1898 hen egg and horse serum albumins were crystallized by Hofmeister¹² and Hopkins and Pinkus¹³ (2). In 1925 Sumner crystalized urease, mentioning the increase of protein activity of urease solutions from its crystals – the first indication of protein purification by crystallization¹⁴. This achievement is followed by crystallization of pepsin by Northrop in 1929¹⁵ and tobacco mosaic virus by Stanley in 1935¹⁶. These achievements were recognized by the 1946 Nobel Prize in Chemistry awarded to these three scientists for the importance of the discoveries in biology.

The discovery of x-ray radiation in 1895 opened a new era in protein crystallization when protein crystals were grown to resolve their molecular structure. The first protein pattern produced with X-ray was of pepsin in 1934 by J.D. Bernal and D. Crowfoot-Hodgkin¹⁷. Since this discovery many macromolecules were crystallized for the purpose of study of their structures. The crystallization methods, as noticed by Giegé and Ducruix⁵, were still empirical since scientists were mostly focused on the development of X-ray diffraction as a new method to study crystal structures.

Nowadays, there are two areas in protein science focused on crystallization. First, the structure of biological macromolecules with high atomic resolution is required for understanding proteins functions and their interactions with other molecules and molecular compounds of living organisms. And, second, the delivery of biopharmaceuticals foresees its potential in development and engineering proteins and macromolecules capable to crystallize. This approach promises to show benefits in delivering biopharmaceutical at high concentration, higher molecular stability, low viscosity of their formulation along with a benefit of slow and controlled release¹⁸. Protein insulin, however, still remains the only protein to be approved in therapeutic use in crystalline state and much more of progress is needed in this area to put forth therapeutic applications of proteins. The formation of protein crystals and protein aggregates are of significant interest for researches investigating the molecular mechanisms of some diseases originating from protein malfunction: anemia¹⁹⁻²², cataract²³⁻²⁵, prion diseases²⁶, and others.

Opposite to crystallization of inorganic compounds, protein crystallization remains a pure empirical process with no exact general approach of obtaining crystals from a randomly taken protein. The reason for this fact lies in the structure complexity of these molecules. Periodical pattern, required for crystal formation, might not be achieved for protein molecules due to a highly complicate combination of electrostatic, hydrophobic and steric interactions of its parts. As a result, conditions of protein crystallization are not predictable, the crystals are extremely sensitive to external conditions, such as temperature, pH, ionic strength. Even when crystallization is achieved the protein crystal growth and crystal structures analysis remains challenging

procedures due to several factors. Protein crystals have high solvent content (from 30 to 80%) which makes them very soft and fragile and the size of protein crystals, when they are obtained, remain quite small due to accumulation of surface stress arresting the further growth. Overall, protein crystallization remains a challenging area of protein research and understanding how crystals grow is a required step for further progress.

As mentioned above, structural biology strictly relies on the ability to grow crystals of sufficient size capable to produce qualitative diffraction patterns. The necessity to use intensive X-rays leads to an extensive damage of protein crystals before the data of acceptable quality is obtained²⁷⁻²⁹. This is especially true for small protein crystals. Chapman et al.³⁰ present a method in which they use X-ray pulses briefer than the timescale of most conventional imaging methods which allows them to determine the structure of crystals of non-sufficient sizes for other imaging techniques. This approach allows molecular structure determination for crystals as small as 200 nm, meaning that the growth process for these crystals is practically eliminated, leaving nucleation as primary step on the way of preparing a study material. Hence, profound understanding of nucleation as a process leading to crystal formation is a crucial requirement for structural biologists on their way to solve structures of living macromolecules.

Unlike inorganic compounds which could be crystallized from vapor, solution or melt, protein crystallization uses different designs of crystallization from a solution. The driving force for a transformation of protein molecules from liquid to solid state is the excess of the free energy of the initial system phase (solvent) in comparison to the final state (crystal and solvent). In other words, ΔG_{cryst} has to be negative for crystal formation to be possible:

$$\Delta G_{cryst} = \Delta H_{cryst} - T\Delta S_{cryst} < 0, \quad (1-1)$$

where ΔH_{cryst} is the enthalpy change of crystallization, ΔS_{cryst} is the entropy change of crystallization, and T is absolute temperature. The classical nucleation theory was described by J.W. Gibbs in 19th century^{31,32}.

Nucleation is the first step in crystallization. It is the example of the first order phase transition, when the derivative of the free energy of the system with respect to a thermodynamic variable experiences a discontinuity across the phase boundary, which means, that the coexisting phases would have different entropies, $S_{crystal}$ and $S_{solvent}$. In its turn that leads to the fact that the system must absorb or release heat during the phase transition, i.e., the system has a non-zero enthalpy of transformation

$$\Delta H_{cryst} = T(S_{crystal} - S_{solvent}) = H_{crystal} - H_{solvent} . \quad (1-2)$$

The origin of this matter is the rearrangement of solvent molecules, participating in crystallization, mostly water molecules.

The structuring of water molecules around building material of crystals in solutions is an important factor in the thermodynamics of crystallization³³. Both scenarios were proven to be true, structured water molecules can be released when the molecules or ions join the growing crystal³⁴ or additional water molecules can be also trapped during crystal formation³⁵. Both ways of reorganization of water molecules have significant enthalpic and entropic consequences. The enthalpy effect significantly depends on the system of study and this term may vary from crystal to crystal. Enthalpic contributions in crystallization are mostly weakly negative and derive from the side chain interactions between protein molecules within the crystals replacing the protein-solvent interactions. The change in entropy of crystallization is due to two terms

$$\Delta S_{cryst} = \Delta S_{protein} + \Delta S_{solvent}, \quad (1-3)$$

where $\Delta S_{protein}$ is loss in rotational and translational freedom of the protein in the crystal lattice if compared to solution (negative term), $\Delta S_{solvent}$ is a gain in the solvent entropy due to the release of water molecules from both, hydrophobic and polar, surface residues (positive term, usually, $5 \text{ cal mol}^{-1} \text{ K}^{-1}$ per one water molecule released).

As mentioned above, protein crystals might have large quantities of water entrapped. Hence, in case of proteins the entropy contribution due to solvent trapping or release can be significant and determinant in the direction of the phase transition. In the example of apoferritin crystallization it was found that enthalpy of crystallization and the related energy of pair interactions in the solution are close to zero and crystallization is mostly driven by the increase of solvent entropy due to release of up to two water molecules in each of the 12 intermolecular crystal contacts³⁴. This entropy gain is roughly estimated as $+160 \text{ J K}^{-1}$ per mole of protein with lowering the free energy by about -47 kJ mol^{-1} . In a different example of crystallization of the human mutant hemoglobin C, the positive contribution of the enthalpy of crystallization has to be counterweighed by a negative enthalpy of the phase transformation. It is achieved by the release of about 10 water molecules per crystal contact and net gain of free energy upon crystallization about -20 kJ mol^{-1} ³⁵.

Nucleation is the first step in the formation of a new thermodynamic phase and whether the newly created phase continues to grow or dissolves depends on the energetic considerations of the system and bears stochastic character. Generally, nucleation is distinguished either as homogeneous or heterogeneous. For simplicity, we may start with

a spherical nucleus of radius r nucleating in a bulk solution. This scenario is related to homogeneous nucleation. The change of the net free energy of the system is

$$\Delta G = \Delta G_{bulk} + \Delta G_{solution} = -\frac{4}{3}\frac{\pi r^3}{\Omega} \Delta\mu + 4\pi r^2 \gamma, \quad (1-4)$$

where Ω is the volume per molecule, $\Delta\mu$ the free energy response to molecules transferring from liquid to solid phase, γ is the interfacial free energy. The two terms in the equation above signify two distinct contributions to the overall transformation³⁶. The first term is the free energy difference between the solid and liquid phases, i.e., the volume free energy. Its value depends on the experimental conditions and for the temperatures below the equilibrium solidification temperature the value $\Delta\mu$ is positive, making the overall negative contribution to the bulk free energy. The second energy term results from the formation of the new interface between crystal and solution. Associated with this term the free surface energy γ is positive since the molecules of the nucleus acquire uncompensated bonds of higher energetic cost. This term remains positive at any possible experimental conditions.

The contributions of the volume and surface free energies and the resulting net Gibbs free energy ΔG are illustrated in Fig. 1-1. Each component of the net free energy is a monotonic function of the nucleus radius while their combination first increases, passes through a maximum and then monotonically decreases. Physically, this behavior means that the formation of a new solid phase, crystal nucleus, is associated with the increase of the free energy. When the net gained free energy reaches its maximum ΔG^* the further increase of nucleus size lowers the free energy of the system. This characteristic size r^* is called critical nucleus size and all nuclei that due to random fluctuations in its dimensions reached this size will continue their growth to

macrocrystals. The nuclei smaller than the critical size have two pathways. Random fluctuations may help them to exceed the critical parameter and then their existence will follow the previously described scenario, or, if they stay below the critical size, they will dissolve as thermodynamically expensive aggregates due to the excess of their surface free energy contribution $4\pi r^2$.

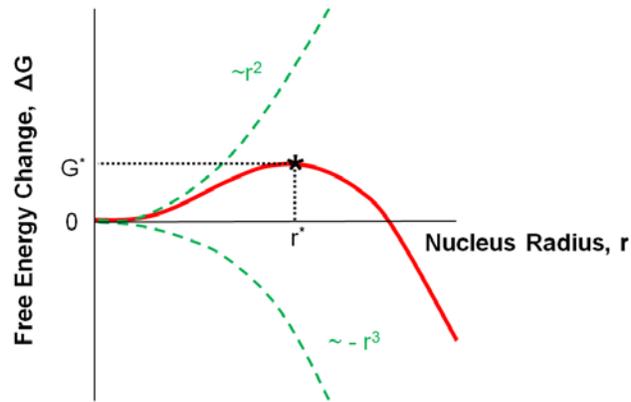


Figure 1-1. The change of the system free energy due to a formation of crystal nucleus.

To solve for the critical radius r^* we may take the derivative at the maximum of ΔG and we get

$$r^* = -\frac{2\Omega\gamma}{\Delta\mu} = \frac{2\Omega\gamma}{k_B T \sigma} \quad (1-5)$$

where k_B is the Boltzmann constant and σ is the supersaturation of the system. Hence, the critical net Gibbs free energy is

$$\Delta G^* = \frac{16\pi/3 \gamma^3}{(\Delta\mu)^2} = \frac{16\pi/3 \gamma^3}{3(k_B T \sigma)^2} \quad (1-6)$$

For the nucleation at the foreign surface, heterogeneous nucleation, the critical nucleus radius takes the form

$$r^*_{hetero} = \frac{2\Omega\gamma'}{k_B T \sigma'} \quad (1-7)$$

where modification to the surface free energy γ' is of the form

$$\gamma' = \gamma_{sol-cryst} \left(1 - \frac{\gamma_{sol-surf} - \gamma_{surf-cryst}}{2\gamma_{sol-cryst}} \right), \quad (1-8)$$

where indices *sol*, *surf*, *cryst* stand for solution, surface and crystal components of the interface. The free energy of the crystal-surface interface is less than that of the surface-solution, making the term in the brackets less than one. As a result, the value of r^*_{hetero} at the foreign surface is smaller than for nucleation in bulk, homogeneous nucleation.

The critical nucleus size controls the probability of nucleus formation at any given timescale³³. Equation (1-6) determines the kinetics of nucleation. Similarly to any other kinetically-limited chemical processes, the nucleation probability is proportional to the exponential of the barrier height normalized by $k_B T$. This nucleation rate was described by Nielsen and Abraham^{37,38}

$$J_n = A \exp\left(-\frac{\Delta G^*}{k_B T}\right), \quad (1-9)$$

where preexponential factor A depends on many parameters. For the rate of nucleation from solutions it is proportional to the density of the dissolved substance, n , and to the particle flux towards the surface of the crystal nucleus with area $4\pi r^{*2}$ (Chernov)³⁹. In solutions this rate depends on the rates of diffusion and addition of particles to the nucleus. The addition of particles requires breaking of several of their bonds with the solvent, i.e., overcoming the potential barrier. The preexponential factor can be written as

$$A \cong 4\pi r^{*2} n^2 \nu a \exp\left(-\frac{G_{act}}{k_B T}\right), \quad (1-10)$$

where ν is the frequency of attachment of particles to the crystal surfaces, a is the characteristic size of a particle, G_{act} is the activation energy for the addition of particles.

Obtaining the data on nucleation rates is always challenging and laborious measurements. In protein crystallization, the first measurements of nucleation rates were achieved in 2000 by Galkin and Vekilov⁴⁰. Importantly, their measurements could separate heterogeneous nucleation from homogeneous, allowing the extraction of the steady-state rate of homogeneous nucleation from solution. Shortly, the method required a protein solution to be loaded at a temperature chosen to prevent nucleation or liquid-liquid demixing. Then the temperature was lowered to a selected T_1 at which nucleation occurred. After a time period of Δt the temperature was raised from the nucleation temperature T_1 to the growth temperature T_2 . At T_2 , supersaturation was at levels where nucleation rate was almost eliminated but already formed crystals could grow to detectable sizes. This allows separation of the nucleation from the ensuing growth. After the growth stage, the nucleated crystals were counted. To suppress undesired heterogeneous nucleation at the solution-air interface, the protein solution droplets were suspended in inert silicon oil.

The dependences of nucleation rates obtained in⁴⁰ contained certain peculiarities. One of those was the fact, that all measured nucleation rates were about ten orders of magnitude less than the prediction of the classical nucleation theory. The estimation of the homogeneous nucleation rate suggested that the difference between the experimentally determined J and the prediction of the classical nucleation theory is due to a lower preexponential factor.

To solve the puzzle above, the two-step mechanism of nucleation in protein solutions was suggested, proposing that crystal nucleation occurs inside metastable mesoscopic clusters of dense protein liquid. According to this mechanism, crystal nuclei assemble within preexisting protein-rich clusters rather than from molecules in the dilute solution. Since the surface free energy at the interface between the crystal and the solution is significantly higher than at the interface between the crystal and the dense liquid, the barrier for nucleation of crystals from the solution would be much higher. This would lead to much slower nucleation of crystals directly from the solution than inside the clusters. Thus, the protein crystal nucleation follows the two-step mechanism of nucleation because it provides for faster rate of the solution to crystal phase transition and in this way for faster decrease of the free energy of the system, which corresponds to faster increase of the entropy. Numerical estimates of the typical surface free energy of the nucleus suggest that it is on the order of 10^{-4} J m^{-2} ,⁴⁰⁻⁴⁵ (i.e., up to two orders of magnitude lower than γ of a nucleus forming in the solution). Since the nucleation barrier depends on the cube of γ ,⁴⁰ this mechanism leads to orders of magnitude lower barriers and makes nucleation feasible over laboratory timescales.

The evidence for existence of dense liquid clusters comes from monitoring solutions of three hemoglobin variants, oxy-HbA, oxy HbS, and deoxy-HbS, proteins lumazine synthase and lysozyme by dynamic light scattering⁴⁶⁻⁴⁹, discussed in detail in chapter 2, Principal Experimental Techniques. A typical intensity correlation function of a lysozyme solution in the homogeneous region of the phase diagram would reveal two processes. The fast process corresponds to the Brownian motion of single lysozyme molecules which is present at all solution concentrations. The related hydrodynamic

radius could be determined by Einstein-Stokes equation and yields the value of about 1.5 nm and matches well the diameter of a lysozyme molecule of 3.2 nm^{46,50}. The slower process takes longer times in comparison to monomers diffusion. Its amplitude increases with higher lysozyme concentrations⁵⁰. This characteristic time could come from either compact lysozyme clusters suspended in the lysozyme solutions, or from single lysozyme molecules embedded in a loose network structure constraining their free diffusion. Since the measured previously low-shear viscosity of lysozyme solutions is equal to those determined using high shear rates⁵¹, no loose networks of lysozyme molecules exist in these solutions and the conclusion is that long times in correlation functions indeed correspond to lysozyme clusters. Since the clusters appear immediately after the solution preparation, it was reasonable to assume that these clusters are liquid. The detailed study of clusters state is discussed in chapter 9, The Liquid Nature of Protein-Rich Clusters.

The cluster diameters vary from ca. 100 nm for the relatively small lysozyme⁵⁰ to several hundred nanometers for larger proteins^{52,53}.^{52,53}. The results on the behavior of clusters of dense liquid in solutions of hemoglobin and lumazine synthase are discussed in^{48,52,54} 68, 70, 72. The cluster radius stays relatively steady over a prolonged periods of time. It was found that with all studied proteins so far, the clusters exist in broad temperature and protein concentration ranges.

The number density n_2 of the dense liquid clusters and the fraction of the total solution volume φ_2 they occupy are evaluated from the amplitudes A_1 and A_2 of the intensity distribution function⁵⁴ as

$$n_2 = \frac{A_2}{A_1} \frac{1}{P(qR_2)f(C_1)} \frac{\left(\frac{\partial n}{\partial C_2}\right)_{T,\mu}}{\left(\frac{\partial n}{\partial C_1}\right)_{T,\mu}} \left(\frac{\rho_1}{\rho_2}\right)^2 \left(\frac{R_1}{R_2}\right)^6 n_1 \quad \text{and} \quad (1-11)$$

$$\varphi_2 = \frac{A_2}{A_1} \frac{1}{P(qR_2)f(C_1)} \frac{\left(\frac{\partial n}{\partial C_2}\right)_{T,\mu}}{\left(\frac{\partial n}{\partial C_1}\right)_{T,\mu}} \left(\frac{\rho_1}{\rho_2}\right)^2 \left(\frac{R_1}{R_2}\right)^3 \varphi_1, \quad (1-12)$$

where $P(qR_2)$ is the shape factor for the clusters of radius R_2 , q is the scattering vector, $C_{1,2}$ is the concentration of monomers and clusters correspondingly, $f(C_1)$ is a virial expansion containing the coefficients of the solution osmotic compressibility, $\left(\frac{\partial n}{\partial C_{1,2}}\right)_{T,\mu}$ are the refractive index increments at constant temperature and chemical potential of monomers and clusters correspondingly, $\rho_{1,2}$ are the densities of monomers and clusters. Experimental data showed that the clusters occupy volume fraction up to 10^{-3} ⁵⁵.

To evaluate the lower boundary of the lifetime of the lysozyme clusters, it was noted that cluster decay processes contribute a q -independent component to the overall rate sensed by DLS⁵⁶,

$$\Gamma_2 = \Gamma_0 + D_2 q^2 \quad (1-13)$$

and can be distinguished from cluster diffusion. Here, Γ_0 is the rate of cluster decay, D_2 is the cluster diffusion coefficient. The q -dependent, diffusion component indeed dominates the DLS signal. Using

$$\Gamma_0 \ll D_2 q^2, \quad (1-14)$$

a lower bound for cluster lifetime is estimated as $\frac{1}{\Gamma_0} \approx 15 \text{ ms}$ ⁵⁰.

Analysis of the lumazine synthase solutions revealed that the clusters exist at all times and only when their volume fraction exceeds 10^{-8} they can be detected with DLS. In contrast to the macroscopic dense liquid the clusters are metastable with respect to the low-concentration solution, the characteristic time is limited to $\sim 10 \text{ s}$, after which they decay. In addition to detection by dynamic light scattering, clusters of lumazine synthase

were directly imaged by atomic force microscopy which confirmed their macroscopic lifetimes^{48,52}. The new studies of protein clusters conducted with Brownian microscopy method, (discussed in details in chapter 2, Principal Experimental Techniques), suggests that clusters of lysozyme and glucose isomerase reveal lifetimes greater than their characteristic diffusion time over the distance of 5 μm . In all discussed scenarios, the lifetimes of the clusters significantly exceeds the equilibration times of the protein concentration at sub-micrometer length scales, i.e., 10^{-5} s ³³. Thus, the compact clusters represent a metastable phase separated from the bulk, dilute solution by a free energy barrier.

Attempts to rationalize the finite size of clusters have focused on a balance of short-range attraction, due to van der Waals, hydrophobic or other forces, and screened Coulombic repulsion between like-charged species^{57,58}. While small clusters, tens of particles or so, naturally appear in such approaches, large clusters are expected only if the constituent particles are highly charged, with hundreds or so elementary charges. Such high charges are feasible for micron-size colloidal particles, however proteins in solutions are known to carry somewhat around 10 elementary charges per molecule as an order of estimate. While these theories have been successfully applied to aggregation in colloidal suspensions⁵⁹⁻⁶¹, a distinct mechanism is at work in protein systems, where clusters contain as many as 10^6 molecules⁴⁶. The alternative approach of forces governing cluster formation is discussed in details in chapter 5, Lack of Dependence of the Sizes of the Mesoscopic Protein Clusters on Electrostatics, where the combination of hydrophobic and water-structuring interactions is suggested as a principal net effect leading to cluster formation.

Recently, a mechanism was proposed where the clusters consist of a concentrated mixture of protein complexes and monomers⁴⁶. In this mechanism, cluster formation results from an interplay between monomer influx, complex formation, and subsequent complex outflow and decay. The complexes are characterized by non-typical tertiary or quaternary structures, depending on the protein. In the case of monomeric proteins, such as lysozyme, these complexes likely represent an ensemble of dimers made of partially unfolded monomers, including domain-swapped dimers^{62,63}. In oligomeric proteins, such as hemoglobin and glucose isomerase (both are typically tetrameric), the complexes may be oligomers whose order is likely higher than its typical value, while the degree of unfolding is insignificant. The domain swapping mechanism and the residues participating in this process are discussed in chapter 3, Molecular Level View of Oligomerization in Protein Solutions, along with nuclear magnetic resonance characterization of lysozyme clusters. The mechanism of cluster formation in glucose isomerase solutions, including the hypothesis of the new species formation, is discussed in chapter 6, Alternative Mechanisms of Cluster Formation on the Example of Protein Glucose Isomerase. Shape change, either by domain swapping or misassembled oligomers, emerges as a crucial precondition for the formation of the clusters. Thus, contrary to intuitive views of crystallization, limited protein conformational flexibility becomes a necessary facilitator for crystal nucleation⁶⁴⁻⁶⁷.

In my dissertation thesis I present our recent advances in understanding of the protein clusters nature. I discuss the mechanism of the formation of new species contributing to cluster existence, the forces governing their formation. I present the newly obtained results on the liquid nature of protein clusters and with a new method of

Brownian microscopy discuss the challenges of diffusion characterization of micron-size clusters. Finally, I attempt to explain why the two-step mechanism of nucleation is selected on the example of protein insulin. In the concluding chapter I show the evidence of the two-step-mechanism of nucleation visualized for the first time with oblique illumination dark field microscopy on the example of nucleation of proteins glucose isomerase and lysozyme.

CHAPTER 2

PRINCIPAL EXPERIMENTAL TECHNIQUES

The requirements for the techniques employed in the investigation of protein-rich liquid clusters are determined by the clusters properties. For the several proteins studied so far, the cluster radii have been in the range 50–500 nm, the fraction of the solution volume occupied by the cluster population has been in the range up to 10^{-3} with the lower boundary limited by restrictions of methods. The cluster volume fraction and size indicate that the average separation between clusters in solution is of the order of micrometers. The cluster size and separation precludes the use of small-angle neutron and X-ray scattering: these two methods detect structures with characteristic length scales in the angstrom and nanometer range^{61,68,69}. The fraction of protein held in the clusters is about $10\times$ the cluster volume fraction meaning that the nuclear magnetic resonance (NMR) signal from the clusters may be too weak to be detected.

Brownian microscopy (BM) and dynamic light scattering (DLS) rely on visible light scattered by the monomers and clusters. According to the Rayleigh law, the intensity scattered from an object is proportional to the sixth power of its radius. Hence, the clusters, which are about two orders of magnitude larger than the monomers, provide a scattering intensity stronger by 12 orders of magnitude. These considerations make DLS and BM particularly well suited for the detection and characterization of the mesoscopic protein-rich clusters.

2-1. Dynamic light scattering

The method of dynamic light scattering relies on light scattered by concentration fluctuations. Since the rate of intensity decay is determined by the diffusion of the scatterers, this rate yields the diffusion coefficient of the scatterers and, using the Einstein–Stokes relation and the viscosity of the medium, their size⁷⁰. The concentration of individual species is determined from the intensity that each of them scatters.

The rate of intensity variation $I(t)$ is determined from the intensity correlation function $g_2(\tau)$ of the scattered light. This function is defined from the intensity at two times, t and $(t - \tau)$, as⁷¹

$$g_2(t) = \frac{\langle I(t)I(t-\tau) \rangle_{\Delta t}}{\langle I \rangle_{\Delta t}^2}, \quad (2-1)$$

where $\langle \rangle_{\Delta t}$ signifies averaging over time Δt and $\langle I \rangle_{\Delta t}$ is the average intensity. The normalized correlation function $g_2(t) - 1$, illustrated in Fig. 2-1, can be represented as the square of the sum of exponential members characterizing the scatterers with different diffusion rates Γ_i . The characteristic diffusion times τ_1 and τ_2 and the amplitudes A_1 and A_2 of the monomers and clusters, respectively, are shown on the graph.

Our dynamic light-scattering experiments are aimed at identifying mainly two scatterers: single molecules and larger clusters. Hence⁷²,

$$g_2(\tau) - 1 = \left[A_1 \exp\left(-\frac{t}{\tau_1}\right) + A_2 \exp\left(-\frac{t}{\tau_2}\right) \right]^2, \quad (2-2)$$

where $\tau_1 = 1/\Gamma_1$ and $\tau_2 = 1/\Gamma_2$ are the characteristic times of the diffusion of scatterers, whose contribution to the scattered light has amplitudes A_1 and A_2 .

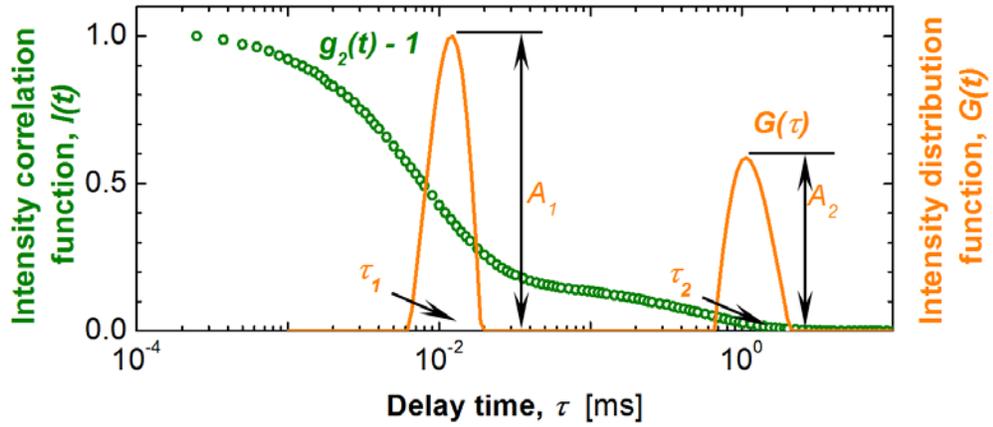


Figure 2-1. Example of the normalized correlation function of the scattered light $g_2(t) - 1$ and the intensity distribution function $G(t)$ of a lysozyme solution.

The characteristic times τ_1 and τ_2 and the amplitudes A_1 and A_2 , also shown in Fig. 2-1, are readily determined from the distribution function $G(\tau)$ defined by the expression

$$g_2(\tau) - 1 = \left[\int G(\tau) \exp\left(-\frac{t}{\tau_1}\right) d\tau \right]^2. \quad (2-3)$$

Hence, $G(\tau)$ is calculated by numerically inverting the Laplace transform with $(g_2 - 1)^{1/2}$, using a software package based on the CONTIN algorithm^{73,74}.

To calculate the equivalent hydrodynamic radii from the values of the relaxation times τ_1 and τ_2 , the Stokes–Einstein relation is used, modified with $\Gamma_i = \tau_i^{-1} = D_i q^2$,

$$R_i = \frac{k_B T q^2}{6\pi\eta_i} \tau_i, \quad (2-4)$$

where $i = 1, 2$ for single molecules or clusters, respectively, k_B is the Boltzmann constant, T is the absolute temperature, η_i is the viscosity to which a diffusing object i is exposed, D_i is its diffusion coefficient and the scattering vector $q = 4\pi n/\lambda \sin(\theta/2)$, where n is the refractive index of the solvent, λ is the wavelength of the laser beam and θ is the scattering angle. The amplitudes A_1 and A_2 are the basis for the determination of the concentration n_2 and volume fraction ϕ_2 of the clusters⁵⁴.

CONTIN algorithm tends to produce questionable results in protein solutions analysis⁷⁵. To avoid uncertainties, we also fit experimental data by cumulative method, in which the intensity distribution function was modeled as a sum of two Dirac delta functions corresponding to the monomers and clusters respectively,

$$G(\tau) = G^{monomer}(\tau_1) + G^{cluster}(\tau_2) = A_1\delta(\ln t - \ln \tau_1) + A_2\delta(\ln t - \ln \tau_2), \quad (2-5)$$

where A_1 and A_2 are the amplitudes of the peaks corresponding to the monomers and clusters, respectively, and τ_1 and τ_2 are the respective diffusion times.

We collected DLS data by ALV light scattering device equipped with He-Ne laser ($\lambda = 632.8$ nm, 35 mW) and ALV-5000/EPP Multiple tau Digital Correlator (ALV-GmbH, Langen, Germany). During the experiment, protein solutions were held in cylindrical cuvettes of volume from 0.5 to 1 ml at 22^oC. The autocorrelation functions were acquired at 90^o for 60 s during 30 mins unless indicated otherwise. If the long-term evolution of the scattering objects in a solution was of interest, numerous data sets can be collected in sequence for up to several days or even longer.

2-2. Static light scattering

Pairwise interactions in protein solutions can be characterized using static light scattering (SLS) technique. We performed SLS measurements on the same ALV device used for DLS. The scattered intensity we collected at 90^o and the results we showed as Debye plots. For molecules in the dilute solution regime the simplified scattering equation is

$$\frac{Kc}{R_\theta} = \frac{1}{M_w} + 2B_2c, \quad (2-6)$$

where $R_\theta = I_\theta/I_0$ 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, $dn/dc = 0.199 \pm 0.003 \text{ ml g}^{-1}$ is the refractive index increment of the solutions. This parameter was determined for each solution composition, using a Brookhaven differential refractometer operating at $\lambda = 620 \text{ nm}$ and calibrated with KCl solutions in water 25°C , Fig. 2-2. The value dn/dc was measured in two solvents, in water, where the ionic strength $I \cong 0 \text{ mM}$, and in 60 mM HEPES at $\text{pH} = 7.8$, where $I = 40 \text{ mM}$. Measurements did not reveal significant difference between the two solvents: both data sets yield $dn/dc = 0.199 \text{ ml/g}$.

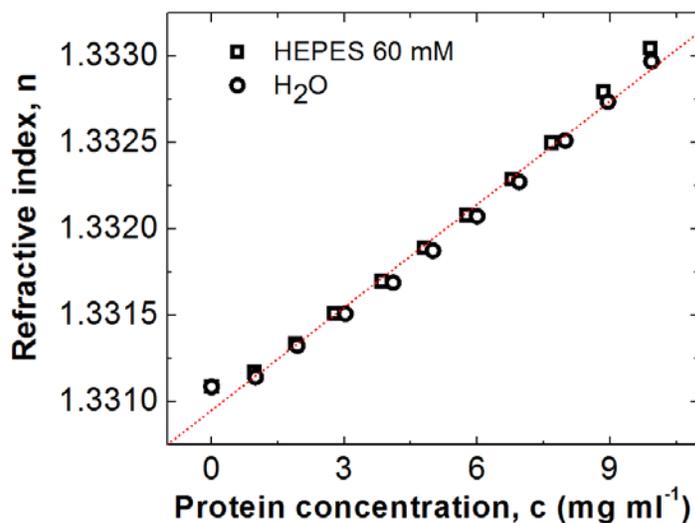


Figure 2-2. The refractive index increment dn/dc of lysozyme solutions.

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.

2-3. Brownian microscopy

Brownian microscopy (BM) is a relatively new method but it is already intensively used to study clusters behavior^{65,66}. We used Nanosight LM10-HS microscope (Nanosight Ltd, currently Malvern Instruments, USA) 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. The solution was illuminated by a laser beam configured so that it does not enter the objective lens of an observation microscope, Fig. 2-3 a. The observation volume is determined by the focal depth of the objective lens and the view field of the microscope, and is typically $120 \times 80 \times 5 \mu\text{m}^3$ (width \times length \times height).

All species in solution scatter the incident light. The intensity scattered by a protein 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_l is the monomer radius), so the clusters are well seen on the background of monomers even at a relatively high protein concentration, Fig. 2-3 b. A 20 \times 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 camera settings and in our experiments it was about 25 fps unless it is noted otherwise. Each frame of the movie is an image of clusters as bright white spots on a dark background. Careful observation reveals that the cluster spots consist of concentric fringes that result from optical geometry.

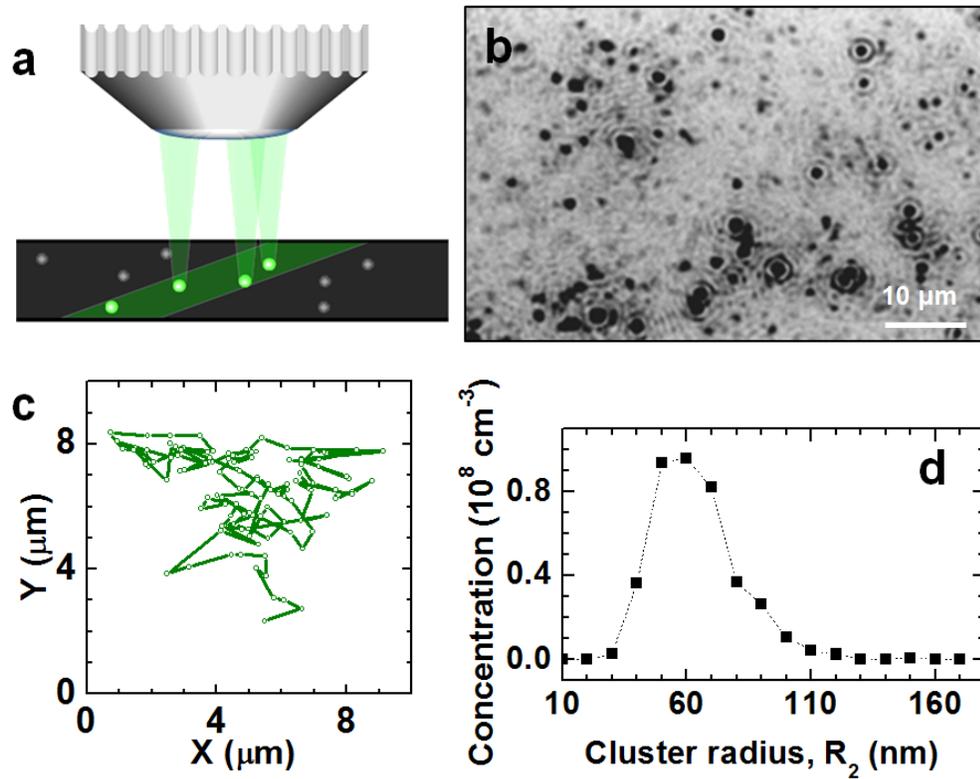


Figure 2-3. Cluster characterization by Brownian microscopy. **a**, Schematic of the BM setup. **b**, A typical BM image shown as a negative. **c**, An example of cluster trajectory. **d**, Distribution of cluster sizes.

The accompanying software package determines the center of these spots in each frame of the movie and builds contiguous cluster trajectories, Fig. 2-3 c. 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. The results are output as the concentration of clusters of a certain size as a function of this size, Fig. 2-3 d.

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 $D_2 \approx 10^{-12} \text{ m}^2 \text{ s}^{-1}$ would be detectable in a focal plane with depth 5 μm for about 25 s. Hence, the objects recorded for < 1 s are likely not clusters and we did not consider them as parts of the cluster population in the determination of the cluster parameters.

The BM method was tested using a solution of latex particles of 200 nm radius in water and was found to faithfully reproduce the particle size and concentration. The distribution obtained with BM are consistent and agree within 10% with the sizes and concentration of the clusters in the same solution determined from DLS data.

CHAPTER 3

MOLECULAR LEVEL VIEW OF OLIGOMERIZATION IN PROTEIN SOLUTIONS

The presence of clusters in protein solutions is confirmed by several studies^{48,52,54}, though the mechanism of cluster formation remains elusive. Pan et al.⁴⁶ propose, from thermodynamics considerations, the formation of new protein species contributing to clustering behavior. They make an attempt to explain the appearance of new species in solutions by partial unfolding of the protein monomers and their oligomerization.

It was clearly established that some proteins reveal domain swapping phenomenon⁶³. This behavior is a sequence of the complex protein structure. The long chain of amino acids tends to arrange itself in order to reach the minimum of the free energy thus implying existence of a definite tertiary structure for each particular protein. Some of the forces contributing to this structure are more stable than others (in comparison, covalent bonds or disulfide bridges are more stable than those that are determined by hydrophobic/hydrophilic or electrostatic interactions). Within one molecule there are well stabilized regions, domains, linked together by a relatively short sequence, hinge. Under special conditions the attractive interaction between the domains may be weakened and the protein will take partially unfolded conformation. As the new conformation will possess higher free energy than the native state of the protein (due to the exposure of the hydrophobic residues) unfolded monomer will be energetically unstable. If two (or more) partially unfolded proteins happen to be in a close proximity the domain swapping can take place. The domain of one molecule will stabilize itself by interaction with the opposite domain from the other molecule. As a result the dimer (or

higher order oligomer) will form out of the initially unfolded monomers. The mechanism of domain swapping is schematically illustrated in Fig. 3-1 b.

The free energy of the new interaction in domain swapped molecules will be lower than for unfolded monomers but still higher than in case of native conformation. As a result the new formed species may show metastable state rather than stable one. This is given In Schlunegger et al ⁶³ present a profound discussion of the domain swapping mechanism, consider the structures that lead to domain swapping, give the general scheme of the process, and show the examples of proteins revealing this behavior.

In the swapped oligomer only hinge sequence has a conformation different from the one in the native state of the molecule. The structure of the hinge is a critical parameter which determines the possibility of the domain swapping⁶³. The length of the hinge has to be enough to keep the monomer in its native state, and the particular sequence of the hinge should be such that allow the formation at least the metastable dimer (oligomer).

Our model system of study is protein lysozyme. The lysozyme folded chain, illustrated in Fig. 3-1 a, contains two stabilized regions, called domains, linked by a relatively short and flexible sequence, called hinge⁷⁶. Two domains of the molecule are shown in pink and orange. The active center is highlighted in red and the hinge is colored in yellow. Yellow spheres represent the disulfide bridges.

Even though there are no reports of domain swapped lysozyme oligomers, theoretical investigations reveal an internal molecular vibrational motion^{76,77}, 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 (^1H - ^2D) rates in native vs denatured states of the molecule⁷⁸ 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.

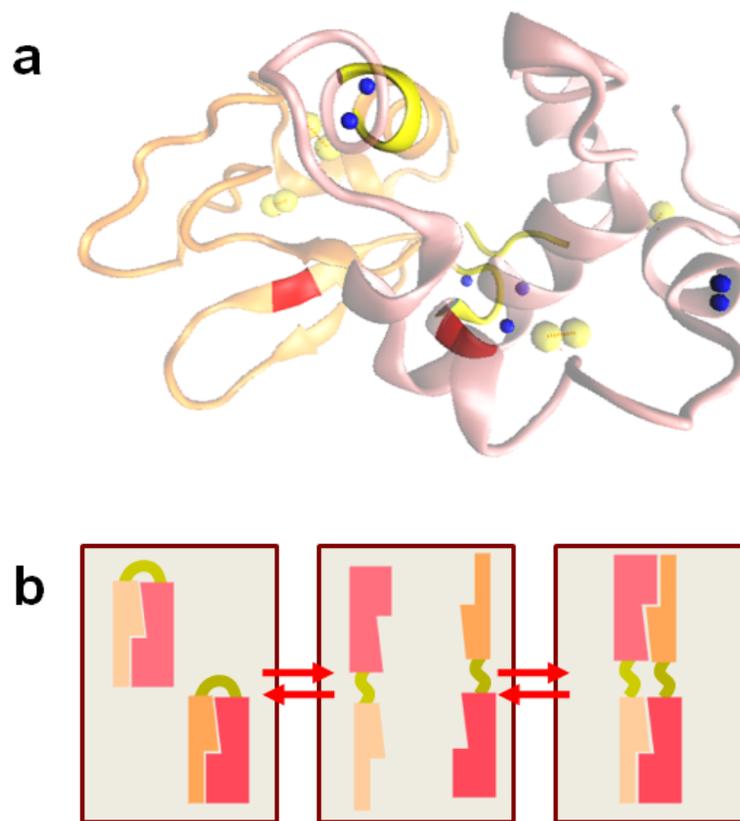


Figure 3-1. Formation of protein oligomers. **a**, Protein lysozyme structure (RCSB PDB: 2VB1). **b**, Schematic representation of the domain swapping mechanism.

3-1. Materials and methods

We used lyophilized 6× crystallized lysozyme powder from Seikagaku (Japan). Two protein solutions of high and low clusters volume fractions were prepared using D₂O-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, Fig. 3.2.

In both solutions, hydrogen-deuterium exchange was allowed to proceed for 1 and 72 hours before it was stopped by quenching the samples to pH 3.8 with a mild solution of acetic acid in D₂O⁷⁹. The 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.

3-2. Monomer flexibility as revealed by nuclear magnetic resonance

The amino acid residues with unexchanged hydrogens yielded signal in ¹H-¹⁵N HSQC 2D spectra, based on the natural abundance of ¹⁵N in the protein. Residues assignment was done by BMRB 4831⁸⁰. The data on the rates of amide hydrogen exchange are shown on corresponding spectra Fig. 3-3. Red labels indicate H-atoms present in dilute, but absent in concentrated solution, black labels – those present in both; unlabeled peaks do not correspond to amide hydrogens.

The results 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. 3-1 as blue spheres.

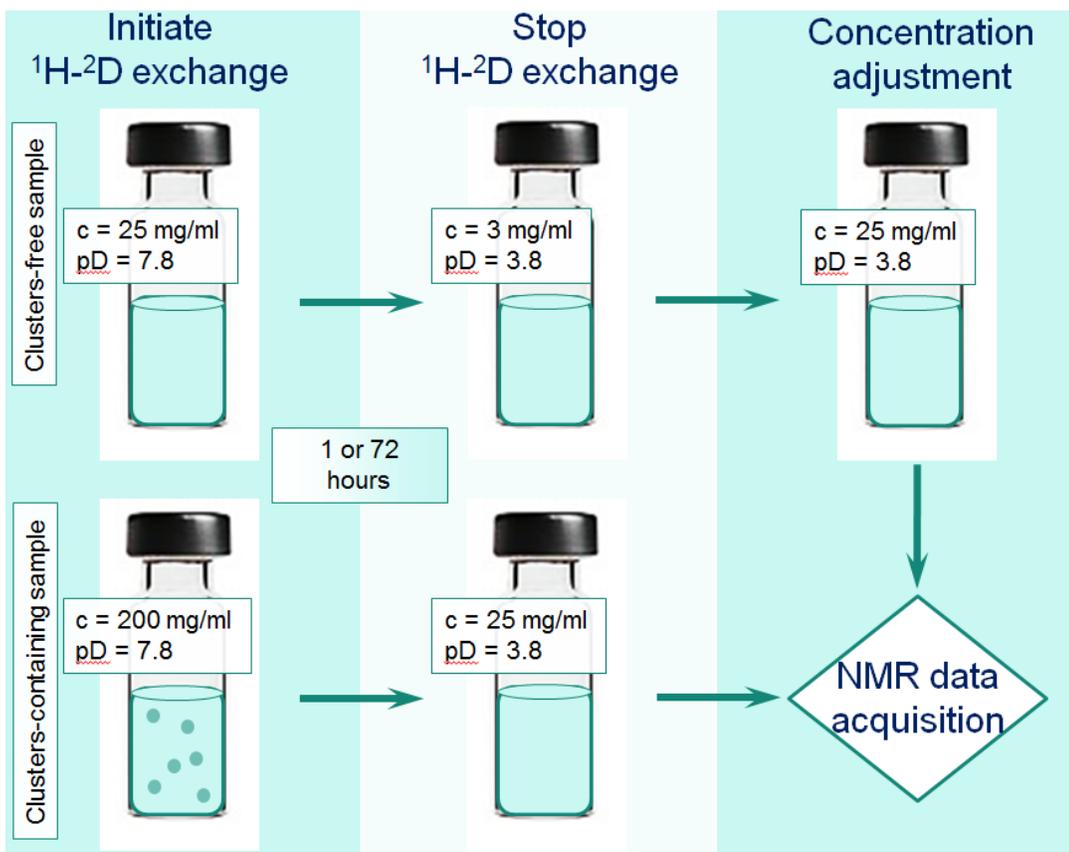


Figure 3-2. Schematic of NMR experiment. Two samples were prepared with initiated ^1H - ^2D initiated. After 1 or 72 hours from preparation ^1H - ^2D was stopped. Both solutions were tested for remaining ^1H - ^{15}N signal by NMR.

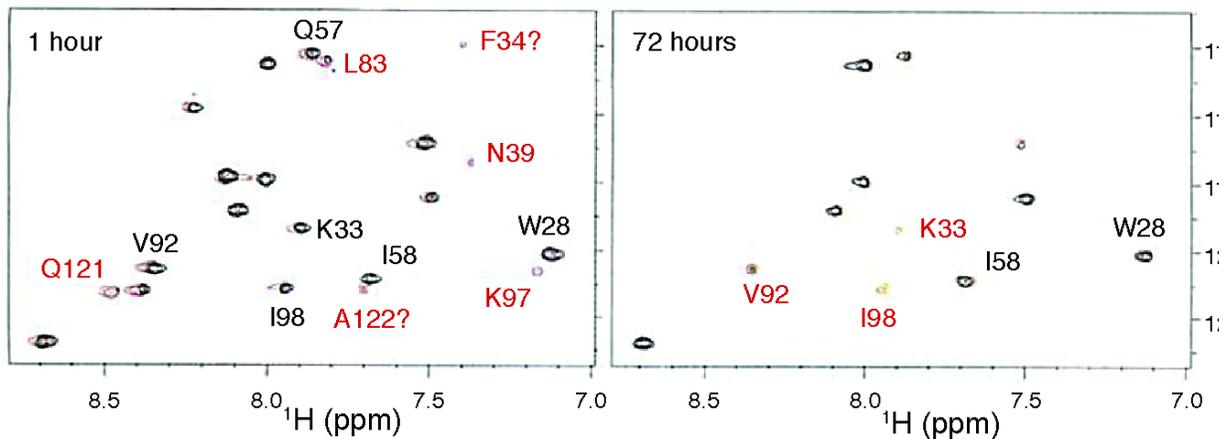


Figure 3-3. ^1H - ^{15}N HSQC spectra of a dilute (in red) and concentrated (black) lysozyme solution at 1 (left) and 72 hours (right) after solution preparation.

3-3. Conclusions

The faster ^1H - ^2D exchange rate in the high concentration solution is counterintuitive for a homogeneously distributed protein and suggests that the protein in the clusters is partially unfolded. Importantly, Fig. 3-1 a shows that the ^1H -atoms exposed in the high concentration solution are near the hinges between the two lysozyme structural domains⁷⁶. These results support the notion that partial unfolding of the domains followed by dimerization, as schematically depicted in Fig. 3-1 b, may be a part of this protein's cluster mechanism.

CHAPTER 4

TERMINAL SIZE IN CLUSTER EVOLUTION

Pan et al.⁴⁶ suggested thermodynamical model describing the existence of liquid clusters in protein solutions. The model states that clusters consist of non-equilibrium mixture of monomers and dimers (or higher order oligomers), and the diffusion reaction scheme accounts for complex formation, complex decay and diffusion of these two populations, monomers and oligomers. Following this reasoning, cluster radius R_2 can be expressed as

$$R_2 = \sqrt{\left(\frac{2k_1 n_1^{out}}{D_1} + \frac{k_2}{D_2}\right)}, \quad (4-1)$$

where k_1, k_2 are rates of dimers (a simple case of oligomers) formation and dimers decay, D_1, D_2 are diffusion coefficients of monomers and dimers respectively, n_1^{out}, n_2^{out} are monomers and dimers number density.

The condition of the local equilibrium far from the core of the cluster is

$$k_1(n_1^{out})^2 = k_2 n_2^{out}. \quad (4-2)$$

Assuming low dimer concentration, $\frac{n_2^{out}}{n_1^{out}} \ll 1$ (experiments show that the volume fraction of clusters is of the order of $10^{-6} - 10^{-7}$) the cluster size becomes

$$R_2 = \sqrt{\left(\frac{D_2}{k_2}\right)}. \quad (4-3)$$

Hence, the cluster radius is determined only by its diffusion coefficient and the rate of dimers decay.

4-1. Materials and methods

The experiments were carried out with the protein lysozyme purchased from Seikagaku (6× crystallized, Japan). Lyophilized powder was directly dissolved in 20 mM HEPES buffer pH 7.8 adjusted with KOH pellets. Protein concentration was adjusted to 85 mg ml⁻¹. The protein solution was filtered into a quartz cuvette (10 mm diameter, Helma) through 0.22 μm filter (PES, Lightlabs, USA). All solutions were kept close to prevent evaporation. Experiments were carried out at 22⁰C, in between the measurements solutions were kept in water thermostat at 22⁰C. DLS measurements were performed as described in chapter 2. The viscosity measurements were carried out as in⁵⁴.

4-2. Results

The previously observed evolution of the cluster size⁵⁰ is in apparent contradiction with the main result of the model, and gradual increase of the clusters size could manifest the possibility that the studied objects are not mesoscopic clusters but domains of a stable phase slowly evolving towards a macroscopic final size. There are four arguments refusing this scenario.

First, the cluster volume fraction preserves low values despite the increase of protein concentration⁵⁰. This observation forbids the principle of macroscopic phase separation.

Second, all tested solutions remained homogeneous over a prolonged observation time of 30 days. No stable phases, liquid or solid, were revealed over this time.

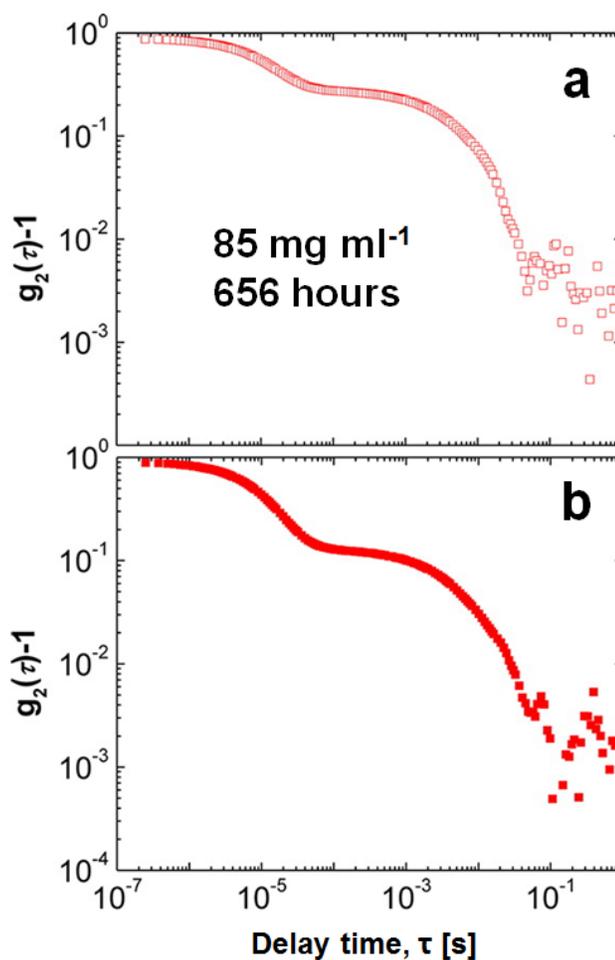


Figure 4-1. Intensity correlation functions from: **a**, a lysozyme solution with shown concentration at shown time after solution preparation, **b**, an identical solution containing latex spheres with $R_{\text{spheres}} = 200 \mu\text{m}$.

Third, if the clusters consisted of a stable dense liquid, their size and volume fraction would increase dramatically faster than the observed evolution. The fast nucleation and growth of the droplets in tested 4% NaCl and 50 mM Na acetate buffer conditions is consistent with previous microscopy observations^{81,82}. Fourth, we experimentally proved that clusters are constrained to their mesoscopic sizes and extremely low volume fractions even after extended periods of equilibration with the solution.

Fig. 4-1 a displays the correlation function from a solution similar to studied in⁵⁰ at 656 h after solution preparation. A shoulder with characteristic $\tau_1 = 20 \mu\text{s}$ corresponds to lysozyme monomers and a shoulder with $\tau_2 = 26 \text{ ms}$ to clusters. This function reveals a $\sim 3\times$ slower diffusion of clusters compared to those, observed after the solution preparation. To understand this slowing down, we monitored in parallel the dynamics of latex spheres of 212 nm radius (carboxylate-modified polystyrene microparticles, OptiLink, USA) suspended at 10^{-5} volume fraction in an identical solution. The correlation function from the solutions with the suspended spheres is illustrated in Fig. 4-1 b. A shoulder at $\tau_1 = 20 \mu\text{s}$ corresponds to lysozyme monomers and a shoulder at $\tau_2 = 30 \text{ ms}$ to the latex spheres. Since the spheres scatter light significantly better than the clusters, light scattered from the clusters is not detected. The correlation function indicates that the solution viscosity has increased by $\sim 3\times$, from 1.45 to 4.5 mPa s. The characteristic diffusion time of the spheres in Fig. 4-1 b is similar to that of clusters in Fig. 4-1 a, hence, the cluster radius is about $R_2 = 200 \text{ nm}$. Using this radius, we determine the cluster volume fraction from the intensity scattered by them as $\varphi_2 \approx 2 \times 10^{-6}$.

4-3. Conclusions

Substantial amount of information is already available on clusters and in all studies of their behavior they are assumed to represent a new aggregation phenomenon in homogeneous protein solutions. Fair thermodynamic considerations, do not prove the uniqueness of clusters formation and a new argument was needed to validate the assumption that clusters are the new population in protein solutions and not the droplets of a dense liquid phase. Here, we presented the experimental results which showed that the cluster population reaches its terminal size upon a prolonged ripening and the clusters

do not coalesce in a new dense liquid phase. Along with other arguments in detail explained in Li et al.⁵⁰, this statement closed the question whether clusters are just nuclei of phase separation or the independent population in protein solutions in a favor of the last assumption – clusters are the new type of aggregates in homogeneous protein solutions and not the domains of a stable phase.

CHAPTER 5

LACK OF DEPENDENCE OF THE SIZES OF THE MESOSCOPIC PROTEIN CLUSTERS ON ELECTROSTATICS

The protein clusters result from interplay of monomer influx, oligomer formation, and subsequent oligomer outflow and decay. Reaction-diffusion schemes are derived within the hydrodynamic description by connecting the rate of approach to equilibrium and the degree of deviation from equilibrium⁸³. 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 $D_{oligomer}$ and decay rate constant $k_{oligomer}$ ⁸⁴.

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 tested. Here we probe effects of Coulomb interactions on the properties of the mesoscopic clusters as a test of the colloid clustering scenario.

From a fundamental perspective, Coulomb forces determine protein 3D structure^{85,86}, substrate binding^{87,88}, enzyme activation^{89,90}, signal transduction⁹¹, etc. Importantly, Coulomb forces govern two major classes of protein aggregation: amyloid fibrillation⁹²⁻⁹⁴ and crystallization^{95,96}. 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 oligomers⁹⁷, 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. 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.

5-1. Materials and methods

Reagents and solutions. We purchased lyophilized lysozyme from Affymetrix. We also used KCl (Fisher), $(\text{NH}_4)_2\text{SO}_4$ (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 $\epsilon=2.64 \text{ ml mg}^{-1} \text{ cm}^{-1}$ at 280 nm⁹⁸. We prepared a stock solution of $\sim 150 \text{ mg ml}^{-1}$ 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 $(\text{NH}_4)_2\text{SO}_4$. For static light scattering measurements the dialyzed solutions were diluted to an initial concentration $\sim 35\text{-}40 \text{ mg ml}^{-1}$. All experiments were done at 22°C. Prior to all measurements, the solutions were filtered through 0.22 μm PES syringe filters (Lightlabs, USA).

Solutions with $\text{pH} < 7.8$ 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) at $\text{pH} = 7.8$, and titrating this solution to the desired pH with 0.10 M HCl.

Estimation of solution ionic strength. The pKa of HEPES is 7.5, hence at $\text{pH} = 7.8$, 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. $0.667\times$ 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 $\text{pH} = 7.8$ 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 $\text{pH} = 7.8$ to +15 at $\text{pH} = 3.8$ 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 $100 \text{ mg ml}^{-1} = 6.8 \text{ mM}$. Hence, the ionic strength is 64 mM at $\text{pH} = 3.8$. Accounting for the lower protein charge (9.5 at $\text{pH} = 5.0$ and 8.5 at $\text{pH} = 6.5^{99}$), $I = 45$ and 42 mM at $\text{pH} = 5.0$ 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 $\text{pH} = 3.8, 5.0$ and 6.5, respectively. Lower I leads to higher D_1^{conc} and φ_2 , as shown in a results section, 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). 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 algorithm¹⁰⁰ and a modified cumulant method introduced in Li et al.¹⁰¹. 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^{55,101}. 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 using carboxylate-modified polystyrene particles (OptiLink, USA) with diameter $0.424 \text{ }\mu\text{m}$ suspended in 100 mg ml^{-1} solution of lysozyme in HEPES buffer at pH 7.8; the data are shown in Fig. 5-1. Lower curve is the dependence of viscosity on the concentration of HEPES; viscosity is not affected by the addition of NaCl, KCl, and

(NH₄)₂SO₄, used to adjust the ionic strength, or KOH and HCl, used to adjust pH. Upper curve is the dependence of viscosity on the concentration of urea in 20 mM HEPES at pH = 7.8. Lines are just guides for the eye.

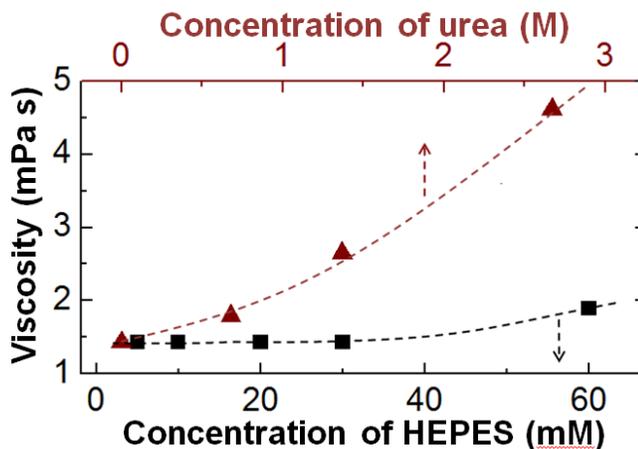


Figure 5-1. The viscosity of 100 mg ml⁻¹ lysozyme solutions hosting the protein-rich clusters.

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 was assumed to be fully pairwise. To facilitate sampling of mutual orientations of two molecules, with collaboration with Dr. Lubchenko, Chemistry Department, UH, we modeled a protein molecule as a dielectric sphere, as in the Kirkwood-Tanford model¹⁰². The sphere radius was chosen at 1.7 nm so that its volume matches that of an actual lysozyme molecule. A charged residue was represented by 1, 2, or 3 point charges depending on the number of distinct charged atoms in the residue. For instance, an

(NH₃)⁺ 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 adopted $\varepsilon = 2$ inside and $\varepsilon = 78$ outside the protein molecule, as in water. In addition, we partially accounted 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 was set to 22^oC.

We modeled 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.

$$E_{mol} = \begin{cases} E_{>}, & r_s > r_2 \\ E_{<}, & r_s < r_1' \end{cases} \quad (5-1)$$

$$E_{>} = -\frac{A_H}{12} \left(\frac{1}{(x+1)^2} + \frac{1}{(x^2+2x)} + 2 \ln \left(\frac{x^2+2x}{(x+1)^2} \right) \right), \text{ and} \quad (5-2)$$

$$E_{<} = 4\varepsilon \left[\left(\frac{\sigma}{r_s+\delta} \right)^{2\alpha} - \left(\frac{\sigma}{r_s+\delta} \right)^\alpha \right], \quad (5-3)$$

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 $3.1 k_B T$.^{103,104} 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 Table 5-1.

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 virial coefficient, B_{22} , is computed as^{105,106}:

$$B_{22} = -\frac{2\pi N_A}{M^2} \langle \int_0^\infty (e^{-\frac{E_{PMF}}{k_B T}} - 1) r_c^2 dr_c \rangle, \quad (5-4)$$

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

Table 5-1. Values of the parameters of the intermolecular interactions.

Variable	Value	Variable	Value
a	12	c_1	16.32
σ	1.64	c_2	90.04
r_1	0.33	c_3	-186.64
r_2	0.73	c_4	173.09
c_0	-0.25	c_5	-60.31

5-2. Model system

Our model protein, lysozyme, has an isoelectric point at $\text{pH} = 11.35^{107}$, which is one of the highest documented for any protein¹⁰⁸. As a result, even at the highest $\text{pH} = 7.8$ probed here, protonation of basic and acidic surface amino acid groups leads to a significant +8 net charge of the lysozyme monomer⁹⁹; 17 positive and nine negative groups were identified at this pH in ref.⁹⁷. At the lowest $\text{pH} = 3.8$ tested here, the net charge increases to +15⁹⁹. This high net positive charge, illustrated in Fig. 5-2, amplifies the significance of the Coulomb forces for aggregation behaviors of lysozyme and makes this protein a suitable model system for the present work. In Fig. 5-2 PBD structure file 2VB1 was used. The protonation state of each acid or basic residue was evaluated at the chosen pH with PROPKA 3.0. The electrostatic map was computed online with the Adaptive Poisson-Boltzmann Solver (APBS) and drawn using PyMOL (www.pymol.org).

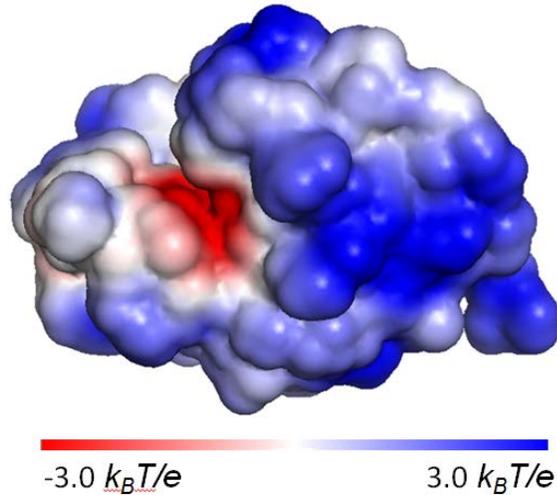


Figure 5-2. The distribution of electrostatic potential of the solvent-accessible surface of a lysozyme molecule at pH = 7.8.

5-3. 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 screening¹⁰⁹, and the solution pH, which determines the protein's charge. We characterize the intermolecular interactions in terms of three parameters.

(1) The second osmotic virial coefficient B_2 , which is an azimuthally and spatially averaged characteristic of the pairwise interaction potential¹¹⁰. We obtain B_2 from the slopes of Debye plots, determined by static light scattering and presented in Fig. 5-3. The ionic strength in these solutions increases from 3.3 mM to 40 mM in Fig. 5-3 a; and from 13.3 mM to 313 mM in Fig. 5-3 b and c. At low c , $Kc/R_\theta = M_w^{-1} + 2B_2c$, where B_2 is the second osmotic virial coefficient and $M_w = 14,300 \text{ g mol}^{-1}$ 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 high ionic strength > 100 mM the effect of electrolyte concentration on B_2 and the intermolecular interactions becomes weak.

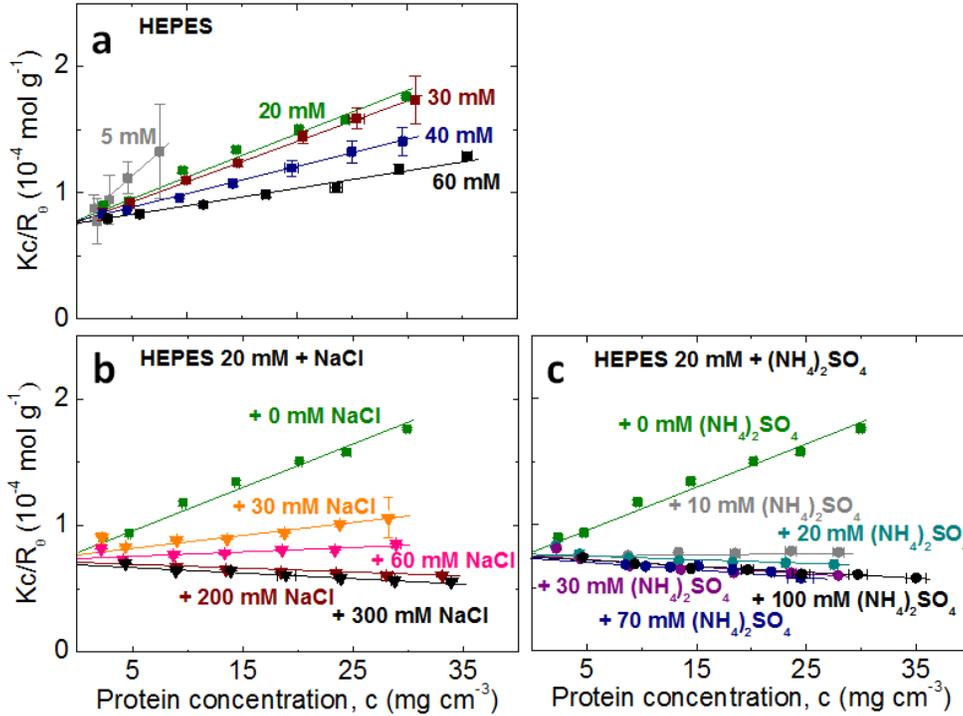


Figure 5-3. Characterization of the intermolecular interaction in lysozyme solutions at increasing ionic strength. Debye plots $Kc/R_\theta(c)$ of lysozyme solutions. **a**, in HEPES buffer; **b**, in 20 mM HEPES with added electrolytes.

(2) The diffusivity of protein monomers in dilute solutions D_1^{dilute} . (3) The same diffusivity in concentrated solutions D_1^{conc} . D_1^{dilute} and D_1^{conc} were determined from the faster shoulder in the autocorrelation function of the intensity of light scattered off the solution, seen in Fig. 5-4. All solutions contain 100 mg ml^{-1} lysozyme in HEPES buffer at $\text{pH} = 7.8$. The autocorrelation functions 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.

The viscosities of all solvents used in this study were practically independent of the salt concentration and identity. Hence, D_1^{dilute} and D_1^{conc} trends in Fig. 5-5 b indicate the response of the protein dynamics to variations in intermolecular interactions, where D_1^{dilute} , measured at 9 mg ml⁻¹, open symbols and D_1^{conc} , measured at 100 mg ml⁻¹, solid symbols. The Stokes-Einstein diffusivity (or self-diffusivity) $D_0 = 1.20 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ 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. We note that while B_2 and D_1^{dilute} account for the interactions at long intermolecular separations, D_1^{conc} is weighted towards short separations.

The dependences of B_2 , D_1^{dilute} , and D_1^{conc} on the ionic strength I are displayed in Figs. 5-5 a and b. We have varied I from 3 to 333 mM by increasing the concentration of HEPES buffer or adding NaCl, KCl, or (NH₄)₂SO₄. The values of B_2 computed using the model are also shown. $B_2(\text{hs}) = 4 V_M N_A M_w^{-2} = 2.35 \times 10^{-4} \text{ mol m}^3 \text{ kg}^{-2}$ for hard spheres ($V_M = 2.0 \times 10^{-26} \text{ m}^3$, molecular volume; N_A , Avogadro's number; $M_w = 14.5 \text{ kg mol}^{-1}$, lysozyme molecular weight) is shown for comparison. 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^{dilute} , and D_1^{conc} at increasing I are consistent with the expectation that the Debye screening due to free ions significantly weakens the Coulomb repulsion. At $I > 100 \text{ mM}$, B_2 becomes lower than its value for hard spheres, while D_1^{dilute} drops below the value of lysozyme's self-diffusivity. Both observations

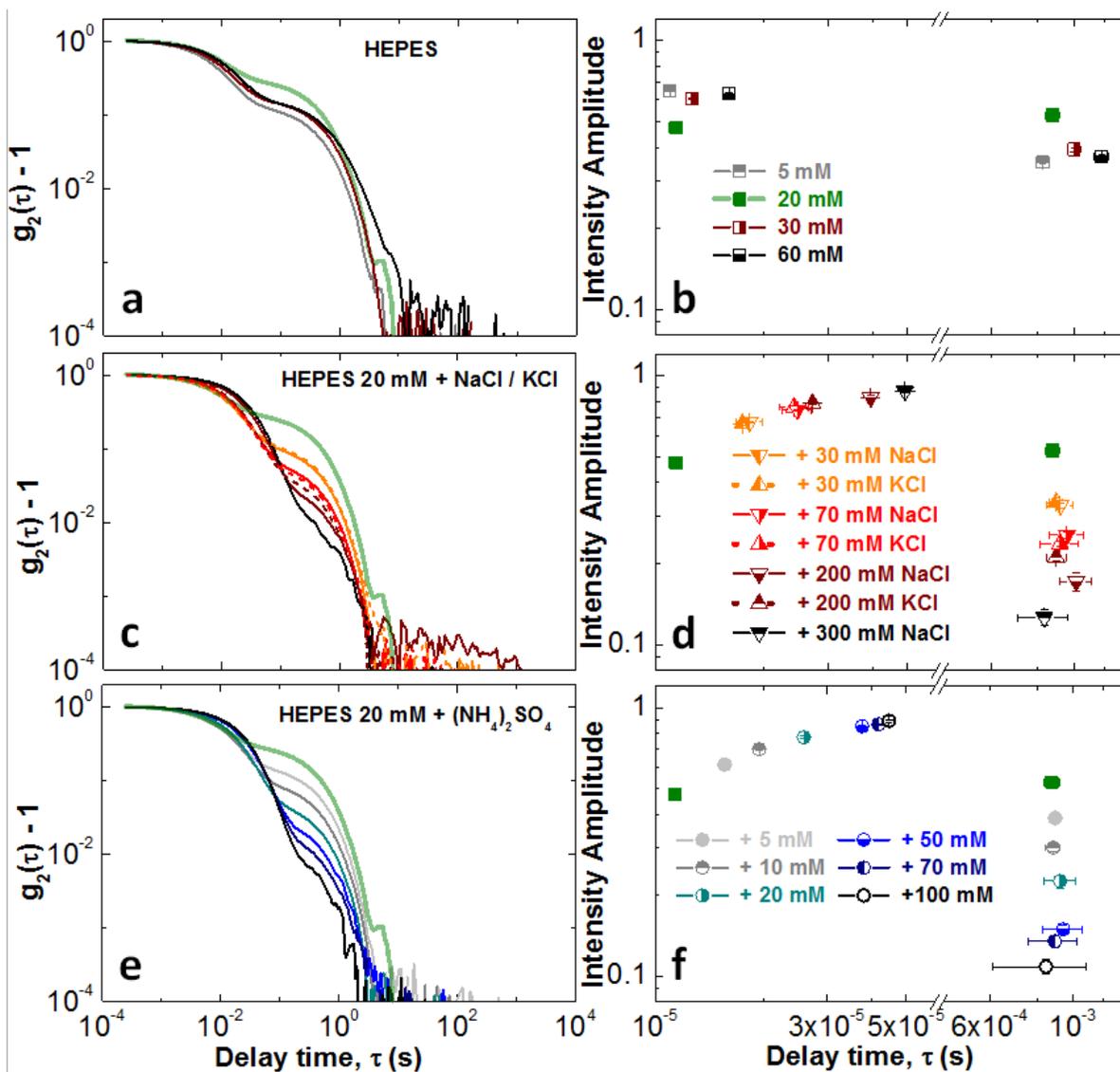


Figure 5-4. Monomers and clusters in lysozyme solutions at varying ionic strength and salt identity, characterized by DLS. **a, c, e** Normalized autocorrelation functions $g_2(\tau) - 1$. **b, d, f**, The amplitudes of monomers and clusters.

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^{conc} exhibits a stronger dependence on I than D_1^{dilute} , indicating that Coulomb repulsion is more sensitive to electrolyte concentration at short than at long separations.

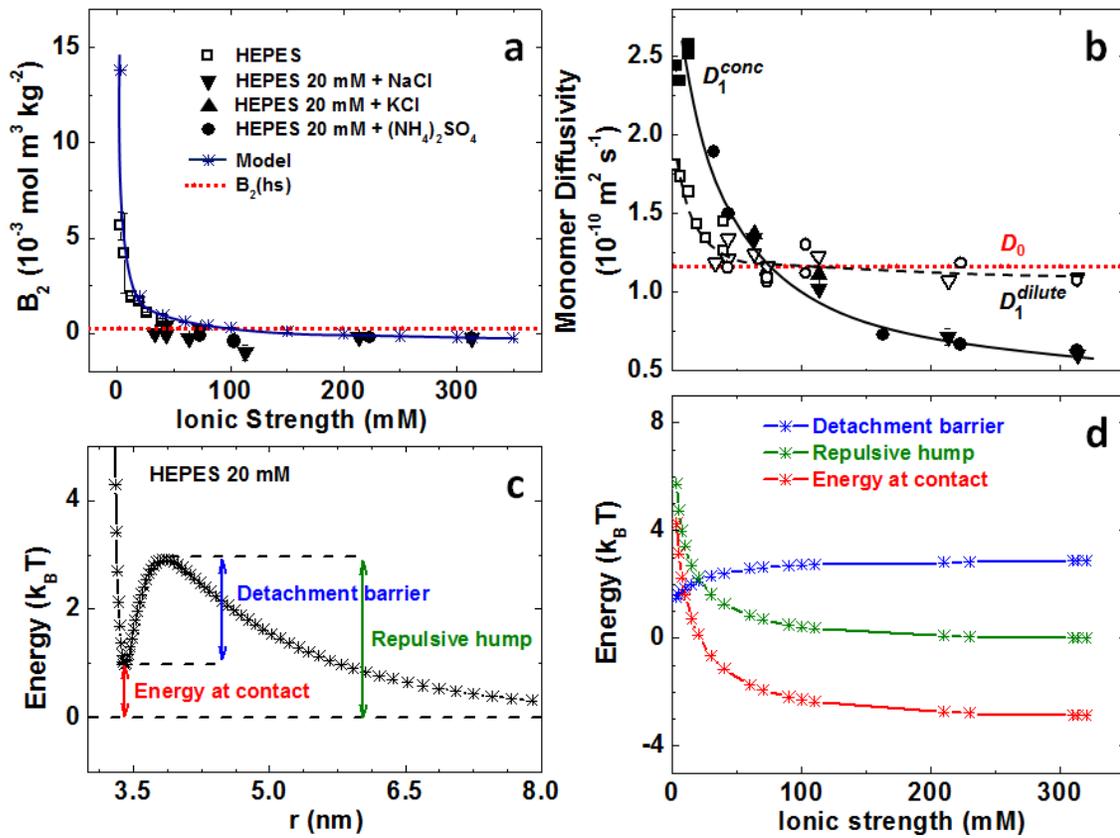


Figure 5-5. Characterization of the intermolecular interactions in solution at pH = 7.8.

The dependences of B_2 , D_1^{dilute} , and D_1^{conc} on pH, displayed in Figs. 5-6 a and b, reveal that the values of the three parameters decrease as pH increases. This is expected: since 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^{conc} in Fig. 6 b is stronger than in D_1^{dilute} , again implying that Coulomb repulsion is more sensitive to the protein's charge at short than at long separations.

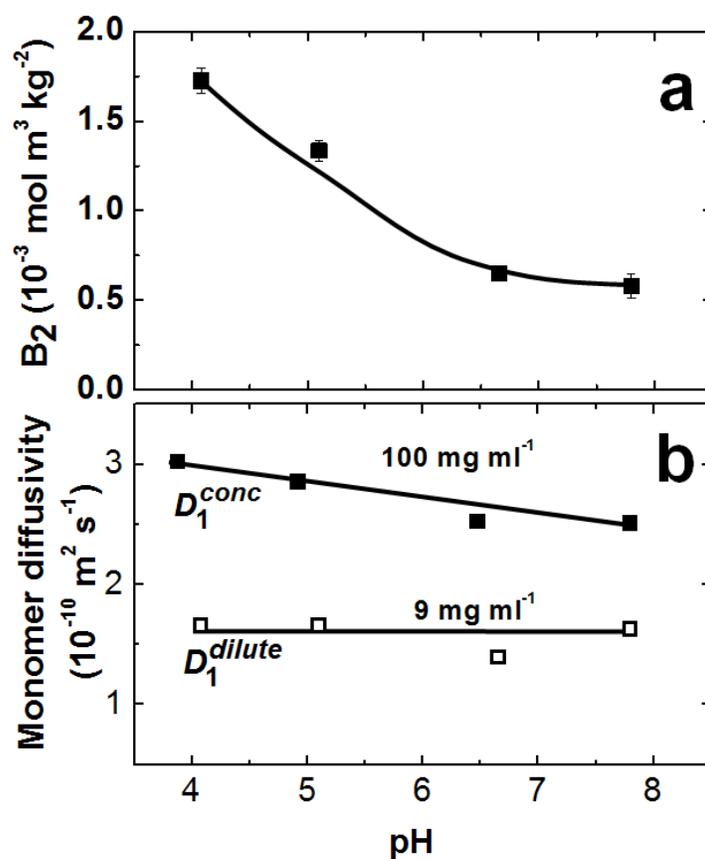


Figure 5-6. pH effects on the intermolecular interactions. **a**, Dependence of the second osmotic virial coefficient B_2 on the solution pH. **b**, Dependencies of the monomer diffusivities in dilute and concentrated solutions.

While the effects of pH on B_2 , D_1^{dilute} , and D_1^{conc} link directly to the decrease of the molecule's charge at higher pH, the correlation between B_2 and I displayed in Fig. 5-5 a requires additional discussion. We employ a computational model following Chan *et al.*⁹⁷. We represent every protein molecule as a sphere with discrete charges as illustrated in Fig. 5-7. The positive and negative surface charges are shown in red and blue, respectively. In Fig. 5-7 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 ⁹⁷. In Fig. 5-7 b we show 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.

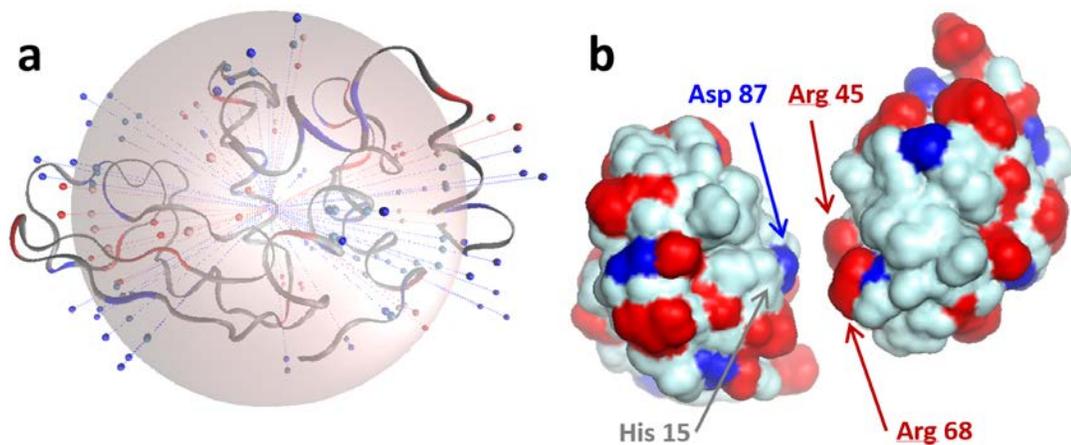


Figure 5-7. The charged groups on the surface of a lysozyme molecule. PDB structure file 2VB1 was used.

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. 5-7 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 weights⁹⁷. The resulting angular-averaged potential of mean force (PMF) at $I = 13.3$ mM and $\text{pH} = 7.8$ is shown in Fig. 5-5 c. Three characteristics of this PMF relevant to the discussion of aggregation include the energy at contact, association barrier, i.e., repulsive “hump”, and detachment barrier; their dependences on I are displayed in Fig. 5-5 d.

The values of B_2 computed using these pairwise potentials are shown in Fig. 5-5 a. The predictions of the model agree well with the experimental data at high ionic strengths, and slightly underestimate 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.

5-4. 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 cluster population φ_2 to the solution ionic strength I are displayed in Figs. 5-8 a and b, where in Fig. 5-8 a we show the 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. In Fig. 5-7 b, cluster volume fraction φ_2 , is determined from the same DLS autocorrelation functions as R_2 , as a function of the solution ionic strength. In Fig. 5-7 c-d, we show the evolutions of cluster size R_2 in Fig. 5-7 c and volume fraction φ_2 in Fig. 5-7 d, determined by DLS and BM. The results of

the two methods are similar, within their errors; the inherent error of the φ_2 determination may be up to 50 %¹⁶.

R_2 is independent of I , within the experimental error, while φ_2 decreases approximately four-fold as I increases from 3 to ca. 100 mM and saturates at higher I values. Figs. 5-8 c and d reveal that the values of R_2 and φ_2 are relatively steady in time (the slow R_2 growth likely reflects an Ostwald-like ripening of the clusters¹¹¹). The effects of pH on R_2 and φ_2 are displayed in Fig. 5-9 a and b. Similarly to the trend in Fig. 5-8 a, the cluster radius R_2 depends weakly on solution pH; several repetitions of this experiment revealed no pH dependence. The cluster volume fraction φ_2 increases by ca. 4× as pH increases from 3.8 to 7.8.

The apparent increase in the cluster volume fraction φ_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. 5-6 b are consistent with these expectations. However, Figs. 5-8 and 5-9 also reveal at least three anomalous cluster behaviors.

First, the behaviors of R_2 and φ_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. The second anomaly is 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. 5-5 a and b and 5-6 a and b. The third

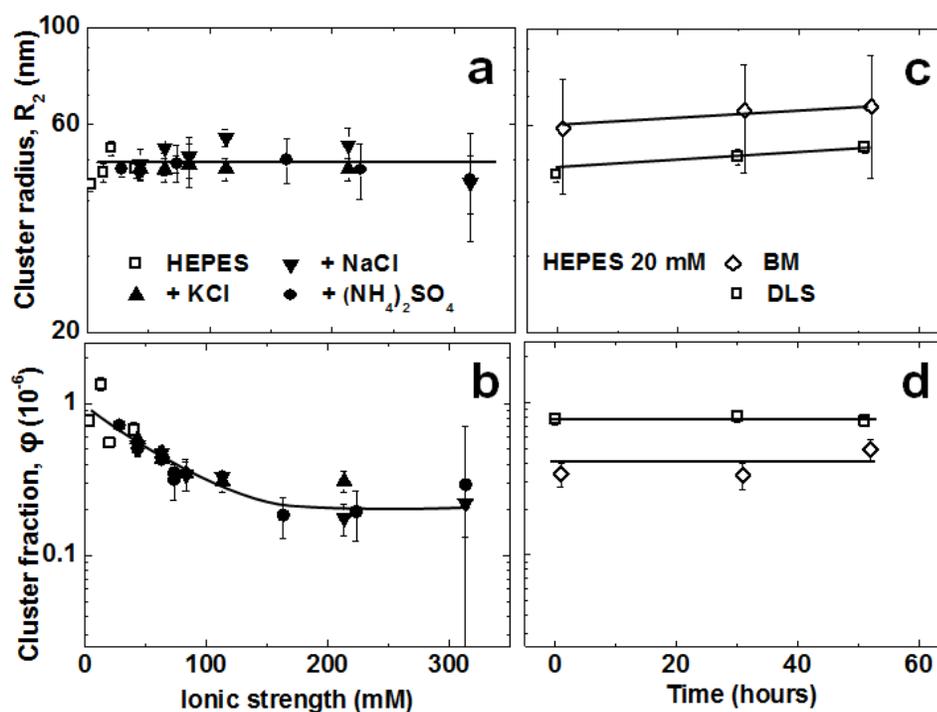


Figure 5-8. Populations of protein-rich clusters in 100 mg ml⁻¹ lysozyme solutions in HEPES buffer at pH = 7.8 characterized by dynamic light scattering (DLS) and Brownian microscopy (BM).

puzzling behavior is the decreasing cluster volume fraction ϕ_2 at high ionic strength I in Fig. 5-8 b. This contradicts the expectation that weaker repulsion in concentrated solutions, revealed by the D_1^{high} trend in Fig. 5-5 b, should lead to cluster stabilization and, hence, a higher cluster volume fraction, similarly to the observations at higher pH in Fig. 5-9 b. Note that the three anomalous behaviors contradict general rules of phase transformations and solution thermodynamics and not a specific model of cluster formation. Hence the conclusions on the cluster mechanism that emerge from their resolution, discussed below, are not confined to such a model.

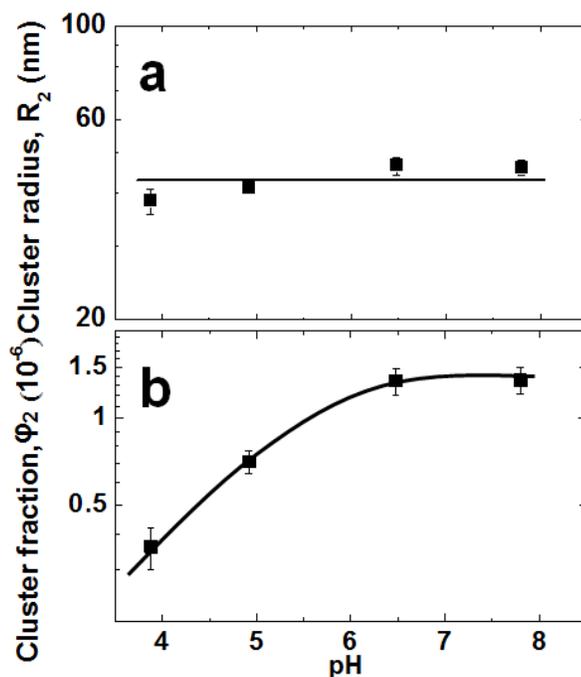


Figure 5-9. pH effects on the cluster characteristics, the cluster radius R_2 in **a**, and the cluster volume fraction φ_2 in **b**.

The decoupled behaviors of R_2 and φ_2 at increasing I indicate that R_2 and φ_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 clusters⁸⁴, while φ_2 reflects the high free energy cost of bringing together positively charged molecules^{84,111}. 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^{112,113}, 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 sub-section.

Another example of Coulomb-independent behavior is presented by proteins of the γ -crystalline family¹¹⁴. 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 γ -crystalline retain a size of about 100 nm as pH is varied from 6.8 to 10 and the NaCl concentration, from 150 to 350 mM¹¹⁴. (The cluster disaggregation at pH and NaCl concentration values outside these ranges¹¹⁴ may be due to a protein-specific mechanism 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^{84,111,114-117}, 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.

5-5. Water-structuring interactions and partial protein unfolding

The anomalous decrease of φ_2 at high I in Fig. 5-8 b is akin to salting-in, 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 Figs. 5-5 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 ca. 500 mg ml⁻¹⁸⁴ and the intermolecular separation is shorter than 1 nm⁸⁴. Possible candidates are water-structuring forces that operate at similarly short separations¹⁰⁹. They are classified either

as hydration, due to water structuring at polar surface patches and augmented by the presence of ions and other kosmotropes^{109,118}, or hydrophobic, due to water layering along non-polar surface patches¹⁰⁹. 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 unfolding¹¹⁹⁻¹²¹. The current consensus appears to be that urea is a universal denaturant since it interacts favorably with the peptide backbone¹²². The aminoacid side chains assist the action of urea by additional preferential interaction with it and by diluting the effective concentration of the backbone amides¹²³⁻¹²⁵. The interactions of urea with the backbone and side chains involves intercalation and destruction of the water structures (chaotropic action)¹²⁶. Ethanol forms homogeneous solutions with water at concentrations below ca. 2.8 M¹²⁷. Similarly to urea, ethanol is a chaotropic agent, however, it accumulates in the vicinity of nonpolar aminoacid residues and disrupts adjacent water structures: it strips off as many as 16 bound water molecules from the lysozyme surface¹²⁸ and may form hydrogen bonds to its hydroxyl groups¹²⁹. 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 α -helixes and β -sheets at concentrations as high as 2.5 M¹²⁹.

We have characterized the effects of urea and ethanol at concentrations up to ca. 2.5 M on the protein interactions in terms of the second osmotic virial coefficient B_2 (determined from plots in Fig. 5-10) 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. 5, 6 and 8, urea and ethanol significantly affect the buffer viscosity at the applied concentrations). 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. In Fig. 5-10 a-b we show variation of the second virial coefficient B_2 (left ordinate, closed symbols) and of the product of monomer diffusion coefficient D_1^{conc} in 100 mg ml⁻¹ solutions and the buffer viscosity η_{buffer} (right ordinate, open symbols) as functions of the concentration of urea in Fig. 5-10 a and ethanol in Fig. 5-10 b. Inset in Fig. 5-10 a is the native structure of lysozyme and its solvent-accessible surface. Locations at which the peptide backbone is exposed to the solvent are highlighted with arrows. In Fig. 5-10 c-d we show the response of the cluster radius R_2 (left ordinate, closed symbols) and volume fraction φ_2 (right ordinate, open symbols) to increasing concentrations of urea in Fig. 5-10 c and ethanol in Fig. 5-10 d. The results in Fig. 5-10 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 aminoacid residues. In addition to them, 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. 5-10 a; the 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.

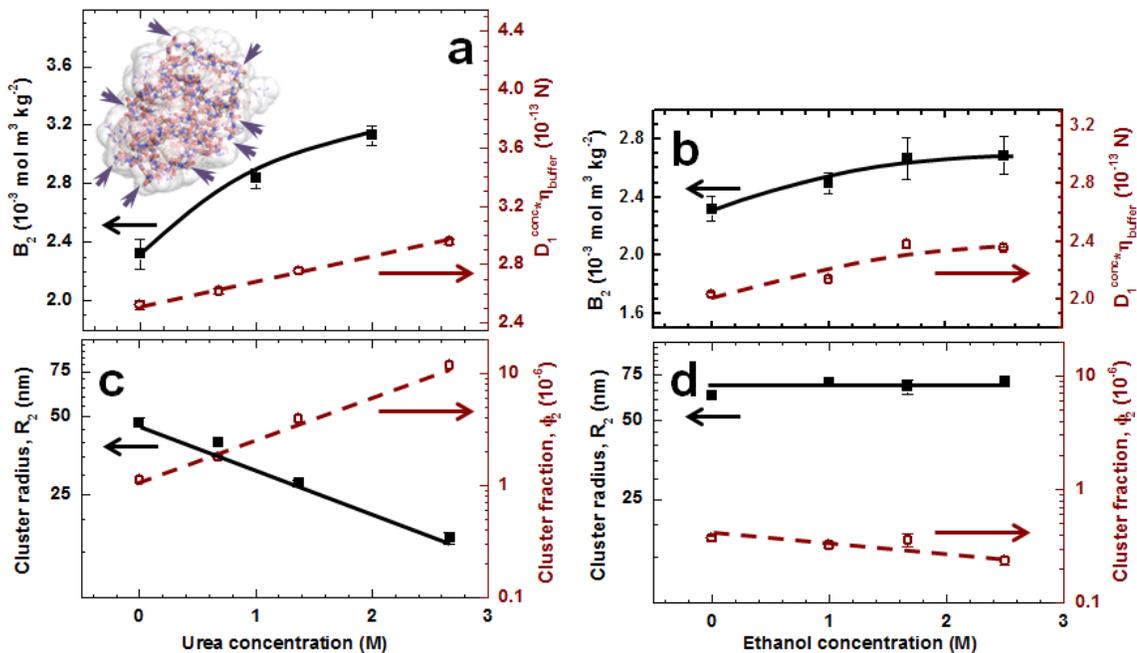


Figure 5-10. The role of hydrophobic interactions and partial unfolding in cluster formation.

While with ethanol the increase in B_2 (ca. 15%) is comparable to that of $D_1^{\text{conc}} \eta_{\text{buffer}}$, with urea the increase in B_2 (ca. 40%) is stronger than that of $D_1^{\text{conc}} \eta_{\text{buffer}}$ (ca. 20%). As discussed above, $D_1^{\text{conc}} \eta_{\text{buffer}}$ is weighted towards short intermolecular separations. Hence, this discrepancy indicates that urea boosts more repulsion at the long range. Since the discrepancy is not observed with ethanol, we assign it to urea-enhanced partial protein unfolding. It exposes to the solvent nonpolar sidechains that in the native structure are tucked inside. Urea does not interact with the

exposed sidechains. 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. 5-10 c demonstrate that the addition of urea reduces the cluster radius R_2 about three-fold, while increasing the cluster population volume fraction φ_2 by an order of magnitude. The addition of ethanol does not affect R_2 and weakly lowers φ_2 , Fig. 10 d. The decoupled behaviors of R_2 and φ_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 φ_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 φ_2 (Fig. 5-11) were practically identical to those in directly prepared 1.25 M urea, indicating that cluster formation and its constituent processes are reversible. The data in Fig. 5-11 was obtained from 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. 5-10 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.

Second, we monitored the evolution of the cluster population over 24 hours. We found (Fig. 5-12) that R_2 and φ_2 did not change from the values established within 30 min

after the addition of urea, Fig. 5-10 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 urea⁸⁴.

The R_2 and φ_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 formation, according to which R_2 and φ_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 $k_{oligomer}$ and lead, according to Eq. (1), to smaller clusters. Since ethanol does not interact with the backbone, it does not affect R_2 . The responses of φ_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 $D_1^{conc}\eta_{buffer}$ trends in Fig. 6 a) exposes hidden nonpolar aminoacid residues. Since the attractive hydrophobic interactions between them are short-ranged, this stabilizes the cluster phase more than the dilute solution and increases φ_2 .

Looking back at the effects of electrolytes, we note that one of the used salts, $(\text{NH}_4)_2\text{SO}_4$, 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 proteins¹³⁰.

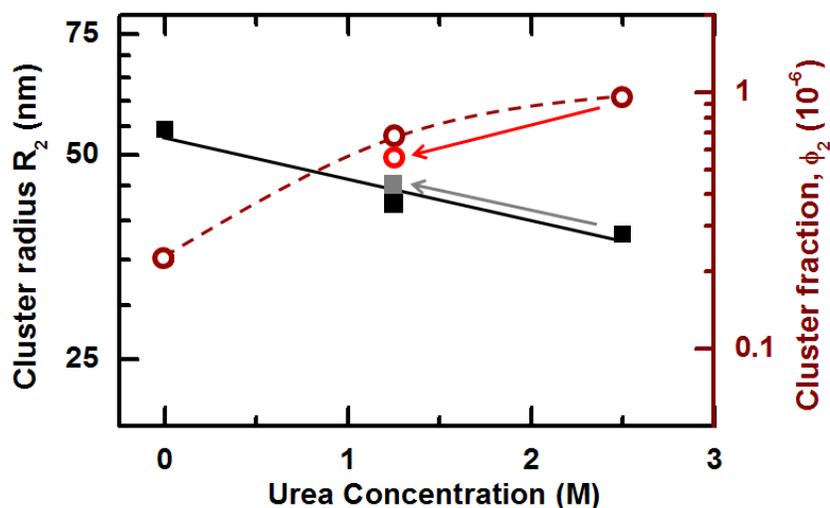


Figure 5-11. The response of the cluster radius R_2 (left ordinate, closed symbols) and volume fraction ϕ_2 (right ordinate, open symbols) to increasing or decreasing concentrations of urea.

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 $(\text{NH}_4)_2\text{SO}_4$ is fully described by the charges of its constituent ions via the ionic strength I .

The responses of R_2 and ϕ_2 to the presence of urea and ethanol in Fig. 5-10 are not dramatic, implying that the cluster formation mechanism have not been modified by these two additives. These responses identify partial protein unfolding as the likely force 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 aminoacid residues, hidden in the native conformation, enables hydrophobic bonds between backbone segments, and stabilizes the cluster phase

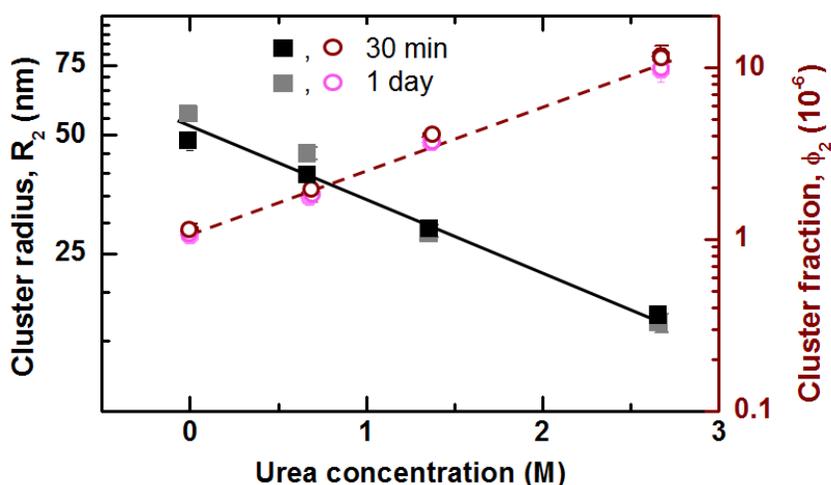


Figure 5-12. The consistency of the radius R_2 and volume fraction ϕ_2 of the clusters in the presence of urea. Solutions were characterized 30 min and 1 day after preparation. The data for 30 min are from Fig. 5-10 c.

through hydrophobic attraction between the exposed nonpolar aminoacid 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 formation⁸⁴, wherein backbone-to-backbone contacts support transient oligomers.

5-6. Purging of lysozyme solution with gases induces shear stress, crucial for lysozyme cluster formation

Purging (or 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 interface¹³¹ or because of the induced shear stress.¹³² Meanwhile, the gas bubbled through the sample causes the change of chemical composition of solutions depending on the gas nature. Our tests aimed to test the significance of hydrophobic forces after the partial unfolding during bubbling and the possibility of participation of chemical bridges in clusters formation, such as -S-S- and -NH₂-OCO-NH₂-

bridges. The increase of oxidative potential of a solution, which can be achieved by bubbling of oxygen through the sample, would break these bridges and lead to a different cluster properties. We performed experiments in which lysozyme solutions were bubbled with different gases: air (laboratory pipeline), He, N₂, O₂ (Praxair, USA) at different flow rates ranging from 0.2 to 0.5 L/min over the course of 2.5 hours. The viscosity of the solutions does not change over the bubbling time and remains 1.42 cP.

We started our tests with characterization of monomers response to applied changes, monomer diffusivity D_1 and the second virial coefficient B_2 , Fig 5-13. In Fig. 5-13 a we show the response of the monomer diffusivity at the highest gas flow rate of 0.5 L/min over the 2.5 hours of bubbling. The first point represent the standard, i.e., the value of monomer diffusivity D_1 before bubbling. In Fig. 5-13 b we display static light scattering experiment data on a standard solution (no bubbling) and solution after 2.5 hours bubbling at the highest gas flow rate 0.5 L/min. We notice that the bubbling did not change the monomer-monomer interactions characterized by the second virial coefficient B_2 , extracted from the slope of these dependences. As in the case of unperturbed solutions, the second virial coefficient $B_2 = 1.89 \times 10^{-3} \text{ mol m}^3 \text{ kg}^{-2}$. These observations indicate that the majority of protein monomers was not affected by bubbling.

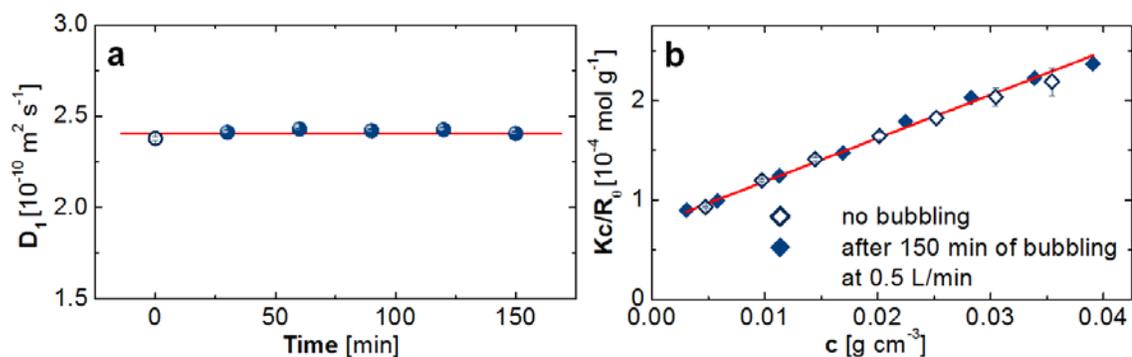


Figure 5-13. Characterization of intermolecular interaction in lysozyme solutions during bubbling test.

Cluster characterization we show in Fig. 5-14 and 5-15. From Fig. 5-14 we notice first, that a created effect on protein-rich clusters does not depend on the nature of the bubbled gas. Clusters characteristics, such as radius R_2 , volume fraction φ_2 , and number density n_2 reveal the same dependences on bubbled time for both, nitrogen and oxygen gases. We also performed tests with bubbling lysozyme solutions with air and helium, obtained dependences reveal the similar trends, from which we conclude, that the chemical identity of the bubbled gas does not affect the resulting clusters characteristics. These results tell us that the proposed possibility of chemical bridges participation in clusters formation is not confirmed. In Fig. 5-14 we also tested the bubbling effect on its preservation. We characterized lysozyme clusters on the second day after bubbling with O_2 and notice that cluster properties remain intact.

We show the evolution of cluster radius R_2 in Fig. 5-15 with increase of the bubbling time. The reference solution with no bubbling was analyzed to verify the steadiness of cluster radius throughout the experiment. We noticed the increase of R_2

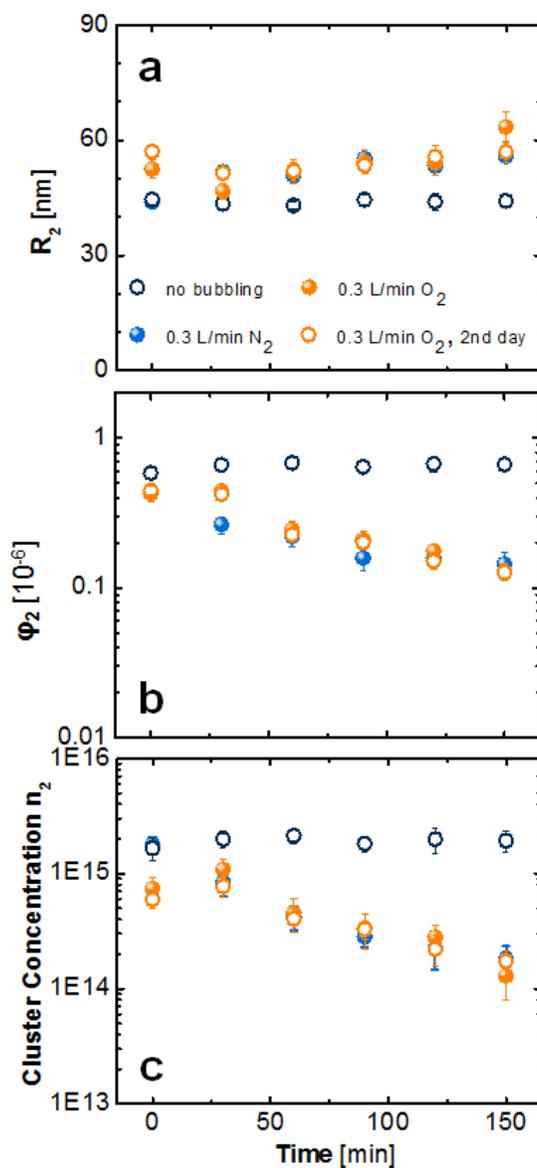


Figure 5-14 Evolution of cluster characteristics during bubbling of N₂ and, O₂, at 0.3 L/min flow rate and preservation of them within 24 hours. **a**, cluster radius R_2 , **b**, volume fraction ϕ_2 , **c**, concentration n_2 .

with increase of the gas flow rate from 0 to 0.3 L min⁻¹ and increase of the bubbling time. This observation corresponds to a proposed cluster formation mechanism due to protein complex formation. Purging of the samples at moderate rates leads to protein unfolding, stabilization of protein oligomers, and increase of cluster radius. The

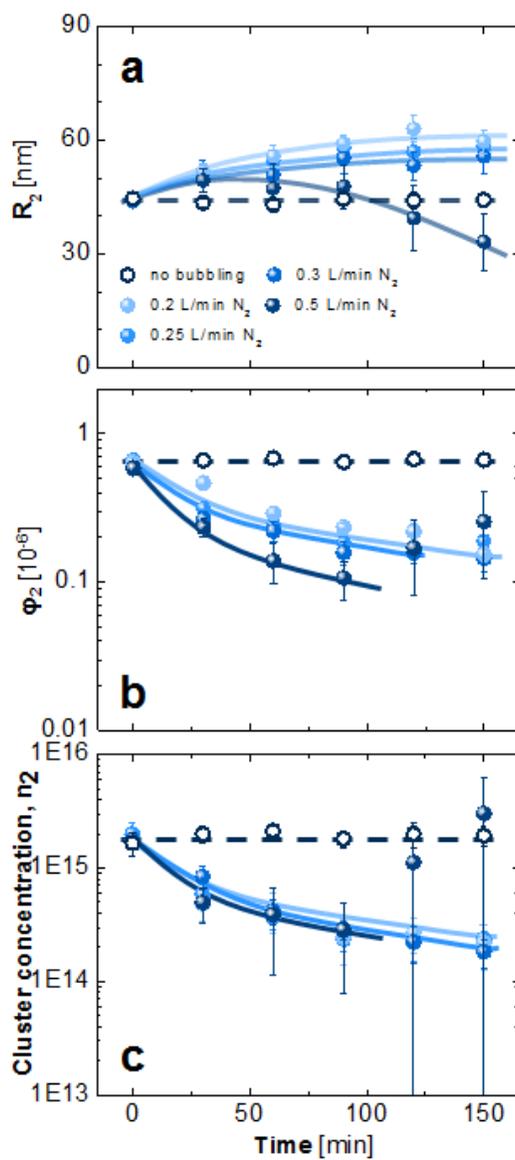


Figure 5-15. The effects of shear flow on cluster formation. The response of the cluster radius R_2 to solution bubbling with N_2 .

decrease of cluster radius at the highest probed flow rate 0.5 L min^{-1} can result from the increase of shear stress rates in solutions which heads to protein denaturation and irreversible aggregation.

5-7. Conclusions

The present 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 Coulomb independence of the clusters size distinguishes them sharply from the two other classes of clusters observed in protein solutions. The mesoscopic clusters exhibit 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. Tests with urea and ethanol and purging lysozyme solutions with gases suggested the importance of hydrophobic and water-structuring interactions in clustering behavior, supporting the hypothesis of domain swapping mechanism in their formation.

CHAPTER 6

ALTERNATIVE MECHANISMS OF CLUSTER FORMATION ON THE EXAMPLE OF PROTEIN GLUCOSE ISOMERASE

Xylose isomerase (glucose isomerase) is a tetrameric enzyme participating in conversion of a wide range of monosaccharides, such as transformation of D-glucose to D-fructose. Each segment of this protein is a 388 amino acid sequence with a total molecular weight of the tetramer 173 kDa, Fig. 6-1. Isoelectric point of glucose isomerize varies depending on a host organism producing the enzyme and typically ranges from pH 3 to 5 (from Hampton Research guide and^{133,134}). The secondary structure of glucose isomerase is related to a TIM-barrel, when eight α -helices and eight parallel β -sheets alternate along the peptide backbone and fold together forming a barrel-like structure with an active site inside. Two Mg^{2+} ions are located inside TIM-barrel and are important in catalytic activity of this enzyme. The major industrial application of glucose isomerase is production of D-fructose extensively used in food industry as a low calorie sweetener.

Moderate amount of work was done on studying glucose isomerase clusters and their participation in the two-step mechanism of nucleation. Nevertheless, there is no profound research on glucose isomerase clusters behavior and attempts to understand a mechanism of their formation. Our interest in glucose isomerase protein clusters hinges on the peculiar structure of this enzyme. While well studied lysozyme-rich clusters are built from single chain monomers, clusters of glucose isomerase originate from solutions of tetrameric molecule structure.

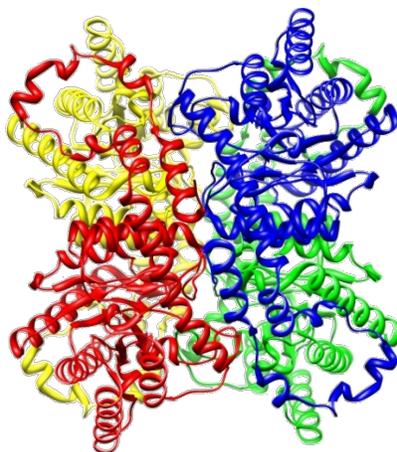


Figure 6-1. The structure of tetrameric glucose isomerase molecule drawn with Pymol. Different units are colored correspondingly.

We test the significance of water structuring interactions and partial protein unfolding on the glucose isomerase clusters formation as well as we induce the shear stress by bubbling of the enzyme solutions with N_2 . The particular choice of these experiments is in the opportunity to relate results to a previously conducted study on lysozyme and reveal similarities or differences which could explain the glucose isomerase cluster formation.

The mechanism of lysozyme clusters we see in formation of oligomers, or in a simple case, dimers. Glucose isomerase possess a native tetrameric structure, thus a different scenario of clustering behavior of this protein can be expected.

6-1. Materials and methods

Glucose isomerase from *Streptomyces rubiginosus* (Hampton Research, USA) was dialyzed against 100 mM HEPES buffer at pH 7.0. The concentration of the protein was determined by UV absorbance at 280 nm using an extinction coefficient $1.042 \text{ ml mg}^{-1} \text{ cm}^{-1}$. The final concentration was adjusted to 1 mg ml^{-1} (justified below) and the stock solutions were filtered through $0.2 \text{ }\mu\text{m}$ PVDF syringe filters (Fisher Scientific,

USA). All experiments were carried out at 22⁰C. Clusters characterization was performed employing Brownian microscopy technique described in detail in chapter 2.

Concentration of glucose isomerase was chosen following two considerations. It had to be low enough to not have a strong monomers background overlapping the signal from clusters, and high enough to produce clusters at a sufficient for Brownian microscopy concentration.

The high concentration of glucose isomerase yields a strong background observed by Brownian microscopy from its monomers due to their large size, 5 nm. This fact prevents the possible glucose isomerase clusters to be resolved with dynamic light scattering and dictates the requirement of keeping the protein concentration at a low level. This fact determined the principal experimental technique, Brownian microscopy since it can register the signal even at the very low scatterers concentration. We determined the final concentration for our tests to be 1 mg ml⁻¹.

To evaluate the effects of partial protein unfolding on glucose isomerase cluster formation, we tested solutions containing urea at 0.5 – 3M concentration. The highest concentration of urea used was chosen following the reasoning to not to cause irreversible unfolding and reveal any possible effect. Shear stress experiments performed by purging of N₂ through the protein solutions required higher clusters concentration in the sample, since we had to be accurate in cluster concentration determination. We tested several conditions for the presence of the high number of clusters at 1 mg ml⁻¹ glucose isomerase concentration. We tested protein dissolved in

- 1) H₂O (almost no clusters),

2) H₂O and 10% ethanol (many clusters with visually narrow distribution of sizes),

3) H₂O and 10% of acetic acid (decreases the pH from neutral to 4, yields many tiny scatterers that could be filtered out, which means that these are not the clusters but irreversible aggregates),

4) 100 mM Hepes pH 7.0 (the so-called standard solution we employed for clusters characterization),

5) 100 mM HEPES pH 7.0 and 10% ethanol (yielded many clusters with low polydispersity),

6) 100 mM HEPES pH 7.0 and 10% of acetic acid (yielded the same result as condition (3)),

7) 100 mM citric buffer pH 4.8 (produced many polydispersed clusters with fluctuating shape which is not suitable for the current experiments),

8) 100 mM citric buffer pH 4.8 and 10% ethanol (produced amorphous precipitate),

9) 100 mM citric buffer pH 4.8 and 10% acetic acid (produced amorphous precipitate as in conditions (8)).

From the tested conditions we selected 100 mM HEPES pH 7.0 and 10% ethanol for the bubbling experiment, since cluster volume fraction was $\sim 3.5\times$ higher in comparison to a standard, 100 mM HEPES pH 7.0, solvent, and clusters radius increased from ~ 50 to ~ 70 nm.

6-2. Pairwise monomers interactions and solution viscosity

Prior characterization of clustering behavior in glucose isomerase solutions we tested the monomers pairwise interactions by evaluating the second virial coefficient B_2 . We inspected three distinct conditions: 100 mM HEPES pH 7.0, 100 mM HEPES pH 7.0 and 10% ethanol, and 100 mM HEPES pH 7.0 and 3 M urea. We performed static light scattering methods as described in chapter 2 and from obtained Debye plots shown in Fig. 6-2 a calculated the second virial coefficient B_2 , displayed in Fig. 6-2 b.

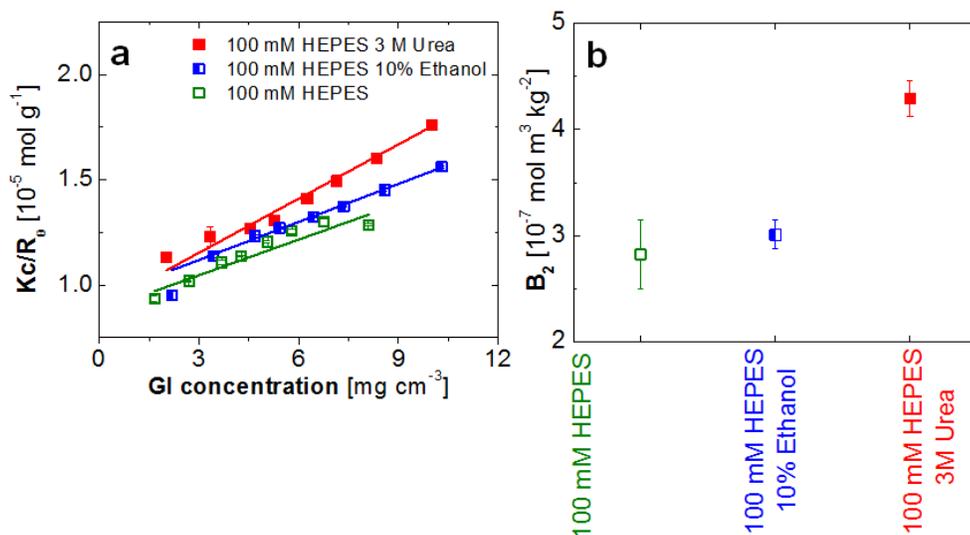


Figure 6-2. The pairwise interactions in glucose isomerase solutions. **a**, Debye plots, **b**, the second virial coefficient B_2 , obtained from **a**.

The values of the second virial coefficient are positive but less than the value of non-interacting hard spheres $B_2^{\text{HS}}=1.6 \text{ mol cm}^3 \text{ g}^{-2}$. Glucose isomerase proteins experience slight attraction, and this attraction attenuates following conditions: HEPES, HEPES and 10% ethanol, HEPES and 3M Urea, displayed in Fig. 6-2b.

Overall, the results on monomer-monomer interactions suggest that variation of solvent composition does not affect the behavior of glucose isomerase monomers much.

This fact allows us to eliminate a possible influence of monomers interactions on clusters formation and study the mechanisms of monomers aggregation into a new mesoscopic metastable phase separately from indirect effect of the solution osmotic compressibility.

We measured protein solutions viscosity in the presence of 10% ethanol and at all tested concentrations of urea. We obtained 1.22 cP of 1 mg ml⁻¹ protein solution viscosity in the presence of 10% ethanol and the dependence of the viscosity in urea containing solutions we show in Fig. 6-3.

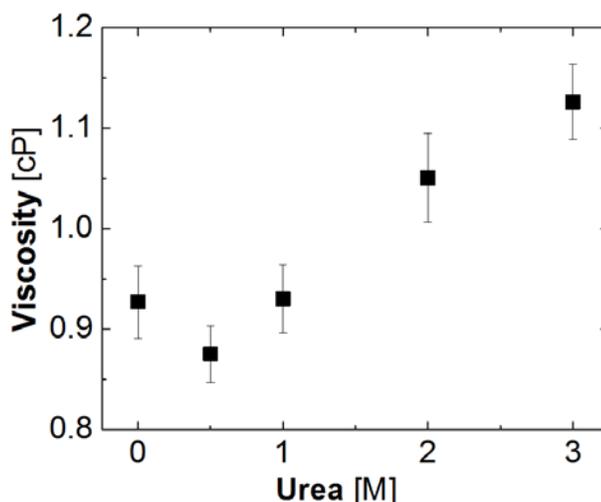


Figure 6-3. Viscosity of glucose isomerase solutions without and in the presence of urea.

6-3. The role of partial protein unfolding and shear-stress influence on cluster formation in glucose isomerase solutions

Following the experiments carried out with protein lysozyme, we tested the clustering behavior in glucose isomerase solutions with different concentrations of urea. The range of added urea concentration was from 0.5 to 3M, below the level when urea causes irreversible protein unfolding. We show the results in terms of clusters radius R_2 ,

volume fraction φ_2 , and concentration n_2 , Fig. 6-4 a-c. We show the measured values of these parameters after the solution preparation as well as after 1 day (24 hours) to study the preservation of possible effects on cluster formation. We see that in the presence of urea, cluster radius R_2 does not change and stays constant at 50 nm. The same trend is observed with cluster volume fraction, φ_2 , which remains unchanged at $\sim 3.5 \times 10^{-7}$, cluster concentration n_2 increases somewhat 20% from the standard value of the solution without urea to solution containing 3 M of urea.

With open symbols in Fig. 6-4 a-c we show the effect of urea on the second day after the solution preparation. We notice that the urea containing solutions, 0.5 and 1 M urea, show the slight increase in cluster radius which is in agreement with the results of the standard solution, where we observe Ostwald-like ripening. At higher urea concentration the cluster ripening seems to be suppressed.

Overall, we observe that in contrast to lysozyme solutions, cluster characteristics do not change significantly. This fact suggests that partial protein unfolding is not a part of clustering mechanism in glucose isomerase solutions.

The next step was to test the effects of shear stress on clusters behavior similar to the experiments conducted with lysozyme solutions. We chose N_2 gas (Praxair, USA) at the rate of 0.3 L min^{-1} . The results of the experiments we show as distributions obtained with Brownian microscopy in Fig. 6-5a-b and as values of clusters characteristics shown in Fig. 6-6 a-c calculated from the maximum values of distribution in Fig. 6-5 a.

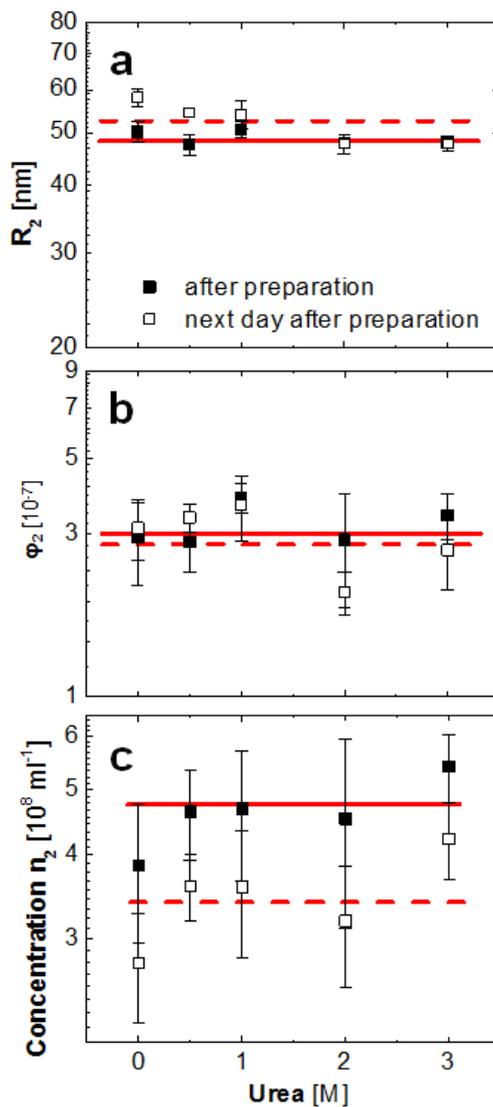


Figure 6-4. Clusters characteristics in glucose isomerase solutions in the presence of urea, **a**, clusters radius R_2 , **b**, clusters volume fraction ϕ_2 , **c**, clusters concentration n_2 .

We noticed that the cluster concentration decreases with bubbling time as follows from logarithmic dependences in Fig. 6-5 a-b. We see the decrease of cluster radius over 20% in the course of bubbling as well as cluster volume fraction and their concentration. These results suggest that shear stress induced by the bubbling of N_2 influence the cluster formation.

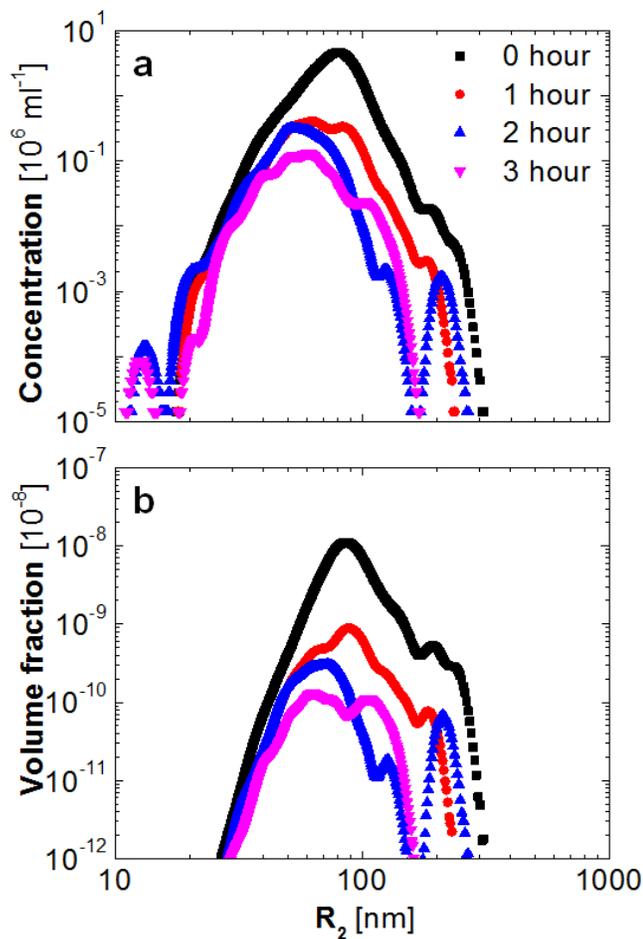


Figure 6-5. Distributions of glucose isomerase clusters population before and after bubbling with N_2 at 0.3 L min^{-1} in terms of concentration **a**, and volume fraction **b**.

6-4. Conclusions

It was proposed that clustering behavior requires the presence of new species in protein solutions. In case of lysozyme we suggest that partial protein unfolding leading to dimerization is the primary mechanism leading to cluster formation. This mechanism is supported by the results of urea and ethanol experiments discussed above. The absence of urea influence on glucose isomerase clusters suggests that partial protein unfolding is not at play in this protein and new species are formed due to some other mechanism.

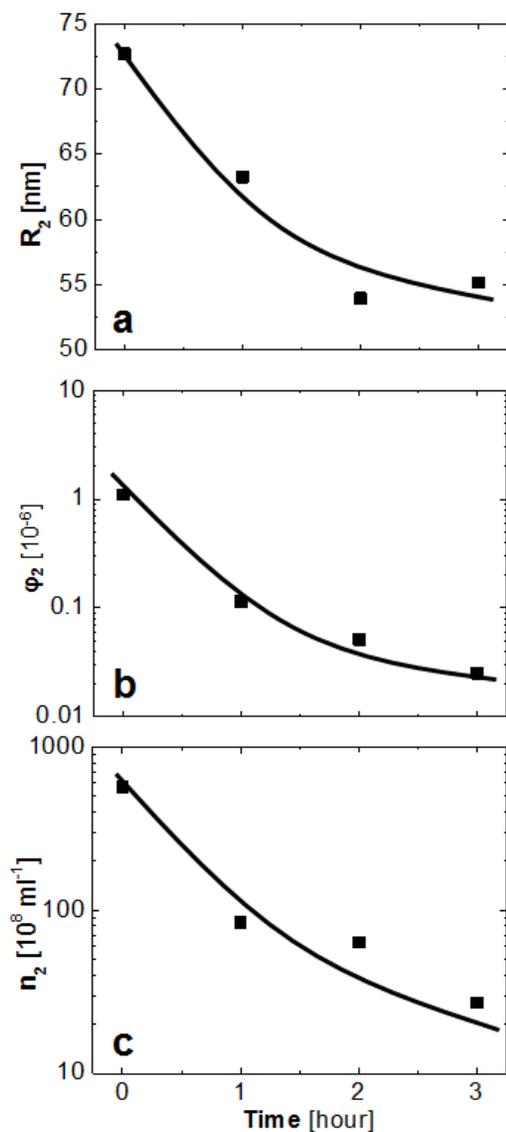


Figure 6-6. Evolution of glucose isomerase clusters characteristics after bubbling: **a**, cluster radius R_2 , **b**, cluster volume fraction φ_2 , **c**, cluster concentration n_2 .

Glucose isomerase monomers exists in tetrameric form. We propose a new mechanism of cluster formation when the new species are misassembled protein monomers, such as trimers and pentamers (higher disordered assemblies would result in higher free energy cost and less probability of formation). Our proposal is in agreement

with obtained results and supports the common clustering mechanism requiring the presence of new species and highlights the uniqueness of a certain protein by the mechanism of the formation of these new species.

CHAPTER 7

ORGANIC ADDITIVES IN CLUSTERS FORMATION

Proteins compose all living organisms and are involved in vast variety of living processes: signal transmission, molecular transport, catalytic reactions and others¹³⁵. This colossal involvement makes these molecules an interesting object of study. There are two major fields of protein aggregation research: first, there is a desire to know how to stabilize these complex molecules and prevent their aggregation, and, second, there is a need to find the conditions under which proteins can crystallize, i.e., to promote a desirable aggregation. Both problems are challenging and have different ways and approaches. The common pathway is to alternate system properties turning them to a desirable outcome. To change properties of the thermodynamic system “protein and solvent” one can start with modification either of the protein molecule or the solvent. The former way includes changing of protein structures – replacing some aminoacids to ensure greater molecule stability and stronger intermolecular interactions. The latter approach is oriented towards alternation of the solvent structure: water bonds, solvent surface tension, dielectric constant the solvent. Thus, modification solvent properties, such as pH or ionic strength, modifies electrostatic interactions, protein solvation, hydrophobic interactions, preferential solvent-protein interactions¹³⁶.

Due to complex protein structure there is no unique approach in protein manipulation. The successful recipes for both, protein stabilization and crystallization, are empirical. In the best case scenario, we have the experimental protocol which allows to achieve the desired result and we can explain why in this specific case the certain protocol works. A tremendous amount of work was done in attempts to explain the

successful use of some crystallization additives in protein science. Some organic substances act as salting-out agents and reduce protein solubility¹³⁶⁻¹³⁸. Alcohol addition to an aqueous solution also increases the second virial coefficient of a protein which means higher intermolecular repulsion¹³⁹. It was shown that in the presence of ethanol lysozyme acquires a more tightly folded conformation¹⁴⁰. Alcohols are capable of changing a dielectric constant of the protein, are responsible for enhanced electrostatic interactions, they break water structure and dehydrate protein^{141,142}. For example, in RNase-A crystallization experiments in the presence of ethanol attractive intermolecular interactions increased¹⁴¹. Authors suggest that this effect might be attributed to the change of the solvent dielectric constant. The enthalpy of lysozyme thermal transition in aqueous solutions changes in the presence of ethanol¹⁴³. Water-miscible organic solvents, such as ethanol and acetone, are generally considered to be good protein precipitants^{144,145}. Both lower dielectric constants of the solution and reduce the solvation power of solvents. A study revealed that protein lysozyme does not distinguish effects from ethanol and acetone as cosolvents and the solvation effect from these components have a non-specific character¹⁴⁶.

Substantial amount of work was done towards understanding the importance of MPD (2-methyl-2,4-pentanediol, hexylene glycol) as a cosolvent in protein behavior control. MPD is a very strong precipitant but not a strong denaturant, it does not denature proteins at crystallization conditions^{136,147}. Addition of MPD promotes strong protein hydration, not disturbing secondary and tertiary structure of the molecule, therefore the self-aggregation is enhanced¹⁴⁷⁻¹⁵⁰. It was revealed that MPD molecules tend to make penetrative contacts with protein molecules¹⁴⁷ stabilizing intermolecular interactions and

helping in protein crystallization. It was found that heptameric structure in the crystal was induced by MPD in α -hemolysin, and it was proposed that MPD can bind at the lipid head binding pocket and facilitates oligomer formation¹⁵¹. In¹⁴⁸ it was suggested that alkanediol may act similarly to surfactants in crystallization of membrane proteins, where solubilization reduces precipitation by aggregation and allows the controlled association of protein molecules into a protein crystal. MPD demonstrates a preference for hydrophobic residues¹⁴⁷⁻¹⁵⁰. In the example of RNase-A crystallization, it affects crystallization through preferential hydration of the protein and at high MPD concentration the protein undergoes phase separation, at conditions close to pH 5.8 it is strongly excluded from protein and leads to crystallization^{136,148}. It was found that in the presence of MPD the structure of RNase-A in crystals is somewhat different from that in solution, or that only one or a few of all native conformations is selected by the crystal lattice¹⁵². Detailed research revealed that MPD molecules are incorporated in the crystal structures and have the specific sites of binding: Phe34, Trp63, Trp123 and C-subsite for a lysozyme molecule, and Trp179, Arg200 in α -hemolysin^{151,153,154}. Michaux et al. note¹⁵⁴ that C-subsite is also a binding site of alcohols. In crystallized lysozyme MPD is found on the surface between two lysozyme molecules, close to Arg 114. The specificity suggests the importance of protein-MPD interactions essential in cooperative behavior¹⁵¹.

Another common additive, glycerol, acts in ways similar to MPD. It does not change the dielectric constant of the solvent significantly but is able to penetrate inside the protein molecule, stabilizing their native structure and increasing protein solubility^{137,155,156}. Glycerol increases the solubility, decreases enthalpy of crystallization, promotes nucleation and can result in a better quality crystals¹⁵⁶. It is also

noted that interfacial energies are reduced in the presence of glycerol and MPD^{156,157}. And for both, glycerol and MPD, preferential interactions can explain their effects on the self-association of proteins and their stability¹³⁶.

Crystallization is only one example of protein self-aggregation. Proteins can also form fibers, filaments, and protein-rich liquid clusters^{46,48,52,54,158}. The later example is encountered in solutions of many proteins^{48,50,52-54,65,75,159} and the protein clusters might possess different characteristics depending on their type¹⁶⁰. Vorontsova et al. give detailed description of the clusters type¹⁶⁰ and focus on studying those clusters, which participate in the protein two-step mechanism of aggregation^{49,53,54,161-163}. Following this mechanism, first dense protein-rich liquid clusters form in the solution and the nucleation event occurs inside these clusters^{46,48,52,54,155,158,164,165}. Thus, the properties of protein clusters populations are crucial for the properties of the crystal population.

Over 20 years ago George and Wilson introduced a concept of crystallization slot of B_2 , osmotic second virial coefficient¹⁶⁶. They propose that the optimal crystallization conditions correlate with a certain window of values for B_2 , which spans from -9 to $0 \times 10^4 \text{ mol m}^3 \text{ kg}^{-2}$. Thus, the possible success of crystallization experiments is expected to correlate with the decrease of intermolecular repulsion characterized by certain values of B_2 .

There are three main questions that we discuss in our study. How crystallization additives affect intermolecular interactions in protein solutions and whether variations in the second virial coefficient B_2 correlates with crystallizability. How these additives change thermodynamics of solutions, and how the additives affect crystal nuclei precursors, protein-rich liquid clusters? Understanding the clusters behavior in solutions

of different chemical compositions might help in answering the question why some crystallization experiments are successful while others are not.

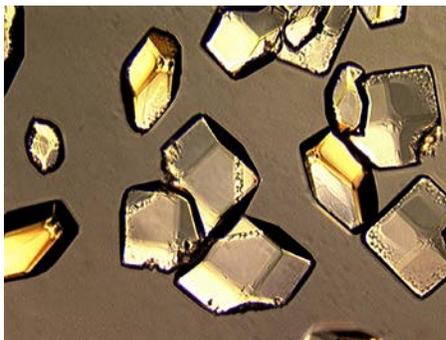


Figure 7-1. Lysosyme crystals grown in 20 mM HEPES pH 7.8 at 250 mg m⁻¹ concentration.

We chose protein lysozyme as our model system. It consists of a single chain and its tertiary structure encompasses two domains. This deviation originates from the vibrational motions of the molecule fragments⁷⁶, where they were distinguished as α - and β -domains linked by hinges regions. The most flexible portion of lysozyme molecule is its β -domain, while α -domain mostly consists of highly stable α -helices. Lysozyme behavior was widely studied - from crystallization under different conditions to formation of protein-rich clusters. Lysozyme crystals we show in Fig. 7-1, they are grown in 20 mM HEPES buffer pH 7.8 at 250 mg ml⁻¹ protein concentration. At these conditions crystallization process takes a prolonged period of time, around 1 – 2 weeks, and it allows us detailed characterization of protein solutions properties before the phase transition occurs. Structures of mentioned above organic additives are shown in Fig. 7-2 along with their chemical formulae. Structures are drawn with PYMOL using .pdb files

from <https://www.chem.purdue.edu/> for ethanol, acetone, and glycerol, and .sdf file from <http://webbook.nist.gov/> for MPD.

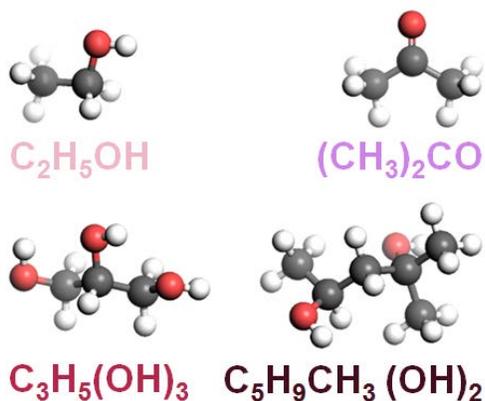


Figure 7-2. Ball and stick representation of organic additives used in this work and their chemical formulae. First row: ethanol, acetone; second row: glycerol, MPD (from the left to the right).

7-1. Materials and methods

Reagents and sample preparation We used chicken egg white lysozyme (Fisher Scientific) as a model protein, HEPES (Calbiochem), KOH (Mallinckrodt Chemicals) to adjust buffer pH, sodium azide (Fisher Scientific) to prevent bacterial growth. Organic solvents, ethanol, glycerol, MPD all were purchased from Sigma Aldrich, and acetone from VWR. To perform solutions viscosity measurements we used Optilink carboxylate-modified polystyrene particles with 0.424 μm diameter.

All tests were performed in 20 mM K-HEPES buffer (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonate), pH 7.8 with 0.01% NaN_3 . To determine protein concentration we employed Coulter DU 800 Spectrophotometer, using the

extinction coefficient of lysozyme $\varepsilon_{\lambda=280 \text{ nm}} = 2.64 \text{ ml mg}^{-1} \text{ cm}^{-1}$ ⁹⁸. Stock solutions of lysozyme were dialyzed over two days prior the experiments in order to remove small molecule salts acquired during an industrial preparation. After the dialysis the solutions were brought to a desired concentration of 20 mg ml^{-1} unless specified otherwise. Before data acquisition, solutions were filtered through PVDF $0.2 \mu\text{m}$ filters (Thermo Scientific) compatible with used organic solvents.

Data collection and processing. Clusters were characterized using dynamic light scattering (DLS) measurements performed on ALV scattering machine equipped with ALV-5000/EPP Multiple tau Digital Correlator and He-Ne laser, $\lambda = 632.8 \text{ nm}$, 35 mW (ALV-GmbH, Langen, Germany). Autocorrelation functions were collected at 90° for 60 s and analysis was performed as described in⁵⁴, applying independently measured solutions viscosity as in⁷⁵. The static light scattering (SLS) data was collected at 90° employing the same scattering equipment as used for DLS. The second osmotic virial coefficients of studied solutions were determined from Debye plots, using the refractive index increment of lysozyme solutions $dn/dc = 0.199 \text{ ml g}^{-1}$ ⁶⁶. All measurements were taken at 22°C .

7-2. Results and discussion

We started characterization of lysozyme solutions in the presence of organic additives by exploring monomer-monomer interactions. Protein monomers diffusivity D is a parameter which can be extracted from dynamic light scattering measurements and serve as a first indication of possible changes in intermolecular forces balance. The results of monomer diffusivity we show in Fig. 7-3, where we demonstrate the changes of diffusivity on concentration of crystallization additives. We account for the changes of

the buffer viscosities η_{buffer} , measured separately, by showing the dependence of the product $D\eta_{\text{buffer}}$. The common trend in shown dependences is the increase of monomers motility upon introducing cosolvents. This behavior can be interpreted by monomer stabilization and acquiring more compact conformation or by increase of monomer-monomer repulsion which also results in increase of monomers dynamics.

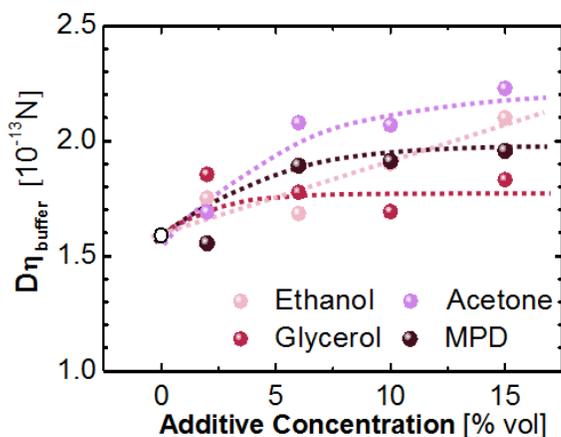


Figure 7-3. Dependence of lysozyme monomer diffusivity on the concentration of organic additives used as cosolvents. The lines shown here are guides for an eye.

To further investigate the changes in intermolecular behavior of the solutions we performed static light scattering measurements which gave us information on the effective osmotic second virial coefficient $B_{2\text{ eff}}$ and its changes in the presence of crystallization additives. The results we show in Fig. 7-4 a. We notice an increase of $B_{2\text{ eff}}$ for all tested cosolvents. Positive $B_{2\text{ eff}}$ and its further increase signifies the importance of repulsive interactions in studied solutions. Both parameters, $D\eta_{\text{buffer}}$ and $B_{2\text{ eff}}$, show that intermolecular interactions depend not only on the concentration of additives but on their chemical identity as well. The order in which crystallization additives modify

system interactions are preserved for both characteristics. Glycerol, MPD, ethanol, and acetone – is the sequence in which the effect of organic cosolvents appears to be pronounced. Different rates of the increase in $D\eta_{\text{buffer}}$ and $B_{2\text{ eff}}$ might originate from protein concentration differences at which these parameters were estimated. The diffusion coefficient was evaluated at 20 mg ml^{-1} protein concentration – conditions used to characterize protein-rich liquid clusters. Second virial coefficient $B_{2\text{ eff}}$ is the parameter mostly accurate for dilute solutions, at zero-limit concentrations.

Deviation from ideal behavior in protein solutions we also characterized by extracting the effective intercept $M_{w\text{ eff}}^{-1}$ from the Debye plots. We show dependence of $M_{w\text{ eff}}^{-1}$ on additive concentration in Fig. 7-4 b and we notice non-linear correlation between the percentage of organic component in the solutions and its influence on the effective molecular weight of the protein.

To further interpret the experimental data from static light scattering we employed Kirkwood and Goldberg theory on multi-component scattering in solutions¹⁶⁷ with at least one macromolecular component. We used the equation applied for solutions with all non-electrolyte components (equation 20 in¹⁶⁷). The refractive index ratio, α , we assumed independent of concentrations c_1 and c_2 , protein and low molecular weight organic solvent (acetone, ethanol, glycerol or MPD) respectively: $\alpha = \frac{dn/dc_1}{dn/dc_2}$, where from now on 1 stands for low molecular weight organic additive and 2 for protein molecules. After substitution of coefficients and simplification we obtained the following expression:

$$\frac{Kc_2}{R\theta} = \frac{1}{M_{2\text{ eff}}} + 2c_2B_{2\text{ eff}}, \quad (7-1)$$

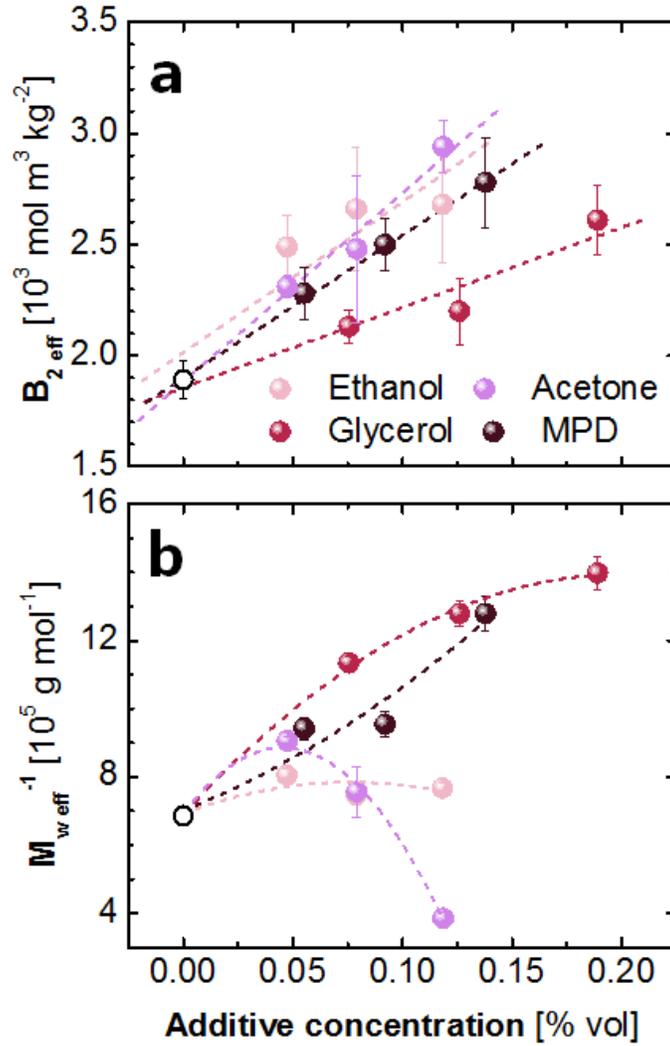


Figure 7-4. Static light scattering characterization of intermolecular interactions in protein solutions.

where from the Debye plots intercept is

$$\frac{1}{M_{2,eff}} = \frac{1}{M_2} + \frac{2\alpha A_{12}}{M_2} c_1 + \left(\frac{4\alpha B_{112}}{M_2} - \frac{2\alpha A_{11}A_{12}}{M_2} + \frac{3\alpha^2 A_{12}^2}{M_2} \right) c_1^2, \quad (7-2)$$

and the effective osmotic second virial coefficient is

$$B_{2,eff} = \frac{A_{22}}{2M_2} + \left[\left(1 + 2\frac{M_1}{M_2}\alpha \right) \frac{B_{212}}{M_2} - \frac{M_2}{M_1} \frac{A_{12}^2}{2M_2} + \frac{\alpha A_{12}A_{22}}{M_2} \right] c_2. \quad (7-3)$$

Coefficients used in these expressions have the following meaning:

$M_{1,2}$ – molecular weights of organic cosolvents and protein respectively,

A_{11} – characterizes pairwise interactions of organic molecules,

A_{12} – stands for interactions of pair organic and protein molecule,

A_{22} – the second virial coefficient for protein-protein interactions,

B_{212} – characterizes protein-protein interactions in the presence of organic molecule.

Since protein concentration in studied solutions was low, the virial coefficient of three-wise protein interactions B_{222} was accounted to be negligible.

Using refractive indices of organic solvents, $\frac{dn}{dc_1, \text{glycerol}} = 1.474 \text{ ml g}^{-1}$ ¹⁶⁸,

$\frac{dn}{dc_1, \text{MPD}} = 1.441 \text{ ml g}^{-1}$ ¹⁶⁹, $\frac{dn}{dc_1, \text{ethanol}} = 1.361 \text{ ml g}^{-1}$ ¹⁷⁰, $\frac{dn}{dc_1, \text{acetone}} = 1.359 \text{ ml g}^{-1}$

¹⁷¹ and corresponding densities, $\rho_{\text{glycerol}} = 1.26 \text{ g cm}^{-3}$, $\rho_{\text{MPD}} = 0.92 \text{ g cm}^{-3}$, $\rho_{\text{ethanol}} = 0.781 \text{ g cm}^{-3}$, $\rho_{\text{acetone}} = 0.794 \text{ g cm}^{-3}$, we fit $B_{2 \text{ eff}}$ as a linear function and $\frac{1}{M_{2 \text{ eff}}}$ as a

second order polynomial and from the coefficients of the fit we calculate the mentioned above virial coefficients.

We show computed coefficients in the Table 7-1, where we provide the values evaluated for solutions containing organic components as cosolvents as well as for system pure buffer-protein composition with no multiscattering effect. It is important to notice that the “real” molecular weight of the protein extracted using Kirkwood-Goldberg theory and the actual molecular weight of lysozyme agree with each other as well as the second virial osmotic coefficient gives identical values for the solutions containing organic additives as well as for pure buffer.

Table 7-1. The virial coefficients calculated using Kirkwood-Goldberg theory of scattering from multicomponent systems.

	$M_{w \text{ lys}}$ [g mol ⁻¹]	A_{12} [g cm ⁻³]	A_{11} [g cm ⁻³]	A_{22} [g cm ⁻³]	B_{212} [10 ³ cm ⁶ g ⁻²]
HEPES	14500	–	–	53.9	–
Glycerol	14500	9.1	13	53.7	6.2
MPD	14300	3.1	-4.3	54.2	0.6
Ethanol	14400	9.7	11.6	54.3	14.6
Acetone	14500	37.8	21.5	58.6	178.1

Similar to⁴⁷, ratio $\frac{Kc_2}{R\theta}$ can be related to chemical potential of the system and characterize thermodynamics of studied solutions. We obtain expressions for the chemical potential of lysozyme μ_2 in the absence of organic cosolvents and in the presence of used organic additives:

$$\frac{\mu_2 - \mu_2^0}{RT} = \ln c_2 + 2B_2 c_2 M_2 \quad \text{and} \quad (7-4)$$

$$\frac{\mu_2 - \mu_2^0}{RT} = \frac{M_2}{M_{2 \text{ eff}}} \ln c_2 + 2B_{2 \text{ eff}} M_2 c_2. \quad (7-5)$$

Here, B_2 and M_2 are from Debye plots from experiments with pure HEPES, $B_{2 \text{ eff}}$ and $M_{2 \text{ eff}}$ from Debye plots from solutions containing organic cosolvents, μ_2^0 is a standard chemical potential of the protein at concentration 1 M. The resulting chemical potential of the system is affected by the protein concentration as well as the chemical composition of samples.

We show the values of the solutions Gibbs free energy increase in Fig. 7-5. The dependences of the chemical potential μ_2 reveal different system responses to the

presence of organic cosolvents based on their identity. In Fig. 7-5 a ethanol up to 15% vol appears not to change solution thermodynamics despite the fact that it changes protein monomers motility and their interactions (Figs. 7-3 and 7-4 a). This behavior suggests a balance of counter effects of multicomponent system interactions, i.e., overall, the increase in $B_{2\text{ eff}}$ is compensated by $M_{2\text{ eff}}$ dependence on additive concentration. Acetone increases lysozyme chemical potential, Fig. 7-5 b, while glycerol and MPD tend to lower the excess of the free energy of studied solutions, Fig. 7-5 c-d.

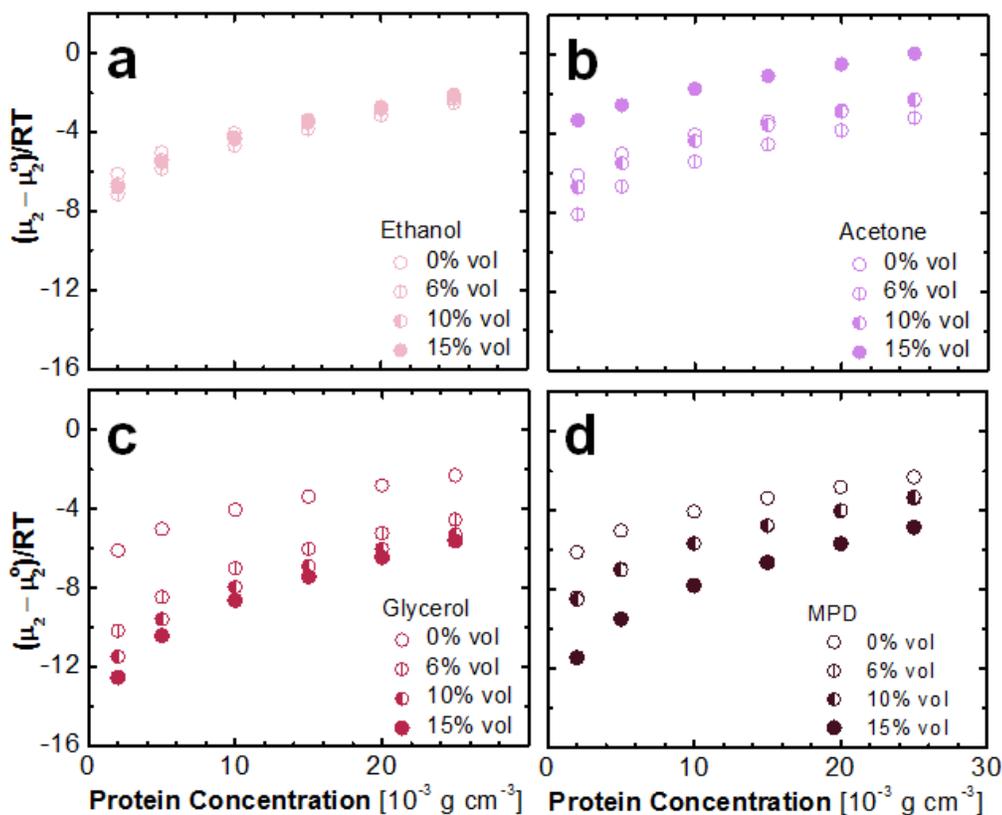


Figure 7-5. The increase of chemical potential μ_2 of lysozyme with its concentration in the presence of organic additives at different volume fractions: **a** – ethanol, **b** – acetone, **c** – glycerol, and **d** – MPD.

We show in Fig. 7-6 a the chemical potential of lysozyme in the presence of cosolvents at 20 mg ml^{-1} protein concentration, conditions used to characterize protein-rich clusters. Ethanol does not change solutions thermodynamics while acetone increases the chemical potential of lysozyme in the solutions and glycerol and MPD lowers it.

Typical crystallization experiments take from several hours to several days. To reveal possible effect of crystallization additives on protein-rich clusters we characterize clustering behavior in solutions 48 hours after samples preparation.

Addition of crystallization agent generally increases cluster radius R_2 , Fig. 7-6 b. The greater increase show solutions containing ethanol, two-fold. Glycerol does not show any increase from 2 to 10% of its concentration, at 15% of glycerol cluster radius even decreases.

Protein-rich liquid clusters are precursors for protein crystallization. It is still elusive whether the crystals nucleate inside clusters or on their surface. The first pathway assumes the total volume fraction ϕ_2 as an important factor for nucleation, the latter – their surface area, Σ_2 . Both parameters, ϕ_2 and Σ_2 , increase in the presence of crystallization agent (Fig. 7-6 b and c, respectively). The increase for different additives has different trend, signifying the dependence of clusters parameters on agent identity. Despite the fact that addition of glycerol leads to cluster radius decrease, the volume fraction and their surface area become greater upon its addition. The absence of correlation between solution thermodynamics and clusters behavior, as for example, with ethanol, when the protein chemical potential does not change and cluster parameters still increase, suggests the importance of protein concentration in evaluation of induced

effects on protein systems. The chemical potential μ_2 was calculated from the experimental data obtained from solutions of low protein concentrations, while the clustering behavior occurs at high protein concentration (concentrations close to concentrations in dense liquid phase).

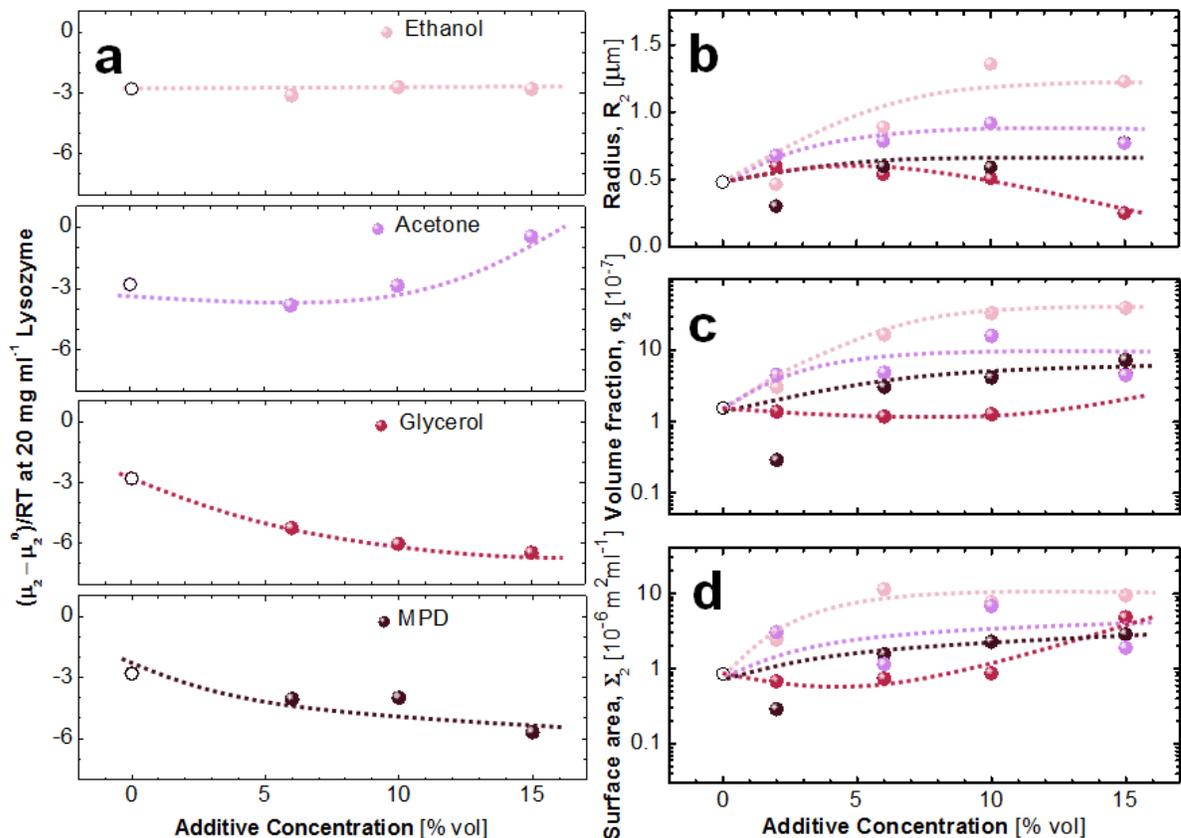


Figure 7-6. Lysozyme response to the presence of organic cosolvents.

Uncoupled behavior of cluster radius and volume fraction clearly noticeable on MPD example from fig. 7-6 b and c, where R_2 drops while φ_2 increases, highlights the difference in governing processes which determine both of these characteristics: cluster size is the kinetically determined parameter while cluster volume fraction is the property affected by solutions thermodynamics.

Overall we found, that clusters are affected by the presence of crystallization agents and are sensitive to the agent identity. Crystallization agents increase cluster fraction ϕ_2 and clusters surface area Σ_2 . Crystallization agents promote crystals growth in protein solutions through their effect on nucleation precursors – protein-rich liquid clusters.

7-3. Conclusions

In this study we investigated the effect of common organic crystallization additives on intermolecular interactions. We found that despite the concept of crystallization window, these additives tend to increase the second virial coefficient $B_{2\text{ eff}}$ manifesting higher intermolecular repulsion. We study the influence of organic cosolvents on protein-rich liquid clusters, protein nucleation precursors. SLS measurements were used to test the changes in solution thermodynamics, chemical potential of lysozyme, in studied conditions. Calculations revealed different behavior of the protein Gibbs free energy on additive identity and its concentration. Ethanol did not change the chemical potential of lysozyme compared to similar conditions in pure buffer while acetone increased it and glycerol and MPD lowered μ_2 .

We tested four common crystallization additives on clustering behavior in lysozyme solutions, ethanol, acetone, glycerol, MPD. We noticed that clusters size and their concentration depend on the additive identity. A general trend shows increase in clusters volume fraction and their total surface area. Here we conclude that crystallization additives successfully used in protein crystallization help the nucleation process through their effect on cluster population.

CHAPTER 8

TRACKING COMPLEX OBJECTS – PROTEIN CLUSTERS WITH TIME-DEPENDENT ASYMMETRIC INTENSITY PATTERNS

Diffusive dynamics constitutes an important part of processes of interest to fields ranging from biotechnology¹⁷²⁻¹⁸⁰ and cell biology¹⁸¹⁻¹⁸⁷ to fluid mechanics^{186,188} and colloid science^{189,190}. Understanding the role of diffusion in natural and engineered processes requires methods to quantify the motion of micron and submicron particles in complex media^{191,192}. Recent technological developments in time-lapse microscopy have greatly improved the imaging field, including characterization of diffusive motion. Nowadays, one can monitor the dynamics of single particles (spheres, living cells, protein complexes, viruses, etc.) with unprecedented detail and single particle tracking (SPT) has provided important insights on particle properties, their interactions with other particles and the environment, and the mechanisms that drive particle motion, and in this way have helped to understand numerous physical and biological processes^{173,184,185,187,193}.

Over the years several computer algorithms for particle tracking from a sequence of microscopic images have been developed¹⁹³⁻¹⁹⁵. On each image, particles are identified and the coordinates of their centers are determined, resulting in a time series of positions. General tracking methods deal with cases in which the intensity profile of a particle is radially symmetric. As such the center is allocated either to the point of maximum intensity, to the intensity centroid, or to the center of radial symmetry. More sophisticated techniques use a variety of fitting algorithms to provide sub-pixel resolution of the particle coordinates. As the peak of the point spread function of a diffraction-

limited spot can, in most cases, be approximated by a Gaussian function, many fitting algorithms use Gaussian profiles^{196,197}. A novel technique, radial centering^{198,199}, uses that the orientation of the gradient at any point on a radially symmetric intensity pattern is in the direction of the image center. This technique does not employ iterative nonlinear fitting and is faster than other centering methods.

Recently, systems have come into focus, which appear to challenge even advanced particle tracking algorithms. For instance, in images of micron-size liquid droplets the recorded intensity patterns are asymmetric and fluctuate with a characteristic time comparable to their diffusion time. Another examples come from particle tracking by dark field oblique illumination microscopy, a method that records the intensity scattered from submicron and micron-size particles^{27,28}, which is hampered by the asymmetry of the scattered intensity pattern. To quantify the diffusive motion of such particles and droplets, here we put forth a fast and easy to implement SPT algorithm. We employ a local spatial cross-correlation function to identify the displacement of a particle between two frames, similarly to an existing algorithm to track large solid particles²⁰⁰. We implement radial centering^{198,199} of the computed cross-correlation functions to evaluate the travelled distance with sub-pixel resolution. Importantly, the proposed algorithm does not rely on identification of the particle center in any single image, but reconstructs the particle trajectory from the displacements between pairs of images.

We demonstrate that this method enables an accurate characterization of dynamic behaviors of particle populations. We employ the proposed algorithms to obtain trajectories of submicron and micron-size particles and droplets of protein dense liquid. From the recorded trajectories, we evaluate the diffusion coefficient of spherical latex

particles of known size and compare its value to the Stokes-Einstein diffusivity. We use the correspondence between the two values as a quantitative indicator of the performance of the method. Applying this criterion, we demonstrate that the proposed cross-correlation method yields a more accurate estimate of the particle diffusivity than several common diffusion coefficient estimators.

8-1. Materials and methods

The spatial cross-correlation method of single particle tracking was tested using aqueous solutions of model polystyrene microspheres and gold nanorods. We used microspheres of three different diameters: 0.1 and 1.0 μm , supplied by OptiLinkTM, and 0.424 μm , from Seradyne. The particles had carboxylate-modified surfaces to provide surface charge that impedes aggregation. Gold nanorods were of 0.1 μm diameter and 1.0 μm length (NanopartzTM Inc., US). The concentrations of particles and nanorods were chosen so that their trajectories did not overlap within the longest data collection time, 100 seconds. The highest volume fraction was 10^{-8} , employed for the smallest 0.1 μm spheres to improve their visibility. These low concentrations minimized particle interactions, eliminated particle aggregation, and ensured that only self-diffusivity and convection contributed to particle motion.

Two protein solutions were tested in our work. Glucose isomerase (Microcrystal Oy, Helsinki, Finland) solutions was prepared at 90 mg ml^{-1} concentration in 100 mM Na-HEPES (N-2-Hydroxyethylpiperazine-N-2-ethanesulfonic, Fisher, US) buffer at pH = 7.0 containing 200mM MgCl_2 (Fisher, US). Lysozyme powder (Affymetrix, US) was dissolved in 20 mM HEPES buffer at pH 7.8 and dialyzed over two days. Tested solutions contained 20 mg/ml lysozyme and 15% v/v ethanol (Fisher, US). Both protein

solutions were aged for 1-2 weeks prior the experiments to ensure large clusters size in the range $1 - 2 \mu\text{m}^{111}$.

Data were collected employing Nanosight TM microscope setup, discussed in chapter 2. We used two cameras: sCMOS, operating at 25 frames s^{-1} , supplied by Nanosight, and Cooke Edge 4.2 with adjustable frame rate. The Cooke Edge 4.2 camera was used at 50 frames s^{-1} . Slower frame rates ($25, 10$ and 5 s^{-1}) were obtained from the original movies by removing a corresponding number of frames.

Centering and tracking algorithms. All tracking algorithms were implemented in MATLAB (version 2013a, MathWorks). No background correction was applied to the images. All images were cropped to the region containing the particle of interest.

1. The coordinates of the intensity centroid (x_c, y_c) were determined as

$$x_c = \frac{\sum_{i=1}^{i_{max}} \sum_{j=1}^{j_{max}} i I_{ij}}{\sum_{i=1}^{i_{max}} \sum_{j=1}^{j_{max}} I_{ij}} \quad (8-1)$$

and

$$y_c = \frac{\sum_{i=1}^{i_{max}} \sum_{j=1}^{j_{max}} j I_{ij}}{\sum_{i=1}^{i_{max}} \sum_{j=1}^{j_{max}} I_{ij}} , \quad (8-2)$$

where I_{ij} is the intensity of pixel (i,j) for an image of i_{max} by j_{max} pixels.

2. The Gaussian center was determined as the maximum (x_c, y_c) of the Gaussian profile

$$I_{ij} = A \exp\left(-\frac{(i-x_c)^2 + (j-y_c)^2}{2\sigma^2}\right) + B , \quad (8-3)$$

fitted to the image of the a particle of interest using a non-linear least-squares fitting procedure with five parameters, x_c, y_c, A, B , and σ .

3. The radial centering technique relies on the fact that in a radially-symmetric intensity pattern the intensity gradient at any point is directed towards the center of the

tracked particle. We followed the procedure by Parthasarathy^{198,199} and the MATLAB code published in the Supplementary Information of ref. ¹⁹⁸.

4. For the spatial cross correlation method we calculated a spatial correlation function of two images as the inverse two dimensional fast Fourier transform (2D-FFT) of the product of the 2D-FFT of the first image and the complex conjugate of the 2D-FFT of the second image. For convenience, the origin of the inverse 2D-FFT was shifted to the center of the image. The result was normalized with the product of the average intensity of both images. The center of the spatial cross correlation function was determined with sub-pixel resolution using the radial centering technique in the vicinity of the maximum.

Diffusion coefficient. The expected diffusion coefficient of the spherical latex particles was calculated using the Einstein–Stokes relation

$$D = \frac{k_B T}{6\pi\eta R}, \quad (8-4)$$

where k_B is the Boltzmann constant, $T = 295$ K is the temperature, and R is the particle radius. The viscosity of water η was taken as 0.957 mPa s at the temperature of the experiments. This estimate yields diffusivity $D = 4.51 \times 10^{-13} \text{ m}^2 \text{ s}^{-1}$.

8-2. Challenges of tracking micron-size objects

The current particle tracking methods may break down when the particle images vary in shape and intensity during monitoring. The variation may reflect deformations of liquid droplets and cells driven by Brownian collisions with the solvent molecules, and rotation of particles with anisotropic shapes or non-uniform optical properties. Asymmetric illumination may exaggerate the shape variations of particles with minor

deviations from sphericity that expose different sides to the beam as they undergo rotational diffusion.

As examples of time-dependent intensity patterns, here we use images of diffusing solid spheres and rods and liquid droplets obtained with oblique illumination dark-field microscopy, oftentimes referred to as Brownian microscopy or Nanosight technology^{101,111,201-204}. In this technique, solution samples are held in a thin cuvette under a microscope. The illuminating beam extends at an angle with the microscope optical axis, adjusted to avoid the microscope objective lens, Fig. 2-3a. Light scattered from particles in the field of view is captured by the objective lens and recorded by a video camera. The reliance on scattered light offers several important advantages. First, particles with refractive index close to that of the solution and with sizes smaller than the diffraction limit are detectable. Second, owing to stronger scattering, reflected in the Rayleigh law, larger particles produce stronger signal and can be detected on a background of smaller scatterers. Last, particles out of the image plane are detectable, leading to better population statistics. The scattered intensity pattern of all particles is significantly larger than their size.

To produce time-dependent asymmetric scattered intensity patterns with oblique illumination microscopy, we use spherical particles with near-micron diameters (these particles are too small for continuous monitoring with bright-field microscopy), which produce asymmetric intensity patterns because of the asymmetric illumination. The variability of the intensity patterns escalates for our additional test objects, non-spherical particles and liquid droplets. Figures 8-1 a and b illustrate the challenge of tracking liquid droplets of near micron size. As a model system, we use mesoscopic protein-rich

clusters that exist in solutions of numerous proteins^{66,117,205,206} and are liquid^{55,65,116}. Evidence for several systems suggests that they may serve as precursors to the nucleation of crystals and other solid aggregates^{115,207,208}. The protein liquid clusters may relate to the non-membrane bound compartments (nucleoli, centrosomes, Cajal bodies, etc.) in several organisms²⁰⁹⁻²¹². The clusters of the proteins lysozyme and glucose isomerase (GI) change shape in the course of their tracking, which leads to asymmetric and variable patterns of scattered intensity, Fig. 8-1 a, dark red corresponds to maximum intensity, blue, to minimum. Time after the start of the recording is indicated on each image. The centers of the clusters obtained with the radial centering technique are indicated. As a result, the radial centering technique, the most advanced of the currently available options, overestimates the shifts of the particle center, Fig. 8-1 a, and produces an exaggerated particle trajectory, Fig. 8-1 b.

To further test the performance of commonly used tracking algorithms (intensity maximum (IM), radial centering (R), intensity centroid (IC), and Gaussian centering (G), described in detail in the Materials and Methods section), we monitor the diffusive motion of four classes of submicron, micron-size, and non-spherical particles. We use spherical latex particle with diameters 0.1, 0.424 and 1 μm and gold nanorods with 1 μm length and 100 nm diameter. The selection of images in Fig. 8-2 a, in which we have indicated the particle centers identified by each of the four techniques (The particle center, identified as intensity maximum (pink), radial center (yellow), intensity centroid (green), and Gaussian center (blue) is indicated on each image. The time after the start of the recording is shown on the images), reveals that the IM, R, and G algorithms yield

consistent results for the submicron size particles: the center locations are identical and the particle trajectories, displayed in Fig. 8-2 b, are similar.

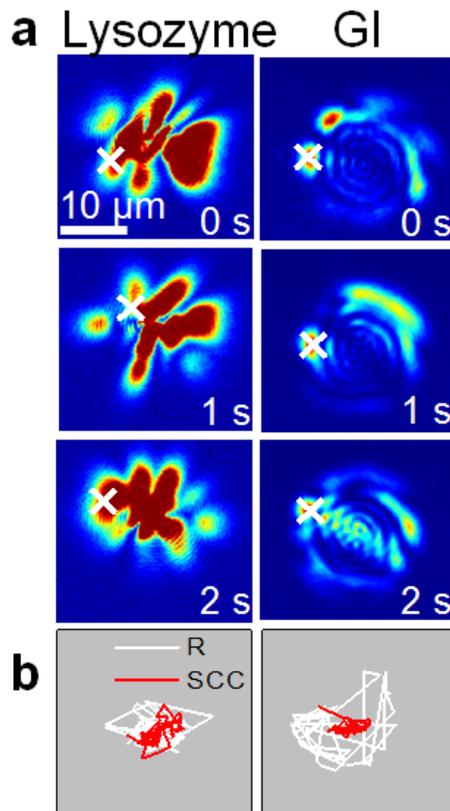


Figure 8-1. Imaging of liquid objects with oblique illumination microscopy. **a.** Images of protein-rich liquid clusters in a lysozyme solution and a glucose isomerase solution. **b.** The trajectories of the clusters imaged in **a**.

The IC method misjudges the center locations owing to the low-intensity of the scattered pattern. The accuracy of this method depends heavily on the background estimation and can be improved by background correction, but this increases the expanded computational time²⁰⁰. All four methods produce disparate center locations and trajectories for spheres of radius 0.424 and 1.0 μm and for 1 μm long nanorods. Inspection of the images in Fig. 8-2 a suggests that the errors of the tested methods are

rooted in the lack of radial symmetry of the intensity patterns, accompanied by fast changes in the location of the intensity maxima. The Gaussian centering method failed to converge for the intensity patterns of 0.424 and 1.0 μm particles, comprised of several concentric rings.

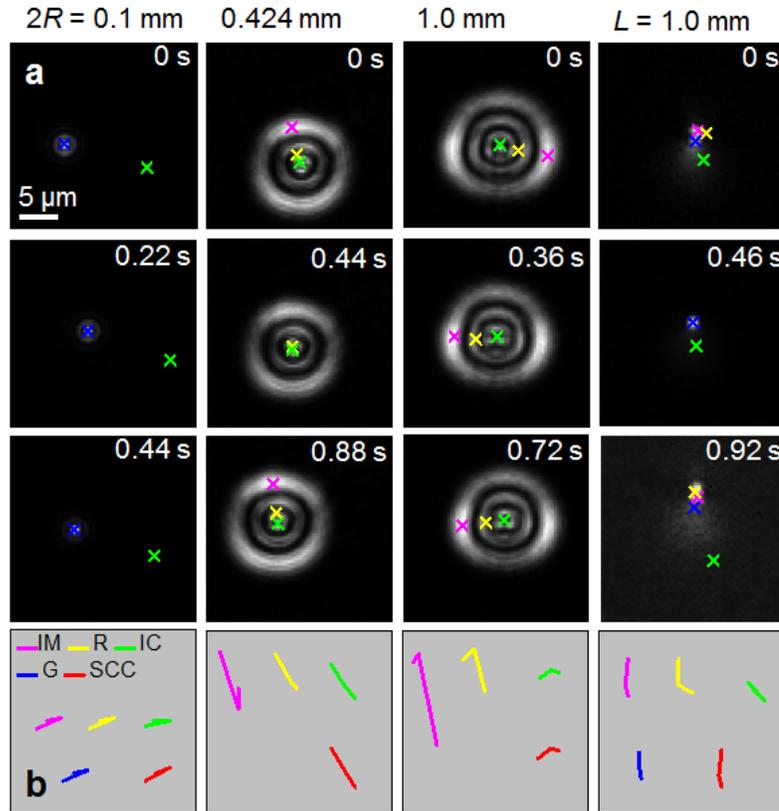


Figure 8-2. Tracking of latex spheres and gold nanorods by currently existing methods. **a.** Gray-scale images of the particles. **b.** Two-dimensional trajectories of the particles traced by the five methods.

8-3. The spatial cross-correlation (SCC) method

We propose to quantify the displacement of single particles using the spatial cross-correlation function of an image, taken at time t , with a reference image. Since the spatial cross-correlation has a single sharp quasi-centrosymmetric peak, the movements of the particles can be studied by tracking the intensity maximum of the spatial cross-

correlation function. The algorithm is illustrated in Fig. 8-3. To calculate the cross-correlation function of the images shown in Fig. 8-3 a, we first compute their spatial fast Fourier transforms (FFT); the respective amplitudes and phases are displayed in Fig. 8-3 b. We then compute the product of the FFT of an image with the complex conjugate of the reference image (the FFT of the reference image is multiplied with its own complex conjugate, resulting in a phase uniformly equal to zero) and the respective products are displayed in Fig. 8-3 c. Inverting the FFT of the products yields the cross-correlation function of the two images, or the autocorrelation function of the reference image, Fig. 8-3 d. The displacement of the intensity maximum of the cross-correlation function from that of the autocorrelation function of the reference image is taken as a measure of the displacement vector $\vec{\Delta r}$ of the tracked particle between the two images. In Fig. 8-3 e we superimpose the reference and the target images and position the tail of the displacement vector $\vec{\Delta r}$ at the apparent center of the reference image. As evidence of the reliability of the SCC method, the vector head points to the center of the target image.

To achieve subpixel resolution, we apply radial centering^{198,199} in the vicinity of cross-correlation maximum. Furthermore, in cases where the shape of the intensity profile changes drastically within a sequence of images, resulting in a decaying peak of the spatial cross-correlation function, a second reference image may be selected. Thus, if for the first group of k images image 1 is used as reference, for the next k images, we use image k as reference. To choose k , we note that large k 's bring down the signal-to-noise ratio since they correspond to a longer particle displacement. On the other hand, small values of k result in a higher and narrower peak of the spatial cross-correlation function, conducive of more accurate center localization.

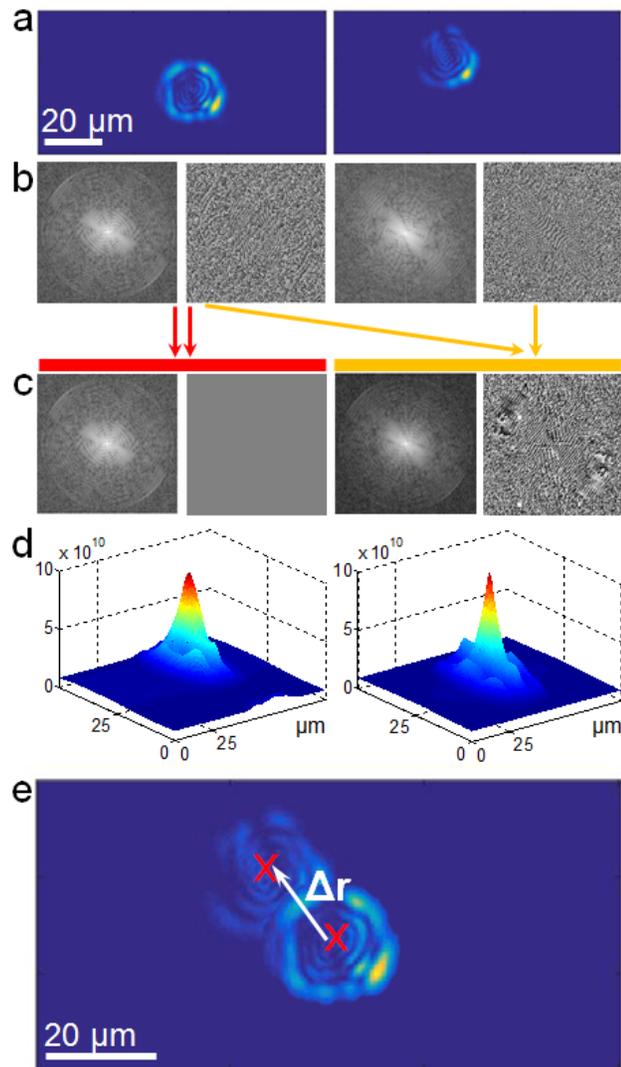


Figure 8-3. Illustration of the spatial cross-correlation (SCC) method of particle tracking.

In Fig. 8-1 b we compare the trajectories of liquid clusters of a variable shape resulting from the SCC method to those, produced by radial centering. In Fig. 8-2 b we plot the SCC trajectories of spherical particles of three sizes and nanorods. For these four classes of diffusing objects, the SCC trajectories are more compact. In view of the unrealistic shifts of the macroscopic clusters centers identified by radial centering in the

images in Fig. 8-1 a, and by the other three tracking methods in Fig. 8-2 a, we conclude the SCC trajectories are better fits to the actual particle motions. The proposed SCC method is robust, cross-correlation functions of overexposed images, such as the those in Fig. 8-1 a, possess sharp intensity peaks that allow accurate determination of the particle displacement.

8-4. The accuracy of the spatial cross-correlation algorithm

To test for the accuracy of the SCC algorithm, we employ it to evaluate the diffusion coefficient D of single spherical particles with known diameter, suspended in a solvent with known viscosity, and compare the resulting D with the theoretical diffusivity from the Stokes-Einstein equation employing the actual particle size. We monitor latex particles of 1 μm diameter freely diffusing in water and construct projections of their trajectories in the image plane. The Stokes-Einstein diffusivity of these particles is $4.51 \times 10^{-13} \text{ m}^2\text{s}^{-1}$. We quantify the mobility of a particle from its net displacement over n steps, squared to account for motion in both positive and negative directions, and averaged over all possible sets of consecutive n steps. This average is called mean squared displacement (MSD) and is denoted as $\overline{d_n^2}$. In current diffusivity evaluators^{45,48-51}, a time series of positions of a particle $\vec{r}_0(x_0, y_0), \vec{r}_1(x_1, y_1), \dots, \vec{r}_N(x_N, y_N)$ are used to evaluate particle displacements along the x and y coordinate axes, Δx_i and Δy_i . The SCC method yields the displacements with respect to the center of a reference image $\Delta \vec{r}_i(\Delta x_i, \Delta y_i)$. From $\Delta \vec{r}_i$, we compute particle displacements between the i -th and $(i + n)$ -th steps, $(\Delta \vec{r}_{n+i} - \Delta \vec{r}_i)$, and, from the displacement we obtain particle MSD, $\overline{d_n^2}$, as

$$\overline{d_n^2} = \frac{1}{N-n+1} \sum_{i=0}^{N-n} \{ (x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2 - \left\{ \frac{1}{N-n+1} \sum_{i=0}^{N-n} (x_{i+n} - x_i) \right\}^2 - \left\{ \frac{1}{N-n+1} \sum_{i=0}^{N-n} (y_{i+n} - y_i) \right\}^2 \}, \quad (8-5)$$

where $i = 0$ denotes the reference image. The two last terms on the right-hand side of the equation are the squares of the average displacements along the x and y axes, respectively, over n steps. If a particle performs a purely diffusive motion, these terms tend to zero for long times and can be omitted. However, in the cases where a particle is subject to non-stochastic motions, such as drift, the subtraction of the two terms ensures that only the stochastic part of the displacement is taken into account²¹³. The averaging in equation 1 and other relations below is unweighted since the recent evidence suggests that weighting does not improve the accuracy of the method²¹⁴.

We evaluate D as the slope of the correlation between $\overline{d_n^2}$ and the duration t_n of n steps. We compute $t_n = n\Delta t$, where Δt is the time between two consecutive images in the sequence, from which the displacement is computed. With this, diffusivity D is the slope of the relation

$$\overline{d_n^2} = 4Dt_n, \quad (8-6)$$

where the coefficient 4 accounts for the two dimensional motion of the recorded particle. Since diffusion in each of the three spatial dimensions is independent, a two-dimensional projection of the trajectory of a particle contains all information about the particle Brownian motions in an isotropic medium.

Besides evaluation of the accuracy of the SCC algorithm, we use data on the diffusive motion of known particles to address several outstanding questions related to the characterization of diffusive dynamics. These are 1) the utility of one-dimensional particle displacement data, 2) the significance of using non-overlapping n step sets for MSD determination instead of classical approach in which the steps are dependent, 3) the optimal duration over which particle displacement should be averaged, i.e., number

frames included in MSD calculation, 4) the optimal frequency of particle monitoring, i.e., frames per second of a captured movie, and 5) on the time duration of particle monitoring, i.e., length of a particle movie.

Oftentimes, projections of Brownian trajectories on a single axis, one dimensional motion, are used to determine the diffusivity from $\overline{d_n^2} = 2Dt_n$, where $\overline{d_n^2} = \overline{y_n^2}$ or $\overline{x_n^2}$. To address the validity of this data reduction scheme, we compute the mean squared displacements along the x and y axes, respectively, as

$$\overline{x_n^2} = \frac{1}{N-n+1} \sum_{i=0}^{N-n} \{(x_{i+n} - x_i)^2\} - \left\{ \frac{1}{N-n+1} \sum_{i=0}^{N-n} (x_{i+n} - x_i) \right\}^2 \quad (8-7)$$

and

$$\overline{y_n^2} = \frac{1}{N-n+1} \sum_{i=0}^{N-n} \{(y_{i+n} - y_i)^2\} - \left\{ \frac{1}{N-n+1} \sum_{i=0}^{N-n} (y_{i+n} - y_i) \right\}^2. \quad (8-8)$$

The displacements, averaged in equations 8-5 and 8-7, are correlated. For instance, $(x_2 - x_0)$ overlaps with $(x_3 - x_1)$, etc. MSDs based on independent non-overlapping displacements, i.e., $(x_n - x_0)$, $(x_{2n} - x_n)$, etc., can be computed as

$$\overline{d_{n,ind}^2} = \frac{1}{N_{n,ind}} \sum_{i=0}^{int(N/n)} \left\{ (x_{(i+1)n} - x_{in})^2 + (y_{(i+1)n} - y_{in})^2 \right\} - \left\{ \frac{1}{N_{n,ind}} \sum_{i=0}^{int(N/n)} (x_{(i+1)n} - x_{in}) \right\}^2 - \left\{ \frac{1}{N_{n,ind}} \sum_{i=1}^{N_{n,ind}} (y_{(i+1)n} - y_{in}) \right\}^2, \quad (8-9)$$

where $int(N/n)$ is the integer part of the ratio N/n , equal to the total number of independent displacements of length n . MSDs based on independent displacements along the x- and y-axes were computed using reductions of equation 8-8 similar to equation 8-7 above.

In Fig. 8-4 a we display the projection in the plane of the image of the trajectory of a latex particle with radius 1 μm . The particle coordinates with respect to the center of

the autocorrelation function of the first image were evaluated using the SCC method from 5000 images collected at constant $\Delta t = 20$ ms. The particle trajectory is highly asymmetric, suggesting the presence of strong drift roughly parallel to the x-axis. The MSDs $\overline{d_n^2}$ over $t_n = n\Delta t$ from 0 to 100 s, computed using equation 8-5, are displayed in Fig. 8-4 b. The MSDs increase monotonically with t_n for $t_n < 48$ s and then decrease and reach values close to zero. We attribute this decreasing branch to overestimation of the contribution of drift with variable velocity in the particle motions in equation 8-1, which assumes that the drift velocity is constant over the monitored period, and convection which does not have a constant rate. The increasing branch at $t_n < 48$ s is far from the straight line predicted by the Stokes-Einstein relation. The reason for the deviation is the limited statistics provided by data for a single particle. Averaging over 20 and more particles yields a nearly perfect linear correlation²¹⁵.

The particle diffusivities D , evaluated from the MSDs in Fig. 7-4 b, are displayed in Fig. 8-4 c. We determined each D value by fitting a segment of the $\overline{d_n^2}$ data to equation 8-6, ranging from the lowest $t_n = 0.02$ to a highest t_n varying from 0.04 to 20 s, and plotted the resulting D as a function of the highest t_n used in the fit. The upper limit of the highest t_n range, 20 s, was chosen somewhat arbitrary to evaluate the effects of longer t_n on the D determination. Evidence presented below suggests that the accuracy of D determinations decreases monotonically at $t_n > 5$ s, significantly shorter than the chosen limit, and no peculiarities occur near this limit. In general, the D values in Fig. 7-4 c deviate from the Stokes-Einstein value. At lag times $t_n < 1$ s, the deviation is $< 10\%$. The best correspondence is achieved at t_n near 1 s. At lag times longer than this, the

deviation increases and reaches up to 30%. At lag times > 10 s, the deviation decreases and reaches below 10% at $t_n = 18$ s.

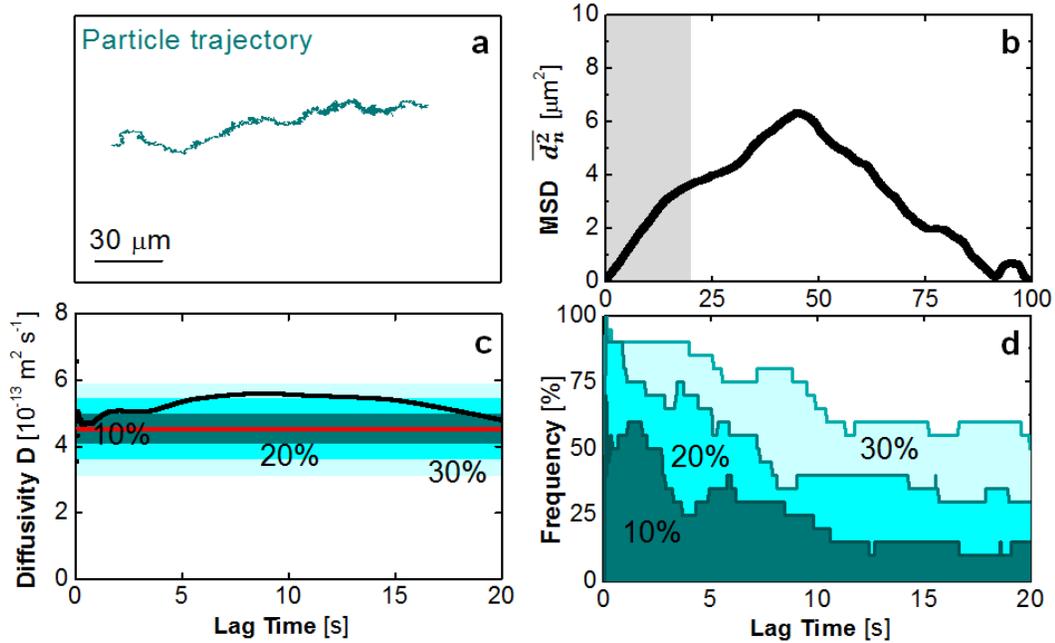


Figure 8-4. Evaluation of the accuracy of the SCC method. **a**, particle trajectory, **b**, MSD computed from **a**, **c**, diffusivity calculated for each lag time from **b**, **d**, distribution of particles diffusivities calculated as in **c**.

To evaluate the statistical significance of the observations on the accuracy of the diffusivity determination at different lag time ranges, observed in Fig. 8-4 c, we carried out identical determinations with total of 20 particles. We divided the diffusivities determined in each t_n range in four groups, those that are within, respectively, 10, 20, and 30% of the Stokes-Einstein value, and those that deviate by more than 30% from it. Clearly, the diffusivities in the first group are a subset of those in the second, and the latter are a subset of the 30% group. In Figure 8-4 d we plot as a function of the upper limit of the t_n range the number of diffusivities in each group scaled with the total number

of tested particles. This ratio is equal to the relative frequency of the respective deviations from the Stokes-Einstein value.

The data in Fig. 8-4 d reveal that for lag times shorter than 3 s about half of the tested particles yielded diffusivities within 10%, and about 85% within 30% of the Stokes-Einstein value. These data allow evaluation of the quality of the procedures for determination of D from the relative frequency of smaller deviations from the Stokes-Einstein value. Thus, D s determined over shorter lag time ranges are more accurate. This may appear counterintuitive since fewer data points seem to produce a better result. In fact, even the shortest $\overline{d_n^2}$ represents an average over the entire particle trajectory. The apparent reason for the higher accuracy of D determined over short lag times is the accumulation of inaccuracies in the localization algorithm, compounded by the imperfect account for drift with time-dependent velocity. A recent theoretical analysis²¹⁶ demonstrated that the ratio

$$\overline{d_1^2}/(4\Delta t) \quad (8-10)$$

may be closer to the exact value of value of D , and its distribution has a lower variance than the least-squares fit used to find the slope of the $d_n^2(t_n)$ correlation. The ratio in equation 8-9 corresponds to first point in Fig. 8-4 c. Determination of D from experimental data may not yield maximum accuracy at the shortest lag time since the shortest displacement is most sensitive to localization error intrinsic to the chosen particle tracking method. The correspondence of the data in Fig. 8-4 d to the theoretical prediction²¹⁶ is a testament to the high accuracy of determination of the trajectories of the individual particles by the spatial cross-decorrelation method.

8-5. Comparison with other particle tracking methods

We compare the performance of the SCC method to that of the common intensity maximum and the recent radial centering^{198,199} techniques of particle tracking. For this, we process with the three methods sequences of 5000 images for 20 particles of diameter 1 μm taken at a rate 50 frames s⁻¹. From the resulting trajectories we compute the MSDs and evaluate the diffusivities for different ranges of the lag time t_n , as illustrated in Fig. 8-4. For each lag time range and each method, we compare the obtained diffusivities with each particle and group the deviations from the Stokes-Einstein value in three groups: below 10, 20, or 30% and above 30%. We plot the frequency of individual particle diffusivity in these ranges as a function of the highest lag time in each determination in Fig. 8-5.

The first three columns in Fig. 8-5 demonstrate that a greater number of particles yield diffusivities closer to the Stokes-Einstein value with the SCC method than with the intensity maximum or radial centering methods. As expected, the radial centering technique produces more accurate D estimates than the intensity maximum method. In addition, the two older techniques exhibit frequency maxima at lag times of several seconds. These maxima contradict the predictions of the theoretical analysis²¹⁶ of diffusivity determination from particle trajectories and the contradiction suggests inaccurate identification of the particle coordinates, similar to the sequence illustrated in Fig. 8-2. Closer inspection of the SCC data at short lag times in the semi-logarithmic plots in Fig. 8-6 a reveals that diffusivities evaluated from MSDs correlations with lag times of about 0.1 s are more accurate than those with the shorter tested lag times, 0.02 s,

0.04 s, etc. This slight discord with the theoretical prediction is likely due to the finite accuracy of particle tracking with the SCC method.

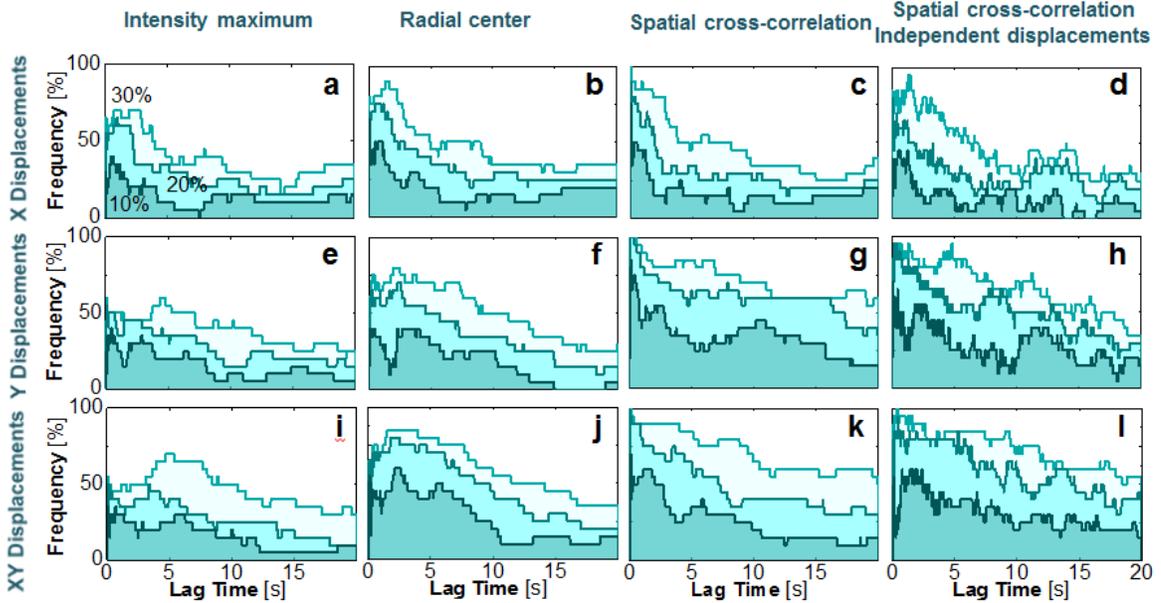


Figure 8-5. Comparisons of particle tracking and MSD evaluation methods.

The data in the rightmost column use MSDs computed from independent displacements using equation 8-8 and SCC-produced trajectories, to evaluate the diffusivity of individual particles. The frequency dependencies on the lag time are noisier, reflecting noisier $d_n^2(t_n)$ correlations that are likely due to significantly reduced statistics^{48,52}. The numbers of particles yielding diffusivities closer to the Stokes-Einstein value with this method is comparable to the numbers of particle processed with the “classical” SCC method with overlapping displacements. This similarity suggests that the use of independent displacements does not produce a significant advantage.

The diffusivities reflected in the statistics in Fig. 8-5 were computed from two-dimensional displacements in the image plane (bottom row) and from displacements exclusively along the x - or y -axes (top and middle row, respectively). The quality of the diffusivity data evaluated from trajectories identified with the intensity maximum technique do not appear to display significant differences between the three methods to compute displacements. The radial centering technique produces highest data quality with two-dimensional displacements. The SCC method allows finer distinctions. While the diffusivities evaluated from the two-dimensional data are more accurate (the initial values of the 10, 20, and 30% curves are higher, and the areas under them are greater) than those evaluated from x -data, the diffusivities evaluated from the displacements along the y -axis are even more accurate (e.g., 75% of the particles yield diffusivities within 10% of the Stokes-Einstein value). Examination of the trajectories of the 20 particles revealed that they are similar to the one in Fig. 8-4 a and exhibit significant drift nearly parallel to the x -direction. The inability to accurately account for this drift biases the displacements in the x -direction, and by inclusion, the two-dimensional displacements. The displacements in the y -direction are relatively free from such bias.

8-6. The frame rate and the length of the movie

The length of a time series of images of a particle may be limited owing to factors such as limited fluorescence time or departure from the field of view. Shorter image sequences may lead to a poor statistics and reduced accuracy of the determination. The data in Fig. 8-6 a characterizes the quality of the diffusivity evaluated with an increasing number of frames from 50 to 5000 recorded at a rate of 50 frames s^{-1} . The data demonstrates that sequences of 1000 images lead to more accurate diffusivity

determinations than those of 50, 100, 250 or 500 images. Extending the sequence length to 2500 and 5000 images does not improve the determination accuracy.

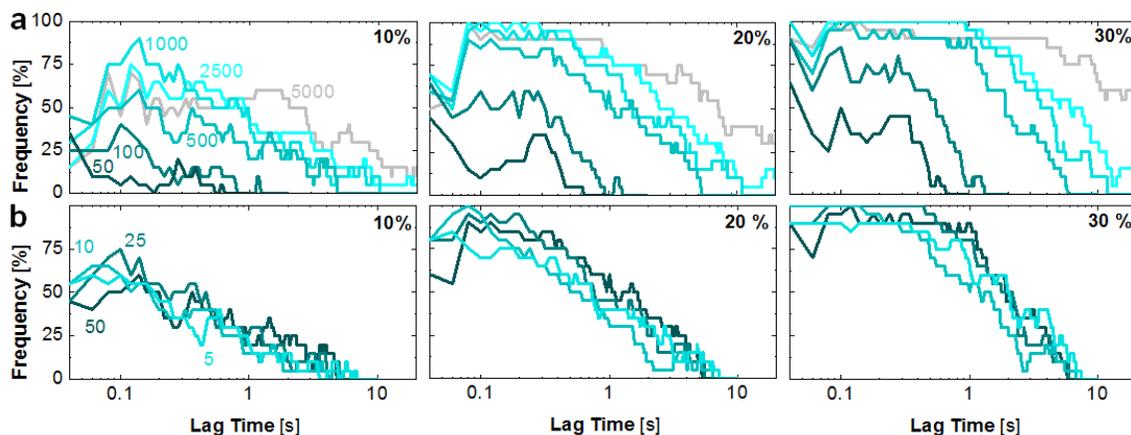


Figure 8-6. The effects of movie length and frame rate.

Slower frame rates ensure longer displacements between two successive images resulting in a high signal to noise ratio of the particle trajectory and should induce more accurate diffusivity data. On the other hand, with a fixed image sequence length, slower frame rates extend the duration of data acquisition and enhance the effects of drift, vibrations, temperature variations and other destructive factors. To evaluate the effects of frame rates on the quality of the diffusivity data, we compare the accuracy of diffusivity determinations from image sequences collected with frame rates varying from 5 to 50 frames s^{-1} . The data in Fig. 8-6 b demonstrates that the frame rate does not affect the accuracy of diffusivity determinations. We conclude that the better signal to noise ratio of slower frame rates is balanced by the perturbations enabled by longer data collection.

8-7. Conclusions

We have developed a new method to track the diffusive motions of single particles producing intensity patterns that vary over a timescale comparable to that of

diffusion. The method constructs particle trajectories in the image plane from the displacements of the peak of the cross-correlation function of an image with respect to a reference image. To increase the accuracy of the displacement determinations, the peak coordinates are identified using a published radial centering algorithm.

We demonstrate the utility of the method for tracking liquid droplets with changing shapes and micron-size particles producing images with exaggerated asymmetry. We evaluate the accuracy of the method by comparing the diffusivity of particles of known size determined by this method to the value predicted by the Stokes-Einstein relation. We compare the results with the intensity maximum and radial centering methods. We show that the diffusivity evaluations using trajectories determined with the spatial-cross-correlation (SCC) method are significantly closer to the expected value than those using the other two methods.

We address several open questions on the characterization of diffusive behaviors. We show that in the presence of drift, one dimensional trajectories in direction perpendicular to the dominant convective flow yield more accurate diffusivity values. We show that MSDs determined from non-overlapping displacements do not yield more accurate diffusivities than classical methods employing overlapping displacements. We find that more accurate diffusivity determinations result from mean squared displacement (MSD) for lag times up to about 0.1 s.

The sequences of 1000 images lead to more accurate diffusivity determinations than those of 50, 100, 250 or 500 images. Extending the sequence length to 2500 and 5000 images does not improve the determination accuracy. We show that with constant movie length, the frame rate does not affect the accuracy of diffusivity determinations.

We envision applicability of the SCC tracking method to all classes of objects with variable image shapes, cells, liquid droplets, particles of anisotropic shapes or optical density, particles with non-uniform fluorescent labeling, and others.

CHAPTER 9

THE LIQUID NATURE OF PROTEIN-RICH CLUSTERS

To address the question whether the protein clusters are liquid or solid, we carefully observed the time dependence of their size evolution, evaluated from the lag time of dynamic light scattering (DLS) signal. In numerous data sets^{48,52,54} two features of the evolution of cluster population are revealed: 1) the clusters appear immediately after solution preparation and 2) their mean radius R_2 remains relatively unchanged within several hours of observation. Both features are incompatible with solid clusters. Under the tested conditions protein solids take times of the order of 1 h to nucleate and grow to several hundred nanometers, and their size is not constrained but increases to dimensions visible with the naked eye within commensurate times^{48,217,218}. With these observations we can conclude that the clusters corresponding to the slow shoulder of the DLS correlation function consist of dense liquid in which the molecules move with respect to one another.

The liquid nature of lumazine synthase clusters was revealed by three atomic force microscopy observations⁹⁷. 1) The clusters shrink in height as they rest on the crystal surface, 2) the layers originating from the clusters merge continuously with each other and with the underlying lattice, and 3) the velocity of the layers originating from a cluster is the same as the velocity of the other layers on the surface of the crystal. If these were crystalline clusters, the probability of them landing on (001) plane downwards and rotating to a perfect register with the underlying lattice would be negligible. Numerous examples of microcrystals landing on surfaces of growing

crystals out of register and being incorporated with major stacking defects have been reported^{219,220}. Disordered solid clusters would not shrink in size with the observation times, and the generation of new layers would be likely to be accompanied by the creation of a strained shell that would delay the spreading of the layers started by clusters¹⁶⁴.

In the previous chapter, with a novel method of oblique illumination scattering microscopy we developed a new tracking method of large clusters with asymmetric dynamic intensity pattern, based on a spatial cross-correlation. The same method of the spatial cross-correlation can be used also to demonstrate the liquid nature of protein clusters and show that the characteristic time of the spatial cross-correlation function maximum could only be explained by shape instability of the liquid clusters.

9-1. Materials and methods

We studied clusters of protein lysozyme (Affymetrix) solutions, dialyzed over two days in 20 mM HEPES buffer pH 7.8. Protein clusters were allowed to age to reach the macrosize when the intensity pattern observed with oblique illumination microscopy represented the irregularly shaped and fluctuating bright spot. To compare the intensity patterns of the clusters with the patterns from spherical objects, we monitored the behavior of 1 μm latex microspheres (OptiLink).

We employ oblique illumination microscopy (Nanosight LM10) to monitor the dynamics of individual clusters and microspheres. To ensure visibility of 100 nm fluorescent particles (Fisher Scientific) without interference with the scattered light from protein clusters we used a green-wave filter, supplied with Nanosight microscope.

9-2. Intensity patterns of protein clusters are indicators of their liquid nature

Observations of individual clusters revealed that intensity patterns observed with the oblique illumination microscopy of many clusters fluctuates, as shown in Fig. 9-1 b. Latex particles of a similar size do not exhibit the same behavior, Fig. 9-1a. Since particles with spherical symmetry do not show intensity fluctuations, we conclude that the protein clusters have a complex asymmetrical shape. The cluster intensity pattern is a sum of intensity patterns of its parts, which act as independent scatterers. The intensity fluctuations indicate that the clusters are characterized by a complex non-spherical shape which may also be fluctuating.

The method of the spatial cross-correlation, discussed in chapter 8, allows to build trajectories of big scatterers ($>\lambda/2$). With this, the movies of intensity patterns of particles, clusters or latex microspheres, can be recentered, so that translational diffusivity is eliminated. Afterwards, we can analyze the newly centered clusters, or microspheres, by the spatial cross-correlation method again. The center of the resulting function will have the same coordinates through the whole range of analyzed frames, though the height of the spatial cross-correlation function would have a certain behavior depending on the type of a scatterer.

In Fig. 9-2 we show the resulting height of SCC functions from a microsphere (Fig. 9-2 a) and a cluster (Fig. 9-2 b). We see that the maximum intensity of the SCC function peak stays constant for a 1 μm microsphere and shows a decay for a 1 μm cluster.

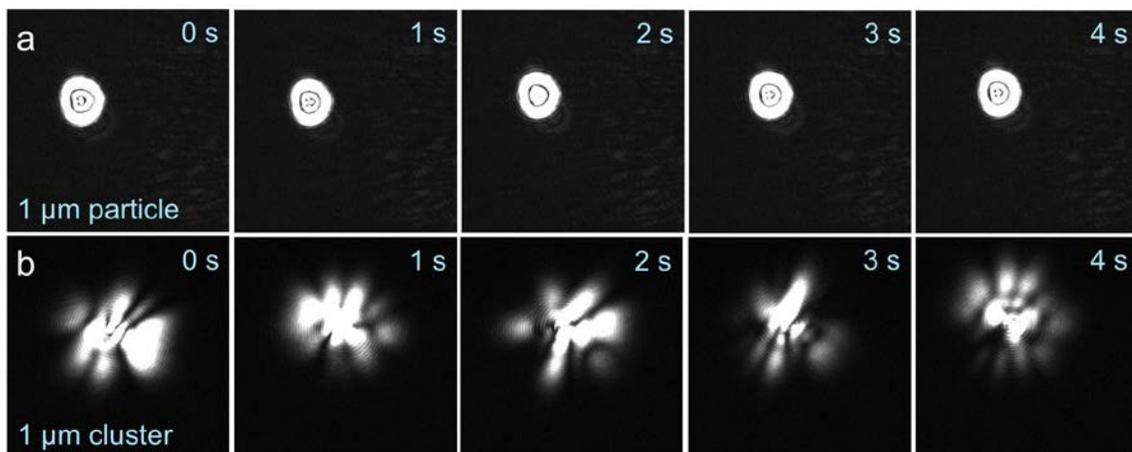


Figure 9-1. Intensity patterns of microspheres (a) and lysozyme clusters (b) observed by oblique illumination dark field microscopy. The intensity spots signify unique properties of protein clusters shape oscillating in time.

Thorough observation of intensity patterns produced by both type of scatterers explains the apparent difference in the maximum intensity behavior. The intensity pattern of a particle remains relatively constant and symmetrical throughout the whole number of frames, while the intensity pattern produced by the cluster is irregular and constantly changing. This observation suggests, that the height of the maximum intensity correlates with the intensity profile of the scattering object and decays, or stays constant, according to its transformation in time. Thus, the symmetrical solid microsphere does not show any decay in its spatial cross-correlation function, and the cluster, which could be either liquid or randomly shaped amorphous solid object produces an SCC maximum decay.

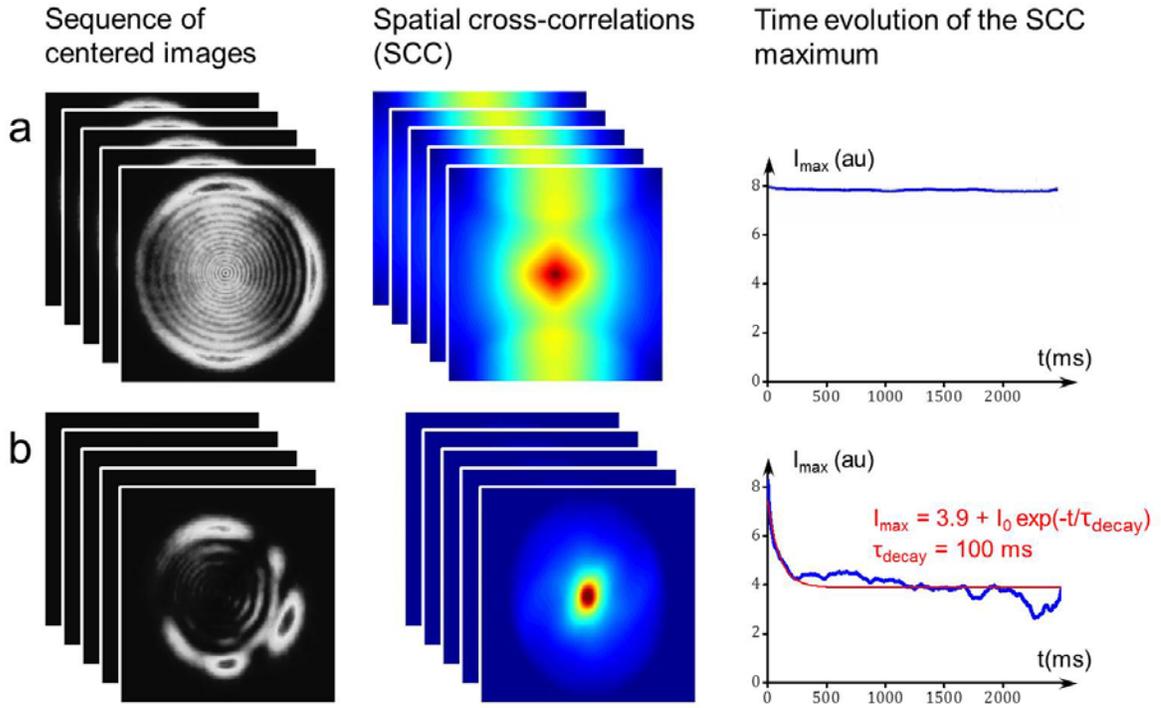


Figure 9-2. Time decay of the spatial cross-correlation (SCC) functions maximum of intensity profile fluctuations of a particle (a) and a cluster (b).

We determined the characteristic time of the SCC decay for the cluster with exponential decay function and we obtained $\tau_{decay} = 100$ ms, as shown in Fig. 9-2 b. If we assume that cluster is a solid object of $R_2 = 1$ μm radius, its rotational diffusivity will be

$$D_r = \frac{k_B T}{8\pi\eta R_2^3} = 1.1 \text{ s}^{-1}, \quad (9-1)$$

where we account for water viscosity $\eta = 1$ Pa s at $T = 22^\circ\text{C}$. Characteristic rotation time t_r in this case is

$$t_r = \frac{\langle\theta\rangle^2}{2} D_r \approx 10 \text{ s}. \quad (9-2)$$

Characteristic diffusion time τ_{decay} is 100-fold smaller than rotation time t_r , which

means that clusters intensity pattern is fluctuation due to a fluctuation of a cluster shape and not to its rotation.

Rayleigh studies on the oscillating droplets suggest an opportunity to extract the information on the intrinsic properties of oscillating object, such as surface tension σ and viscosity η . For the isolated droplet with a small amplitude axisymmetric oscillation and a weak dumping effect we can write

$$r(t) = r_0 + \sum_{n=2}^{\infty} r_n \cos(w_n t) P_n(\cos\theta) \exp\left(-\frac{t}{\tau_n}\right), \quad (9-3)$$

where $r(t)$ is an immediate amplitude of oscillation, r_0 is the actual radius of the droplet, θ is an angle between z-axis and a radial direction, $P_n(\cos\theta)$ is an n^{th} order Legendre polynomial, n is the order of oscillation, and the term corresponding to $n=1$ has the meaning of the translational motion and is omitted. Frequencies w_n and $\frac{1}{\tau_n}$ characterize oscillations regulated by two independent parameters, surface tension γ and viscosity η . The general solution for these constants have the following forms:

$$w_n^2 = n(n-1)(n+2) \frac{\gamma}{\rho r_0^3} \quad (9-4)$$

and

$$\frac{1}{\tau_n} = (n-1)(2n+1) \frac{\eta}{\rho r_0^2}. \quad (9-5)$$

In case of the second order oscillations, $n=2$, relations between the oscillation frequencies and droplet constants are

$$w_n^2 = \frac{8\gamma}{\rho r_0^3} \quad (9-6)$$

and

$$\frac{1}{\tau_n} = \frac{5\eta}{\rho r_0^2}. \quad (9-7)$$

With expected order of clusters parameters, γ and η , we use these expressions

to estimate the characteristic frequencies, w_n and $\frac{1}{\tau_n}$, regulating corresponding oscillations, and compare these values to the decay constant of SCC maximum, $\tau_{decay} = 100$ ms.

Taking $\gamma \propto 10^{-5}$ J/m², $\eta \propto 1$ cPa, $\rho \propto 0.1$ g/cm³, $r_0 = 1000$ nm, we obtain for characteristic times $\frac{1}{w_2} \propto 10^{-25}$ s and $\tau_2 \propto 10^{-11}$ s, which are significantly smaller than the decay time we obtained and the time resolution of the camera we use. This means, that with this approach we most likely cannot extract the intrinsic properties of clusters and the measured decay time $\tau_{decay} = 100$ ms is some characteristic time of cluster shape perturbation. The shape transformations are likely due to Brownian collisions with solvent molecules and the cluster intensity pattern results from a cumulative effect of these collisions, clusters and solvent properties.

9-3. Lysozyme liquid clusters can devour small-size particles and reveal their highly viscous core

In a different experiment we introduced fluorescent particles to a protein solution, containing large lysozyme clusters. A filter between the cuvette and the objective lens was used to obstruct the scattered light and only allowed fluorescent intensity to be registered. We saw that the majority of fluorescent particles diffused independently of each other and several groups of particles moved collectively.

In Fig. 9-3 we show three particles which moved as a group. The separation between them was ~ 2 μ m 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. Hence, the particles were trapped inside the cluster, confirming the liquid state of the clusters.

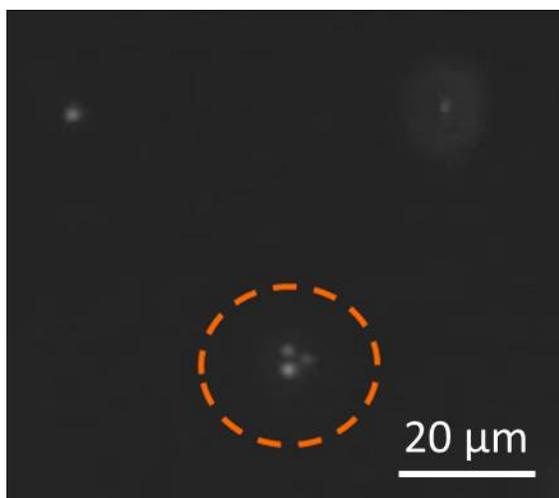


Figure 9-3. Clusters a liquid with a highly viscous core inside. Fluorescent microspheres are observed through a bandpass filter. A dashed circle indicates three particles entrapped inside a cluster.

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 were embedded in a medium of high viscosity.

Lysozyme solutions at high concentrations tend to exhibit gelation^{155,221} arresting the dynamics of the solution and increasing the viscosity. Thus, the observations in Fig. 9-3 suggest the presence of a highly viscous core inside the liquid lysozyme clusters.

9-4. Conclusions

With oblique illumination dark field microscopy we studied the intensity patterns of protein-rich liquid clusters and found that for large ($> \lambda/2$) clusters the intensity snapshots are highly asymmetrical and dynamic. From the height of the spatial autocorrelation function obtained from the centered images of clusters we extract the decorrelation lag time of these fluctuations. We obtain that this time is

smaller than the characteristic rotation time of a similar size solid object. Thus, we conclude that our clusters are liquid and the intensity pattern fluctuations are due to the cluster shape change. Cluster intrinsic properties, such as surface tension γ and viscosity η , with reasonable estimates from similar studies, would result in shape perturbations of extremely high frequencies, which could not be registered with used equipment. It means that the observed frequency of the cluster intensity fluctuations is a net effect of solvent molecules hitting clusters due to their Brownian motion. The possibility of small particles to enter the cluster structure is another confirmation of their liquid nature. Also the arrest of Brownian motility of these particles while entrapped in a protein cluster suggests that clusters possess a highly viscous core. This fact is in agreement with lysozyme solutions studies when this protein tends to gel at high concentrations.

CHAPTER 10

THE TWO-STEP MECHANISM OF NUCLEATION IN PORCINE INSULIN SOLUTIONS

Protein crystallization is notoriously difficult. Oftentimes crystals never form but instead amorphous aggregates, fibrils, dense liquids, and other undesirables appear^{95,222}. The main obstacle is not the slow growth of protein crystals: it was established that despite their large size and irregular shape^{223,224}, the crystal growth rates of many proteins are comparable to those of small-molecule crystals or at most two orders of magnitude slower²²⁵⁻²³¹. A structural impediment to protein crystallization is the lack of suitable intermolecular contact sites that are located on the surface of a protein in a pattern compatible with a crystallographic symmetry group. This problem is overcome through laborious searches towards procedures, conditions, and additives that enable such contacts^{95,222}, including genetic modifications of the protein surface residues²³²⁻²³⁴. The main kinetic impediment to protein nucleation appears to be a relatively large surface free energy γ of the crystal-solution interface. Typical estimates of γ may exceed 0.01 J m^{-2} ^{219,235,236}, which is comparable to values found for small-molecule materials. In combination with the large size of the protein molecules, the high γ leads to a high free energy barrier for nucleation²³⁷⁻²³⁹. In turn, this high barrier imposes high supersaturations at which nucleation may occur, thus expediting the formation of undesired solid phases.

To understand how proteins work around the high surface free energy problem, the two-step mechanism of nucleation was put forth. According to this mechanism, crystal nuclei assemble within preexisting protein-rich clusters rather than from

molecules in the dilute solution^{155,162,165,240}. Owing to the high protein concentration in the clusters, the surface free energy of a crystalline nucleus emerging in them drops off. Estimates of the typical surface free energy of the nucleus suggest that it is on the order of 10^{-4} J m^{-2} ^{217,218,236,241-243}, i.e., up to two orders of magnitude lower than γ of a nucleus forming in the solution. Since the nucleation barrier depends on the cube of γ ^{239,243}, this mechanism leads to orders of magnitude lower barriers and makes nucleation feasible over laboratory timescales.

Porcine insulin is one example of proteins revealing discrepancies of γ , estimated from rates of nucleation and liquid-solid interface. Previous experiments by Luis Filobello showed that the rate of homogeneous nucleation of insulin increases in the presence of acetone. Dynamic light scattering experiments could not resolve a clusters concentration difference in studied solutions. In the current project we employed Brownian microscopy to estimate cluster size and concentration in insulin solutions with and without acetone.

10-1. Materials and methods

Protein solutions were prepared following the protocol developed by L. Filobello which can be summarized as following. Porcine insulin powder (Sigma-Aldrich) with no additional purification was dissolved in 0.02 M HCl (Fisher Scientific) to make a stock solution. The stock was filtrated using a 0.22 μm pore size syringe filter (Millipore Millex-GV) and refrigerated at 4⁰C for later use. The insulin concentration of the stock was determined by measuring the absorbance of ultraviolet light employing a Beckman DU 68 spectrophotometer at a wavelength of 280 nm using quartz cuvettes. Background

measurements were done using 0.02 M HCl. The extinction coefficient used was 1.04 ml mg⁻¹ cm⁻¹.

Before each experiment, a solution with a target concentration was prepared using the insulin stock solution. A protocol to grow 2-zinc rhombohedral insulin crystals was developed based on papers published by Harding et al.²⁴⁴, Schlichtkrull^{245,246}, and McPherson¹² and is described below.

0.1 M zinc chloride (Sigma-Aldrich) was added to an aliquot of insulin solution in 0.02 M HCl. The zinc chloride solution had been clarified with drops of HCl to a pH of 3.23 approximately. Then, in the order given, 0.2 M tri-sodium citrate (Fisher), neat acetone (Fisher, HPLC-grade), and de-ionized water were added. The volume ratios were 10:1:5:3:1 to produce final concentrations of porcine insulin 1 – 4 mg ml⁻¹, HCl 0.01 M, ZnCl₂ 0.005 M, Na citrate 0.05 M, acetone 2.1 M. The final pH of the samples were measured using an Accumet AB 15 pH-meter (Fisher Scientific), yielding values 6.83 – 6.93. In the experiments where acetone was not utilized, the co-solvent was replaced with 0.02 M HCl and the pH was reduced by about 0.2 units.

To examine the Brownian motion of individual clusters in the tested solutions we employed Nanosight LM10-HS microscope (Nanosight Ltd) equipped with green laser $\lambda = 532$ nm and temperature control ($T = 22^{\circ}\text{C}$). The detailed description of Brownian microscopy method is given in chapter 2, Principal Experimental Techniques.

10-2. Insulin-rich liquid clusters help to explain the increase of nucleation rate

In Fig. 10-1 we show the results we obtained from Brownian microscopy method from insulin solutions. The general trend observed is the following: in the presence of

acetone clusters size increases which is indicated by the shift of the red picks in Fig. 10-1 a – d.

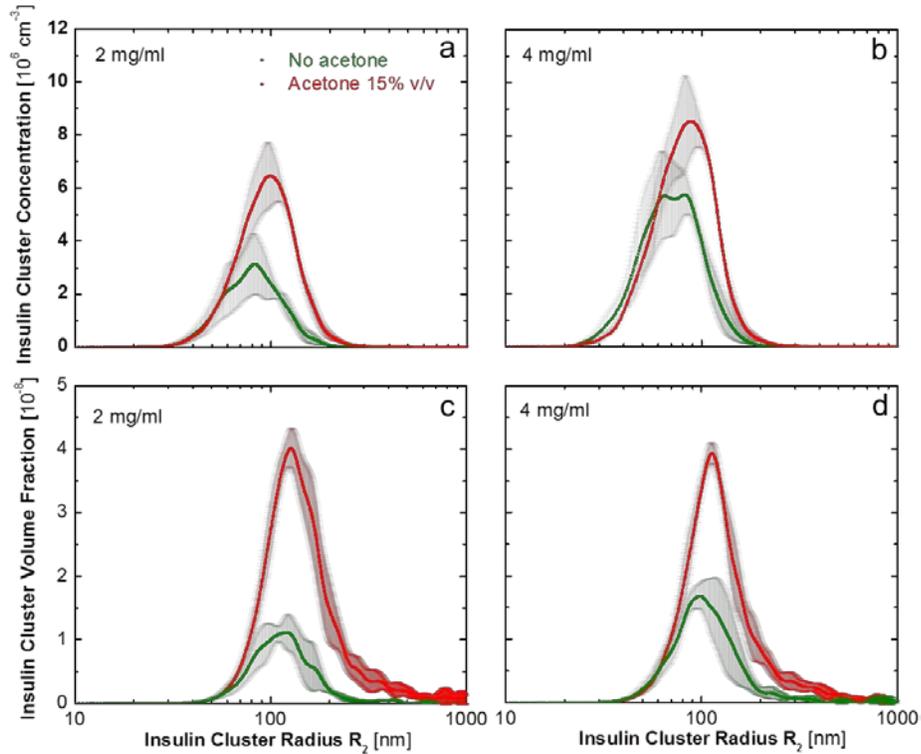


Figure 10-1. Response of porcine insulin solutions to the presence of acetone. Tested protein concentrations: 2 mg/ml (**a**, **c**) and 4 mg/ml (**b**, **d**). The results are shown in terms of clusters concentration (**a**, **b**) and volume fraction (**c**, **d**).

The height of red peaks always exceeds the height of corresponding green ones, which means that clusters number density and volume fraction increases, as well as cluster radius in the presence of acetone. These observations correlate with previously obtained results by L. Filobelo, who showed double increase in homogeneous nucleation rate in insulin solutions in the presence of acetone at 4 mg/ml protein concentration. From Fig. 10-1 d, cluster volume fraction also increases two-fold upon acetone addition.

10-3. Conclusions

We showed that previously observed the increase of homogeneous nucleation rate in solutions containing acetone correlates with cluster concentration in solutions of the same composition. We also estimated the surface free energy from the nucleation rates obtained previously by L. Filobelo. The values are of the order $0.1-0.2 \text{ mJ m}^{-2}$, which is an order of magnitude different from those, estimated from insulin solubility measurements, $\sim 3 \text{ mJ m}^{-2}$ ²⁴⁷. Observations above lead to conclusions that the protein-rich clusters emerge as crucial prerequisites for protein crystal nucleation and that insulin is another protein which follows the two-step mechanism of nucleation with protein-rich liquid clusters – nuclei precursors.

CHAPTER 11

THE TWO-STEP MECHANISM OF NUCLEATION – DIRECT OBSERVATION

Direct imaging of crystal nuclei forming within dense liquid clusters have been provided for two types of systems: colloids, which are larger and move slower than most molecules²⁴⁸, and an ingeniously chosen organic system²⁴⁹. The evidence of the action of this two-step mechanism to the formation of nuclei of protein crystals¹⁶², sickle cell anemia fibers⁴⁹, or amyloid fibrils^{250,251} has been mostly indirect: the two step mechanism was put forth to explain unusual non-monotonic dependencies of the protein crystal nucleation rate on supersaturation and a 10-order of magnitude discrepancy between the nucleation rates predicted by the classical nucleation theory assuming one-step crystal nucleation and the actual data^{162,165}. Nucleation of protein crystals and other ordered solids (e.g., sickle hemoglobin fibers) in stable dense protein liquid has been observed numerous times^{252,253}. Direct observation of protein crystal nucleation inside the metastable clusters is challenging due to a protein clusters size which typically is below the optical resolution limit. To tackle this challenge, in this work we monitored the initial progress of crystallization in glucose isomerase solutions with depolarized oblique illumination dark-field microscopy (DOIDM). Monitoring of supersaturated protein solutions by depolarized optics allows direct detection of crystal shortly after their formation.

11-1. Materials and methods

Glucose isomerase from *Streptomyces rubiginosus* (Hampton Research, USA) was dialyzed against 100 mM HEPES buffer at pH 7.0 containing 200 mM MgCl₂. Lysozyme from hen egg white (Seikagaku, Japan) was dissolved in 100 mM NaAc buffer

without further purification. The concentrations of protein solutions were determined by UV absorbance at 280 nm using extinction coefficient $1.042 \text{ ml mg}^{-1}\text{cm}^{-1}$ for glucose isomerase and $2.64 \text{ ml mg}^{-1}\text{cm}^{-1}$ for lysozyme. All stock solutions were filtered after their preparation through $0.22 \mu\text{m}$ syringe filters. Supersaturation of the studied solutions was achieved by adding $(\text{NH}_4)_2\text{SO}_4$ to glucose isomerase solutions and NaCl to lysozyme solutions. All experiments were carried out at 22°C .

A Nanosight LM10-HS microscope (Nanosight Ltd) equipped with a green laser (wavelength 532 nm) was employed to monitor individual clusters in the tested solutions. The raw data of this method are movies of point spread functions of clusters undergoing Brownian motion. To detect crystals, we modified the commercial set-up by adding a polarizer at the optical entrance of the sample cell and an analyzer at the optical exit, before the objective lens.

11-2. Nucleation of glucose isomerase and lysozyme crystals inside protein-rich clusters in a real time experiment

We tested the liquid nature of the protein-rich clusters, suggested by previous indirect evidence^{48,52,54} with aged solutions of glucose isomerase and lysozyme, which hold protein-rich clusters of relatively large size, $0.5\text{--}1.0 \mu\text{m}$. Oblique illumination dark-field microscopy performed without polarizer and analyzer, exhibits unique intensity patterns readily distinguishable from solid protein aggregates. The patterns are dynamic and highly asymmetrical, with interference fringes spreading from its center. The orientation of these fringes and their intensity vary in time. The minimum of the surface free energy for a liquid cluster corresponds to a spherical shape. However, Brownian collisions with the solvent molecules lead to deviations from this shape resulting in the

highly asymmetric dynamics of the intensity pattern. We conclude that the clusters are liquid and the surface free energy between the cluster and the solution is too low to stabilize a steady sphere. No determinations of the surface free energy of the protein-rich clusters have been performed. However, the free surface energy of dense liquid droplets in lysozyme solutions has been estimated as $\gamma = 4 \times 10^{-5} \text{ J m}^{-2}$ ²⁵⁴. With this γ , the excess free energy of a protrusion of radius $r = 100 \text{ nm}$ can be estimated as $\Omega\gamma/r \approx 10^{-23} \text{ J}$ ($\Omega = 3 \times 10^{-26} \text{ m}^{-3}$ is the molecular volume). This is significantly less than the driving force for shape change, the thermal energy of the solvent molecules $k_B T \approx 10^{-21} \text{ J}$ (k_B , Boltzmann constant; T , temperature). Assuming that the surface free energy of the protein-rich clusters in glucose isomerase solutions is similar to that value, the shape dynamics in response to Brownian collisions with the solvent is feasible and the glucose isomerase clusters are liquid.

To study the nucleation of crystals, we added $(\text{NH}_4)_2\text{SO}_4$ to glucose isomerase solutions and NaCl to lysozyme solutions and loaded the solutions in the DOIDM cell. If the polarizer and analyzer are in parallel orientation, immediately after solution loading we detected bright spots, corresponding to individual clusters; perpendicular arrangements of these two optical elements resulted in completely dark images. This indicates that only amorphous liquid clusters are present in the tested solutions. Forty minutes after the beginning of an experiment we see bright intensity spots with crossed polarizer and analyzer, indicating the presence of objects that rotate the plane of light polarization. Solution filtration prior to loading removed all stray crystals from the monitored volume; hence we conclude that the only particles which capable of light depolarization are protein crystals. This signal indicates protein crystal nucleation.

Further observations revealed three types of particles in the solution: 1. Visible only with parallel polarizer and analyzer with a cluster-like intensity pattern, i.e., large protein clusters. 2. Visible with both parallel and crossed polarizer and analyzer, but with steady shape, we conclude that these are large freely diffusing crystals. 3. Particles visible in both parallel and crossed polarizer and analyzer, but with a cluster-like pattern in parallel mode, and crystal-like pattern in crossed mode. We conclude that these are protein clusters with entrapped protein crystals inside.

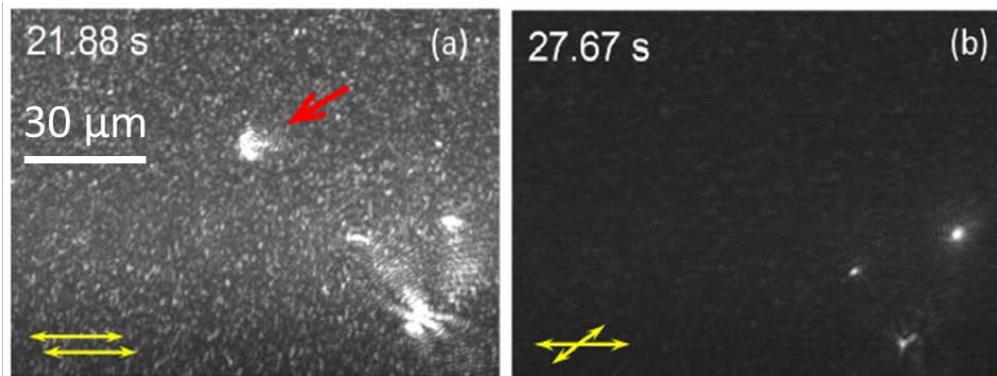


Figure 11-1. Snapshots from DOIDM showing three types of scatterers in a lysozyme solution under crystallization conditions: clusters, crystals, and clusters containing crystals. Orientation of polarizers is indicated in the figure.

In Fig. 11-1 we show two snapshots from a DOIDM movie, where we indicate the position of analyzer and polarizer and time corresponding to these frames. Fig. 11-1 a, is an example of intensity patterns typically observed for large clusters in parallel polarizers (indicated with an arrow). Fig. 11-1 b shows three crystals in crossed polarizers seen as steady bright spots. Comparing the locations of the crystals and three of the clusters, we conclude that these are most probably protein clusters with entrapped protein crystals inside.

11-3. Conclusions

With oblique illumination dark field microscopy we monitored the clusters dynamic behavior in solutions of two proteins, glucose isomerase and lysozyme. We observed the intensity patterns in these solutions typical to clusters of a large size ($> \lambda/2$). We modified the existing oblique illumination microscopy setup by introducing a polarizer and an analyzer. We confirmed that clusters have amorphous structure, i.e., they did not yield any signal when the polarizer and the analyzer were perpendicular to each other. After initiating the crystallization conditions for both of the proteins by adding corresponding salts, after a certain time we registered a signal in crossed analyzer and polarizer setup. This fact means that in the protein solutions we initiated nucleation and microcrystals reached a detectable size. Further, we were able to distinguish between three types of scatterers, which we determined as single protein clusters, single protein crystals and crystals entrapped in clusters. The latter type of scatterers is the first direct observation of the two-step mechanism of nucleation of proteins.

CHAPTER 12

CONSLUSIONS

In this work we aimed to deepen our knowledge on a recently discovered phase in protein solutions, which forms in the homogeneous region of the protein phase diagram and play role of nucleation precursors in the two-step mechanism of protein nucleation – metastable mesoscopic protein-rich clusters. The mechanism of cluster formation and their behavior still remain challenging questions. Why do clusters form? Are clusters just nuclei of a dense liquid phase? What are the forces that hold protein molecules inside the clusters? How can we affect cluster behavior by external conditions? How can we prove that clusters are precursors in the two-step mechanism of nucleation? We addressed these questions in our study and progressed the clusters nature research.

First, we confirmed that the protein-rich clusters are not the nuclei of a dense liquid phase by monitoring the long time evolution of cluster radius. After over 600 hours lysozyme clusters did not mature into a new liquid phase and their size reached the terminal value at ~200 nm. With this we conclude that the protein clusters are not the dense liquid phase and they are in near-equilibrium with the solution, i.e., they are metastable mesoscopic protein-rich clusters.

In 2010 Pan et al.⁴⁶ proposed a mechanism by which the protein clusters exist. Following it, clusters form due to a presence of new species in solution, such as protein oligomers. The origin of those oligomers might result from domain swapping, as for example in case of lysozyme, which monomers consist of a single chain, or misassembled molecules, like glucose isomerase protein, which exists in solutions in

tetrameric state, though under high energetic cost it can have trimers, pentamers or other oligomers (probability of their existence would decrease the further the oligomer is from a tetrameric combination).

We attempted to identify the possible regions of flexibility of the protein monomers most likely participating in cluster formation. We carried out the nuclear magnetic resonance experiments with lysozyme solutions, which allowed us to monitor the ^1H - ^2D exchange rate in solutions with abundant amount of clusters in comparisons to a solution with a low cluster concentration. Protein lysozyme is a profoundly studied molecule, thus there is a significant amount of literature available on its characteristics. We were able to identify the NMR signal peaks according to the literature data and found the residues which exhibit a less protected state in solutions with a high cluster concentration. Locations of these residues on a protein sequence map indeed suggest that these are the regions of possible high monomer flexibility potentially contributing to domain-swapped oligomerization.

The forces controlling clusters meta-stability were still unknown in cluster research. We tested the significance of electrostatics on cluster formation. Coulomb interactions dominate in many areas of biochemical processes and they were the first candidate to be considered on a leading role of clustering aggregation. We discovered that, despite the fact that electrostatic interactions are highly important in protein world, their effect on clusters is minimal. Our experiments demonstrate that the mesoscopic clusters represent 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, discussed in introduction to chapter 5. 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. Since 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 inter-molecular water-structuring interactions may be an essential tool to control the cluster population and in this way enhance or suppress protein crystallization and fibrillization.

The experiments in which the protein solution undergoes bubbling with inert gases support the previous arguments. We notice the high importance of water structuring interactions on the example of lysozyme clusters.

The mechanism of cluster formation has its general idea – clusters require the presence of new protein species in solution to appear. Nevertheless, the particular pathway the clusters form depends on the protein identity. Thus, in case of glucose isomerase we show that the water structuring interactions are not the principal forces governing cluster formation, though the shear stress induced by bubbling with inert gas has a major effect on cluster population. The origin of such response is in the mechanism of oligomerization. Since the native confirmation of glucose isomerase molecules in solutions is tetrameric, the existence of new protein species could be granted by simply misassembled glucose isomerase molecules such as trimers or pentamers. The induced shear rates break high energy oligomers decreasing the lifetime of those and leading to a lower cluster volume fraction and concentration.

Besides the challenging thermodynamics of cluster existence, which makes them an interesting object to study, cluster importance in protein science is also in their role of the two-step mechanism of nucleation. According to this mechanism, clusters play role of nucleation precursors which assist protein crystal nucleation.

Addition of organic compounds, such as ethanol, acetone and others, is known to facilitate protein crystallization, increase the size of crystal, help to produce crystals with minimum defects. All protocols of protein crystallization are the result of empirical research and no general explanation exist why in one case certain cosolvents would help crystallization while in other protein system they will not have any effect. Here, we tested the influence of common organic additives (ethanol, acetone, glycerol, MPD) on protein-rich clusters as potential nucleation precursors. We found that the response of protein monomers behavior is different based on the cosolvents chemical identity, as well as the chemical potential of studied systems shows various trends with increase of the additives concentration.

Protein clusters assist protein nucleation though the exact mechanism of how and where crystals form is still elusive. There are two possible scenarios of crystals birth. 1) They nucleate within the cluster bulk, i.e., follow the homogenous nucleation pathway, or 2) crystals utilize the cluster surface following heterogeneous nucleation mechanism. We found that, despite the chemistry of additives used in our tests, the effect of all cosolvents resulted in increase of clusters volume fraction and their surface area, parameters important for previously noted scenarios. This finding supports the idea of clusters importance in the two-step mechanism of nucleation as essential precursors in protein crystallization.

Brownian microscopy, a relatively new method of tracking microparticles in solutions, made a significant step forward in clusters research. The movies of this method are intensity snapshots of clusters undergoing Brownian motion. When protein clusters rich ca. micron size their intensity patterns exhibit dynamic fluctuations, not typical for solid objects. With oblique illumination dark field microscopy (next step modification of Brownian microscopy method, when the interest is in the studying of individual clusters intensity patterns rather than their collective motion) we showed that protein clusters are liquid and have a highly viscous core inside, as in case of protein lyzosome.

Tracking micron-size clusters is a challenging task, since there were no available computer algorithms capable to deal with dynamic intensity patterns of clusters. We developed a new tracking method, based on spatial cross correlation, which overcomes the difficulties associated with large ($> \lambda/2$) clusters. We tested the new method on the latex particles of known size and showed that the method of the spatial cross-correlation yields superior results in comparison to other common tracking algorithms on the research market.

On the example of protein insulin and the results obtained by L. Filobelo in 2003 we showed why the two-step mechanism of nucleation is selected. Evaluation of the surface free energy cost associated with nuclei formation from nucleation rates of insulin are 10 orders of magnitude different from the surface free energy estimations of insulin solubility. These calculations propose that the two-step mechanism of nucleation is selected because it provides the lower free energy barrier associated with nucleation from clusters than following the direct pathway. The results from Brownian microscopy

experiments confirmed the presence of clusters in insulin solutions as well as they show that in the presence of acetone as a cosolvent, cluster volume fraction is increased, which correlates with the increase of previously measured nucleation rates in identical conditions.

The evidence of the two-step mechanism of nucleation in protein solutions is mostly indirect. Here we showed for the first time that protein crystals appear inside protein clusters when the crystallization conditions are induced.. This became feasible with further modification of the oblique illumination dark field microscopy when we introduced polarizer and analyzer in the setup to distinguish between amorphous and crystalline populations of scatterers. We saw that crystals appear in large size clusters (the size is a limiting factor of the technique, since we will not be able to resolve a nucleated crystal inside a small-size cluster), they grow in them and then fall off of them under the gravity. These experiments were carried on with two different proteins, lysozyme and glucose isomerase, and in both setups we obtained similar results. The videos of clusters with crystals entrapped inside them is the first direct evidence of the two-step mechanism of protein nucleation.

REFERENCES

1. Hunter, M. S.; DePonte, D. P.; Shapiro, D. A.; Kirian, R. A.; Wang, X.; Starodub, D.; Marchesini, S.; Weierstall, U.; Doak, R. B.; Spence, J. C. H.; Fromme, P. *Biophysical Journal* **2011**, *100*, 198.
2. Mueller, M.; Jenni, S.; Ban, N. *Curr Opin Struct Biol* **2007**, *17*, 572.
3. Drenth, J. *Principles of Protein X-ray Crystallography*; Springer Science: New York, 2007.
4. DeLucas, L. J. *Membrane Protein Crystallization*; Academic Press: Burlington, MA, 2009.
5. Ducruix, A., Giegé, R. *Crystallization of nucleic acids and proteins. A practical approach*; Oxford University Press: New York, 1992.
6. Hünefeld, F. L. *Leipsic* **1840**, *160*, quoted by Preyer (11), p.1.
7. Funke, O. *Z. rationell Med* **1851**, *i*, 185.
8. Funke, O. *Z. rationell Med* **1852**, *ii*, 199.
9. Hoppe-Seyler, F. *Archiv für pathologische Anatomie und Physiologie und für klinische Medicin*, *29*, 597.
10. Debru, C. *L'esprit des protéines : histoire et philosophie biochimiques*; Hermann: Paris, 1983.
11. Maschke *Journal für Praktische Chemie* **1858**, *ixxiv*, 436.
12. McPherson, A. *Crystallization of biological macromolecules*; Cold Spring Harbor Lab Press, New York, 1999.
13. Hopkins, F. G. P., S.N. *Jour Physiol* **1898-99**, *xxiii*, 130.

14. Sumner, J. B. *Journal of Biological Chemistry* **1926**, 69, 435.
15. Northrop, J. H. *Science* **1929**, 69, 580.
16. Kay, L. E. *Isis* **1986**, 77, 450.
17. Bernal, J. D. C., D. *Nature* **1934**, 133, 794.
18. Basu, S. K.; Govardhan, C. P.; Jung, C. W.; Margolin, A. L. *Expert Opin Biol Ther* **2004**, 4, 301.
19. Charache, S.; Conley, C. L.; Waugh, D. F.; Ugoretz, R. J.; Spurrell, J. R. *J Clin Invest* **1967**, 46, 1795.
20. Hirsch, R. E. R.-S., C.; Olson, J.A.; Nagel, R.L. *Blood* **1985**, 66, 775.
21. Eaton, W. A.; Hofrichter, J. *Adv Protein Chem* **1990**, 40, 63.
22. Vekilov, P. G. *British Journal of Haematology* **2007**, 139, 173.
23. Berland, C. R.; Thurston, G. M.; Kondo, M.; Broide, M. L.; Pande, J.; Ogun, O.; Benedek, G. B. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, 89, 1214.
24. Asherie, N.; Pande, J.; Pande, A.; Zarutskie, J. A.; Lomakin, J.; Lomakin, A.; Ogun, O.; Stern, L. J.; King, J.; Benedek, G. B. *J Mol Biol* **2001**, 314, 663.
25. Pande, A.; Pande, J.; Asherie, N.; Lomakin, A.; Ogun, O.; King, J.; Benedek, G. B. *Proc Natl Acad Sci U S A* **2001**, 98, 6116.
26. Koo, E. H.; Lansbury, P. T.; Kelly, J. W. *Proceedings of the National Academy of Sciences* **1999**, 96, 9989.
27. Owen, R. L.; Rudino-Pinera, E.; Garman, E. F. *Proc Natl Acad Sci U S A* **2006**, 103, 4912.
28. Henderson, R. *Q Rev Biophys* **1995**, 28, 171.

29. Riekell, C. *J Synchrotron Radiat* **2004**, *11*, 4.
30. Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; DePonte, D. P.; Weierstall, U.; Doak, R. B.; Maia, F. R. N. C.; Martin, A. V.; Schlichting, I.; Lomb, L.; Coppola, N.; Shoeman, R. L.; Epp, S. W.; Hartmann, R.; Rolles, D.; Rudenko, A.; Foucar, L.; Kimmel, N.; Weidenspointner, G.; Holl, P.; Liang, M.; Barthelmeß, M.; Caleman, C.; Boutet, S.; Bogan, M. J.; Krzywinski, J.; Bostedt, C.; Bajt, S.; Gumprecht, L.; Rudek, B.; Erk, B.; Schmidt, C.; Homke, A.; Reich, C.; Pietschner, D.; Struder, L.; Hauser, G.; Gorke, H.; Ullrich, J.; Herrmann, S.; Schaller, G.; Schopper, F.; Soltau, H.; Kuhnel, K.-U.; Messerschmidt, M.; Bozek, J. D.; Hau-Riege, S. P.; Frank, M.; Hampton, C. Y.; Sierra, R. G.; Starodub, D.; Williams, G. J.; Hajdu, J.; Timneanu, N.; Seibert, M. M.; Andreasson, J.; Rocker, A.; Jonsson, O.; Svenda, M.; Stern, S.; Nass, K.; Andritschke, R.; Schroter, C.-D.; Krasniqi, F.; Bott, M.; Schmidt, K. E.; Wang, X.; Grotjohann, I.; Holton, J. M.; Barends, T. R. M.; Neutze, R.; Marchesini, S.; Fromme, R.; Schorb, S.; Rupp, D.; Adolph, M.; Gorkhover, T.; Andersson, I.; Hirsemann, H.; Potdevin, G.; Graafsma, H.; Nilsson, B.; Spence, J. C. H. *Nature* **2011**, *470*, 73.
31. Gibbs, J. W. *Trans Connect Acad Sci* **1876**, *3*, 108.
32. Gibbs, J. W. *Trans Connect Acad Sci* **1878**, *3*, 343.
33. De Yoreo, J. J.; Vekilov, P. G. *Reviews in Mineralogy and Geochemistry* **2003**, *54*, 57.
34. Yau, S. T.; Petsev, D. N.; Thomas, B. R.; Vekilov, P. G. *J Mol Biol* **2000**, *303*, 667.
35. Vekilov, P. G.; Feeling-Taylor, A. R.; Yau, S. T.; Petsev, D. *Acta Crystallogr D Biol Crystallogr* **2002**, *58*, 1611.

36. Callister, W. D. *Materials Science and Engineering, Introduction*; John Wiley & Sons, 2007.
37. Nielsen, A. E. *Kinetics of Precipitation*; Pergamon: Oxford, 1964.
38. Abraham, F. F. *Homogeneous Nucleation Theory*; Academic Press: New York, 1974.
39. Chernov, A. A. *Modern Crystallography III Crystal Growth*; Springer-Verlag: Berlin Heidelberg New York Tokyo, 1984.
40. Galkin, O.; Vekilov, P. G. *Journal of the American Chemical Society* **2000**, *122*, 156.
41. Yau, S. T.; Vekilov, P. G. *Nature* **2000**, *406*, 494.
42. Malkin, A. J.; McPherson, A. *Acta Crystallographica Section D* **1994**, *50*, 385.
43. Mullin, J. B.; Surek, T.; Malkin, A. J.; McPherson, A. *Journal of Crystal Growth* **1993**, *128*, 1232.
44. Vekilov, P. G. *Gold nanoparticles: Grown in a crystal*; Nat Nanotechnol. 2011 Feb;6(2):82-3. doi: 10.1038/nnano.2011.9.
45. Yau, S. T.; Vekilov, P. G. *Journal of the American Chemical Society* **2001**, *123*, 1080.
46. Pan, W.; Vekilov, P. G.; Lubchenko, V. *The Journal of Physical Chemistry B* **2010**, *114*, 7620.
47. Pan, W.; Uzunova, V. V.; Vekilov, P. G. *Biopolymers* **2009**, *91*, 1108.
48. Gliko, O.; Neumaier, N.; Pan, W.; Haase, I.; Fischer, M.; Bacher, A.; Weinkauff, S.; Vekilov, P. G. *J Am Chem Soc* **2005**, *127*, 3433.
49. Galkin, O.; Pan, W.; Filobelo, L.; Hirsch, R. E.; Nagel, R. L.; Vekilov, P. G. *Biophys J* **2007**, *93*, 902.

50. Li, Y.; Lubchenko, V.; Vorontsova, M. A.; Filobelo, L.; Vekilov, P. G. *J Phys Chem B* **2012**, *116*, 10657.
51. Pan, W.; Filobelo, L.; Pham, N. D.; Galkin, O.; Uzunova, V. V.; Vekilov, P. G. *Phys Rev Lett* **2009**, *102*, 2.
52. Gliko, O.; Pan, W.; Katsonis, P.; Neumaier, N.; Galkin, O.; Weinkauf, S.; Vekilov, P. G. *J Phys Chem B* **2007**, *111*, 3106.
53. Sleutel, M.; Van Driessche, A. E. S. *Proceedings of the National Academy of Sciences* **2014**, *111*, E546.
54. Pan, W.; Galkin, O.; Filobelo, L.; Nagel, R. L.; Vekilov, P. G. *Biophys J* **2007**, *92*, 267.
55. Pan, W.; Galkin, O.; Filobelo, L.; Nagel, R. L.; Vekilov, P. G. *Biophys J* **2007**, *92*, 267.
56. Uzgiris, E. E.; Golibersuch, D. C. *Physical Review Letters* **1974**, *32*, 37.
57. Groenewold, J.; Kegel, W. K. *The Journal of Physical Chemistry B* **2001**, *105*, 11702.
58. Sciortino, F.; Mossa, S.; Zaccarelli, E.; Tartaglia, P. *Physical Review Letters* **2004**, *93*, 055701.
59. Liu, Y.; Chen, W. R.; Chen, S. H. *J Chem Phys* **2005**, *122*, 44507.
60. Mossa, S.; Sciortino, F.; Tartaglia, P.; Zaccarelli, E. *Langmuir* **2004**, *20*, 10756.
61. Stradner, A.; Sedgwick, H.; Cardinaux, F.; Poon, W. C.; Egelhaaf, S. U.; Schurtenberger, P. *Nature* **2004**, *432*, 492.
62. Bennett, M. J.; Sawaya, M. R.; Eisenberg, D. *Structure* **2006**, *14*, 811.
63. Schlunegger, M. P.; Bennett, M. J.; Eisenberg, D. *Adv Protein Chem* **1997**, *50*, 61.

64. Vekilov, P. G. *Nucleation and Growth Mechanisms of Protein Crystals in Handbook of Crystal Growth*; Elsevier: Amsterdam, 2015; Vol. 1.
65. Maes, D.; Vorontsova, M. A.; Potenza, M. A. C.; Sanvito, T.; Sleutel, M.; Giglio, M.; Vekilov, P. G. *Acta Crystallographica Section F* **2015**, *71*, 815.
66. Vorontsova, M. A.; Maes, D.; Vekilov, P. G. *Faraday Discussions* **2015**, *179*, 27.
67. Vekilov, P. G.; Vorontsova, M. A. *Acta Crystallogr F Struct Biol Commun* **2014**, *70*, 271.
68. Shukla, A.; Mylonas, E.; Di Cola, E.; Finet, S.; Timmins, P.; Narayanan, T.; Svergun, D. I. *Proc Natl Acad Sci U S A* **2008**, *105*, 5075.
69. Porcar, L.; Falus, P.; Chen, W.-R.; Faraone, A.; Fratini, E.; Hong, K.; Baglioni, P.; Liu, Y. *The Journal of Physical Chemistry Letters* **2010**, *1*, 126.
70. *Physical Chemistry*; 2nd edition ed.; Berry, P. S., Rice, S. A., Ross, J., Ed.; Oxford University Press, 2000.
71. Schmitz, K. S. *An Introduction to Dynamic Light Scattering by Macromolecules*; Academic Press, 1990
72. Schatzel, K.; Brown, W., Ed.; Clarendon Press: Oxford, 1993, p 76.
73. Provencher, S. *Computer Physics Communications* **1982**, *27*, 229.
74. Provencher, S. *Comp Phys Commun* **1982**, *27*, 213.
75. Li, Y.; Lubchenko, V.; Vekilov, P. G. *Review of Scientific Instruments* **2011**, *82*, 053106.
76. McCammon, J. A.; Gelin, B. R.; Karplus, M.; Wolynes, P. G. *Nature* **1976**, *262*, 325.
77. De Simone, A.; Montalvao, R. W.; Dobson, C. M.; Vendruscolo, M. *Biochemistry* **2013**, *52*, 6480.

78. Radford, S. E.; Buck, M.; Topping, K. D.; Dobson, C. M.; Evans, P. A. *Proteins* **1992**, *14*, 237.
79. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Structure, Function, and Bioinformatics* **1993**, *17*, 75.
80. Schwalbe, H.; Grimshaw, S. B.; Spencer, A.; Buck, M.; Boyd, J.; Dobson, C. M.; Redfield, C.; Smith, L. J. *Protein Sci* **2001**, *10*, 677.
81. Shah, M.; Galkin, O.; Vekilov, P. G. *The Journal of Chemical Physics* **2004**, *121*, 7505.
82. Shah, M.; Galkin, O.; Vekilov, P. G. *J Phys Chem B* **2009**, *113*, 7340.
83. Eyink, G.; Lebowitz, J.; Spohn, H. *J Stat Phys* **1996**, *83*, 385.
84. Pan, W.; Vekilov, P. G.; Lubchenko, V. *J Phys Chem B* **2010**, *114*, 7620.
85. Goto, Y.; Calciano, L. J.; Fink, A. L. *Proc Natl Acad Sci U S A* **1990**, *87*, 573.
86. Thornton, J. M. *Nature* **1982**, *295*, 13.
87. Meltzer, R. H.; Thompson, E.; Soman, K. V.; Song, X. Z.; Ebalunode, J. O.; Wensel, T. G.; Briggs, J. M.; Pedersen, S. E. *Biophys J* **2006**, *91*, 1302.
88. Anderson, C. F.; Record, M. T. *Annual Review of Physical Chemistry* **1982**, *33*, 191.
89. Ozkirimli, E.; Yadav, S. S.; Miller, W. T.; Post, C. B. *Protein Sci* **2008**, *17*, 1871.
90. Dean, A. M.; Koshland, D. E., Jr. *Science* **1990**, *249*, 1044.
91. Hemsath, L.; Dvorsky, R.; Fiegen, D.; Carlier, M. F.; Ahmadian, M. R. *Mol Cell* **2005**, *20*, 313.
92. Buell, A. K.; Hung, P.; Salvatella, X.; Welland, M. E.; Dobson, C. M.; Knowles, T. P. *Biophysical journal* **2013**, *104*, 1116.

93. Chiti, F.; Calamai, M.; Taddei, N.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99 Suppl 4*, 16419.
94. Marshall, K. E.; Morris, K. L.; Charlton, D.; O'Reilly, N.; Lewis, L.; Walden, H.; Serpell, L. C. *Biochemistry* **2011**, *50*, 2061.
95. McPherson, A. *Introduction to Macromolecular Crystallography* John Wiley: Hoboken, New Jersey, 2009.
96. Chayen, N. E.; Helliwell, J. R.; Snell, E. H. *Macromolecular Crystallization and Crystal Perfection. IUCr Monographs on Crystallography, 24.* ; Oxford University Press: Oxford, New York, 2010.
97. Chan, Ho Y.; Lankevich, V.; Vekilov, Peter G.; Lubchenko, V. *Biophysical Journal* **2012**, *102*, 1934.
98. Aune, K. C.; Tanford, C. *Biochemistry* **1969**, *8*, 4579.
99. Roxby, R.; Tanford, C. *Biochemistry* **1971**, *10*, 3348.
100. Provencher, S. W. *Comp Phys Communications* **1982**, *27*, 229.
101. Li, Y.; Lubchenko, V.; Vekilov, P. G. *Rev Sci Instr* **2011**, *82*, 053106
102. Tanford, C.; Kirkwood, J. G. *Journal of the American Chemical Society* **1957**, *79*, 5333.
103. Beretta, S.; Chirico, G.; Baldini, G. *Macromolecules* **2000**, *33*, 8663.
104. Nir, S. *Progress in Surface Science* **1977**, *8*, 1.
105. Allahyarov, E.; Löwen, H.; Hansen, J.; Louis, A. *Physical Review E* **2003**, *67*.
106. Neal, B. L.; Asthagiri, D.; Velev, O. D.; Lenhoff, A. M.; Kaler, E. W. *Journal of Crystal Growth* **1999**, *196*, 377.

107. Wetter, L. R.; Deutsch, H. F. *Journal of Biological Chemistry* **1951**, *192*, 237.
108. Richetti, P. G.; Tudor, G. *J Chromatography* **1981**, *220*, 115.
109. Leckband, D.; Israelachvili, J. *Quart Rev Biophysics* **2001**, *34*, 105.
110. George, A.; Wilson, W. W. *Acta Crystallogr Section D* **1994**, *50*, 361.
111. Li, Y.; Lubchenko, V.; Vorontsova, M. A.; Filobelo, L.; Vekilov, P. G. *The Journal of Physical Chemistry B* **2012**, *116*, 10657.
112. Sciortino, F.; Mossa, S.; Zaccarelli, E.; Tartaglia, P. *Phys Rev Lett* **2004**, *93*, 055701.
113. Hutchens, S. B.; Wang, Z.-G. *J Chem Phys* **2007** *127*, 084912
114. Mohr, B. G.; Dobson, C. M.; Garman, S. C.; Muthukumar, M. *The Journal of Chemical Physics* **2013**, *139*, 121914.
115. Galkin, O.; Pan, W.; Filobelo, L.; Hirsch, R. E.; Nagel, R. L.; Vekilov, P. G. *Biophys J* **2007**, *93*, 902.
116. Gliko, O.; Neumaier, N.; Pan, W.; Haase, I.; Fischer, M.; Bacher, A.; Weinkauff, S.; Vekilov, P. G. *J Amer Chem Soc* **2005**, *127*, 3433.
117. Sleutel, M.; Van Driessche, A. E. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, E546.
118. Petsev, D. N.; Vekilov, P. G. *Phys Rev Lett* **2000**, *84*, 1339.
119. Das, A.; Mukhopadhyay, C. *J Phys Chem B* **2008**, *112*, 7903.
120. Caballero-Herrera, A.; Nordstrand, K.; Berndt, K. D.; Nilsson, L. *Biophys J* **2005**, *89*, 842.
121. Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. *Proceedings of the National Academy of Sciences* **2008**, *105*, 16928.

122. Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. *Proc Natl Acad Sci U S A* **2007**, *104*, 15317.
123. Courtenay, E. S.; Capp, M. W.; Record, M. T. *Protein Sci* **2001**, *10*, 2485.
124. Diehl, R. C.; Guinn, E. J.; Capp, M. W.; Tsodikov, O. V.; Record, M. T. *Biochemistry* **2013**, *52*, 5997.
125. Holehouse, A. S.; Garai, K.; Lyle, N.; Vitalis, A.; Pappu, R. V. *Journal of the American Chemical Society* **2015**, *137*, 2984.
126. Sagle, L. B.; Zhang, Y.; Litosh, V. A.; Chen, X.; Cho, Y.; Cremer, P. S. *Journal of the American Chemical Society* **2009**, *131*, 9304.
127. Banerjee, S.; Ghosh, R.; Bagchi, B. *The Journal of Physical Chemistry B* **2012**, *116*, 3713.
128. Deshpande, A.; Nimsadkar, S.; Mande, S. C. *Acta Crystallographica Section D* **2005**, *61*, 1005.
129. Chattoraj, S.; Mandal, A. K.; Bhattacharyya, K. *The Journal of Chemical Physics* **2014**, *140*, 115105.
130. Collins, K. D. *Methods* **2004**, *34*, 300.
131. Caussette, M.; Gaunand, A.; Planche, H.; Colombié, S.; Monsan, P.; Lindet, B. *Enzyme and Microbial Technology* **1999**, *24*, 412.
132. Bekard, I. B.; Asimakis, P.; Bertolini, J.; Dunstan, D. E. *Biopolymers* **2011**, *95*, 733.
133. Kasumi, T.; Hayashi, K.; Tsumura, N.; Takagi, T. *Bioscience, Biotechnology, and Biochemistry* **1981**, *45*, 1087.
134. Liu, S. Y.; Wiegel, J.; Gherardini, F. C. *J Bacteriol* **1996**, *178*, 5938.

135. Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter *Molecular Biology of the Cell*; 4 ed.; New York: Garland Science, 2002.
136. Arakawa, T.; Timasheff, S. N. *Methods Enzymol* **1985**, *114*, 49.
137. Schein, C. H. *Biotechnology* **1990**, *8*, 308.
138. Tayyab M., Q. S., Islam M. *Medical Science Research* **1993**, *21*, 805.
139. Liu, W.; Bratko, D.; Prausnitz, J. M.; Blanch, H. W. *Biophys Chem* **2004**, *107*, 289.
140. Bonincontro, A.; De Francesco, A.; Matzeu, M.; Onori, G.; Santucci, A. *Colloids and Surfaces B: Biointerfaces* **1997**, *10*, 105.
141. Boyer, M.; Roy, M.-O.; Jullien, M.; Bonneté, F.; Tardieu, A. *Journal of Crystal Growth* **1999**, *196*, 185.
142. Deshpande, A.; Nimsadkar, S.; Mande, S. C. *Acta Crystallogr D Biol Crystallogr* **2005**, *61*, 1005.
143. Cinelli, S.; Onori, G.; Santucci, A. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **1999**, *160*, 3.
144. Yoshikawa, H.; Hirano, A.; Arakawa, T.; Shiraki, K. *Int J Biol Macromol* **2012**, *50*, 1286.
145. Nemčovičová, I. *Alternative Protein Crystallization Technique: Cross-Influence Procedure (CIP)*; InTech, 2012.
146. Kovrigin, E. L.; Potekhin, S. A. *Biochemistry* **1997**, *36*, 9195.
147. Anand, K.; Pal, D.; Hilgenfeld, R. *Acta Crystallogr D Biol Crystallogr* **2002**, *58*, 1722.
148. Berger, B. W.; Blamey, C. J.; Naik, U. P.; Bahnson, B. J.; Lenhoff, A. M. *Crystal Growth & Design* **2005**, *5*, 1499.

149. Pittz, E. P.; Bello, J. *Archives of Biochemistry and Biophysics* **1973**, *156*, 437.
150. Pittz, E. P.; Timasheff, S. N. *Biochemistry* **1978**, *17*, 615.
151. Tanaka, Y.; Hirano, N.; Kaneko, J.; Kamio, Y.; Yao, M.; Tanaka, I. *Protein Sci* **2011**, *20*, 448.
152. Pittz, E. P.; Bello, J. *Arch Biochem Biophys* **1971**, *146*, 513.
153. Stauber, M.; Jakoncic, J.; Berger, J.; Karp, J. M.; Axelbaum, A.; Sastow, D.; Buldyrev, S. V.; Hrnjez, B. J.; Asherie, N. *Acta Crystallogr D Biol Crystallogr* **2015**, *71*, 427.
154. Michaux, C.; Pouyez, J.; Wouters, J.; Privé, G. G. *BMC Structural Biology* **2008**, *8*, 1.
155. Galkin, O.; Vekilov, P. G. *Proc Natl Acad Sci USA* **2000**, *97*, 6277.
156. Lu, J.; Wang, X.-J.; Ching, C.-B. *Crystal Growth & Design* **2003**, *3*, 83.
157. Fisticaro, E.; Compari, C.; Braibanti, A. *Biophys Chem* **2011**, *156*, 51.
158. Vekilov, P. G.; Pan, W.; Gliko, O.; Katsonis, P.; Galkin, O. In *Aspects of Physical Biology*; Franzese, G., Rubi, M., Eds.; Springer Berlin Heidelberg: 2008; Vol. 752, p 65.
159. Yearley, Eric J.; Godfrin, Paul D.; Perevozchikova, T.; Zhang, H.; Falus, P.; Porcar, L.; Nagao, M.; Curtis, Joseph E.; Gawande, P.; Taing, R.; Zarraga, Isidro E.; Wagner, Norman J.; Liu, Y. *Biophysical Journal*, *106*, 1763.
160. Vorontsova, Maria A.; Chan, Ho Y.; Lubchenko, V.; Vekilov, Peter G. *Biophysical Journal*, *109*, 1959.
161. Uzunova, V.; Pan, W.; Lubchenko, V.; Vekilov, P. G. *Faraday Discussions* **2012**, *159*, 87.
162. Vekilov, P. G. *Crystal Growth & Design* **2010**, *10*, 5007.

163. Pan, W.; Kolomeisky, A. B.; Vekilov, P. G. *J Chem Phys* **2005**, *122*, 174905.
164. Garetz, B.; Matic, J.; Myerson, A. *Phys Rev Lett* **2002**, *89*, 175501.
165. Vekilov, P. G. *Crystal Growth and Design* **2004**, *4*, 671.
166. George, A.; Wilson, W. W. *Acta Crystallographica Section D* **1994**, *50*, 361.
167. Goldberg, J. G. K. R. J. *J Chem Phys* **1950**, *18*, 54.
168. Hoyt, L. F. *Industrial & Engineering Chemistry* **1934**, *26*, 329.
169. Moutzouris, K.; Papamichael, M.; Betsis, S. C.; Stavrakas, I.; Hloupis, G.; Triantis, D. *Applied Physics B* **2013**, *116*, 617.
170. Kedenburg, S.; Vieweg, M.; Gissibl, T.; Giessen, H. *Opt Mater Express* **2012**, *2*, 1588.
171. Rheims, J.; Köser, J.; Wriedt, T. *Measurement Science and Technology* **1997**, *8*, 601.
172. Wilking, J. N.; Zaburdaev, V.; De Volder, M.; Losick, R.; Brenner, M. P.; Weitz, D. A. *Proc Natl Acad Sci USA* **2013**, *110*, 848.
173. Filipe, V.; Poole, R.; Kutscher, M.; Forier, K.; Braeckmans, K.; Jiskoot, W. *Pharm Res-Dordr* **2011**, *28*, 1112.
174. Thomas, B. R.; Chernov, A. A.; Vekilov, P. G.; Carter, D. C. *J Cryst Growth* **2000**, *211*, 149.
175. Vekilov, P. G. *Cryst Growth Des* **2007**, *7*, 2796.
176. Vekilov, P. G.; Galkin, O.; Pettitt, B. M.; Choudhury, N.; Nagel, R. L. *J Mol Biol* **2008**, *377*, 882.

177. Parodi, A.; Haddix, S. G.; Taghipour, N.; Scaria, S.; Taraballi, F.; Cevenini, A.; Yazdi, I. K.; Corbo, C.; Palomba, R.; Khaled, S. Z.; Martinez, J. O.; Brown, B. S.; Isenhardt, L.; Tasciotti, E. *ACS Nano* **2014**, *8*, 9874.
178. Xu, Q.; Boylan, N. J.; Suk, J. S.; Wang, Y.-Y.; Nance, E. A.; Yang, J.-C.; McDonnell, P. J.; Cone, R. A.; Duh, E. J.; Hanes, J. *J Control Release* **2013**, *167*, 1.
179. Kreyling, W. G.; Hirn, S.; Möller, W.; Schleh, C.; Wenk, A.; Celik, G.; Lipka, J.; Schäffler, M.; Haberl, N.; Johnston, B. D.; Sperling, R.; Schmid, G.; Simon, U.; Parak, W. J.; Semmler-Behnke, M. *ACS Nano* **2014**, *8*, 222.
180. Barua, S.; Mitragotri, S. *ACS Nano* **2013**, *7*, 9558.
181. Brangwynne, C. P.; Koenderink, G. H.; Mackintosh, F. C.; Weitz, D. A. *Trends Cell Biol* **2009**, *19*, 423.
182. Guo, Y.; Jangi, S.; Welte, M. A. *Mol Biol Cell* **2005**, *16*, 1406.
183. Barlan, K.; Rossow, M. J.; Gelfand, V. I. *Curr Opin Cell Biol* **2013**, *25*, 483.
184. Haas, B. L.; Matson, J. S.; DiRita, V. J.; Biteen, J. S. *Molecules* **2014**, *19*, 12116.
185. Perillo, E. P.; Liu, Y. L.; Huynh, K.; Liu, C.; Chou, C. K.; Hung, M. C.; Yeh, H. C.; Dunn, A. K. *Nat Commun* **2015**, *6*.
186. Rabencov, B.; Arca, J.; van Hout, R. *Int J Multiphas Flow* **2014**, *62*, 110.
187. Michalet, X.; Pinaud, F. F.; Bentolila, L. A.; Tsay, J. M.; Doose, S.; Li, J. J.; Sundaresan, G.; Wu, A. M.; Gambhir, S. S.; Weiss, S. *Science* **2005**, *307*, 538.
188. Boucher, D.; Deng, Z.; Leadbeater, T.; Langlois, R.; Renaud, M.; Waters, K. E. *Miner Eng* **2014**, *62*, 120.
189. Pace, H. E.; Rogers, N. J.; Jarolimek, C.; Coleman, V. A.; Gray, E. P.; Higgins, C. P.; Ranville, J. F. *Environ Sci Technol* **2012**, *46*, 12272.

190. Turiv, T.; Brodin, A.; Nazarenko, V. G. *Condens Matter Phys* **2015**, *18*.
191. He, K.; Spannuth, M.; Conrad, J. C.; Krishnamoorti, R. *Soft Matter* **2012**, *8*, 11933.
192. Safari, M. S.; Vorontsova, M. A.; Poling-Skutvik, R.; Vekilov, P. G.; Conrad, J. C. *Physical Review E* **2015**, *92*, 042712.
193. Saxton, M. J. *Nat Methods* **2014**, *11*, 247.
194. Chenouard, N.; Smal, I.; de Chaumont, F.; Maska, M.; Sbalzarini, I. F.; Gong, Y. H.; Cardinale, J.; Carthel, C.; Coraluppi, S.; Winter, M.; Cohen, A. R.; Godinez, W. J.; Rohr, K.; Kalaidzidis, Y.; Liang, L.; Duncan, J.; Shen, H. Y.; Xu, Y. K.; Magnusson, K. E. G.; Jalden, J.; Blau, H. M.; Paul-Gilloteaux, P.; Roudot, P.; Kervrann, C.; Waharte, F.; Tinevez, J. Y.; Shorte, S. L.; Willemsse, J.; Celler, K.; van Wezel, G. P.; Dan, H. W.; Tsai, Y. S.; de Solorzano, C. O.; Olivo-Marin, J. C.; Meijering, E. *Nat Methods* **2014**, *11*, 281.
195. Meijering, E.; Dzyubachyk, O.; Smal, I. *Method Enzymol* **2012**, *504*, 183.
196. Kural, C.; Kim, H.; Syed, S.; Goshima, G.; Gelfand, V. I.; Selvin, P. R. *Science* **2005**, *308*, 1469.
197. Liang, Z. Y.; Xu, N.; Guan, Y. H.; Xu, M.; He, Q. H.; Han, Q. D.; Zhang, Y. Y.; Zhao, X. S. *Biochem Biophys Res Commun* **2007**, *353*, 231.
198. Parthasarathy, R. *Nat Methods* **2012**, *9*, 724.
199. Parthasarathy, R. *Biophys J* **2012**, *102*, 202a.
200. Cheezum, M. K.; Walker, W. F.; Guilford, W. H. *Biophys J* **2001**, *81*, 2378.
201. Filipe, V.; Hawe, A.; Jiskoot, W. *Pharm Res-Dordr* **2010**, *27*, 796.
202. Anderson, B.; Rashid, M. H.; Carter, C.; Pasternack, G.; Rajanna, C.; Revazishvili, T.; Dean, T.; Senecal, A.; Sulakvelidze, A. *Bacteriophage* **2011**, *1*, 86.

203. Montes-Burgos, I.; Walczyk, D.; Hole, P.; Smith, J.; Lynch, I.; Dawson, K. *J Nanopart Res* **2010**, *12*, 47.
204. Vorontsova, M. A.; Chan, H. Y.; Lubchenko, V.; Vekilov, P. G. *Biophysical Journal* **2015**, *109*, 1959.
205. Vekilov, P. G.; Vorontsova, M. A. *Acta Crystallogr F-Struct Biol Cryst Commun* **2014**, *70*, 271.
206. Zhang, F.; Roosen-Runge, F.; Sauter, A.; Roth, R.; Skoda, M. W. A.; Jacobs, R. M. J.; Sztucki, M.; Schreiber, F. *Faraday Discussions* **2012**, *159*, 313.
207. Williamson, T. E.; Vitalis, A.; Crick, S. L.; Pappu, R. V. *Journal of Molecular Biology* **2010**, *396*, 1295.
208. Vekilov, P. G. *Soft Matter* **2010**, *6*, 5254.
209. Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q.-X.; Nixon, B. T.; Rosen, M. K. *Nature* **2012**, *483*, 336.
210. Brangwynne, C. P. *The Journal of Cell Biology* **2013**, *203*, 875.
211. Weber, Stephanie C.; Brangwynne, Clifford P. *Current Biology* **2015**, *25*, 641.
212. Hyman, A. A.; Weber, C. A.; Jülicher, F. *Annual Review of Cell and Developmental Biology* **2014**, *30*, 39.
213. Qian, H.; Sheetz, M. P.; Elson, E. L. *Biophys J* **1991**, *60*, 910.
214. Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. *Journal of Medicinal Chemistry* **2004**, *47*, 558.
215. Pan, W.; Filobelo, L.; Pham, N. D. Q.; Galkin, O.; Uzunova, V. V.; Vekilov, P. G. *Physical Review Letters* **2009**, *102*, 058101.

216. Boyer, D.; Dean, D. S.; Mejia-Monasterio, C.; Oshanin, G. *Phys Rev E* **2012**, *85*.
217. Malkin, A. J.; McPherson, A. *J Crystal Growth* **1993**, *128*, 1232.
218. Malkin, A. J.; McPherson, A. *Acta Crystallogr Section D* **1994**, *50*, 385.
219. Malkin, A. J.; Kuznetsov, Y. G.; Glanz, W.; McPherson, A. *J Phys Chem* **1996**, *100*, 11736.
220. Yau, S. T.; Thomas, B. R.; Galkin, O.; Gliko, O.; Vekilov, P. G. *Proteins* **2001**, *43*, 343.
221. Muschol, M.; Rosenberger, F. *J Chem Phys* **1997**, *107*, 1953.
222. McPherson, A. *Preparation and analysis of protein crystals*; Wiley: New York, 1982.
223. Chernov, A. A.; Komatsu, H. In *Science and Technology of Crystal Growth*; van der Eerden, J. P., Bruinsma, O. S. L., Eds.; Kluwer Academic: Dordrecht, 1995, p 67.
224. Chernov, A. A.; Komatsu, H. In *Science and Technology of Crystal Growth*; van der Eerden, J. P., Bruinsma, O. S. L., Eds.; Kluwer Academic: Dordrecht, 1995, p 329.
225. Vekilov, P. G. *Crystal Growth and Design*, **2007**, *7*, 2796.
226. Vekilov, P. G. *Cryst Growth Des* **2007**, *7*, 2239.
227. Vekilov, P. G. In *Methods in Molecular Biology, vol 300: Protein Nanotechnology, Protocols, Instrumentation, and Applications*; Vo-Dinh, T., Ed.; Humana Press: Totowa, NJ, 2005, p 15.
228. Vekilov, P. G. In *Methods in Enzymology volume 368: Macromolecular Crystallography, Part C*; C.W. Carter, J., R.M. Sweet, Eds.; Academic Press: San Diego, 2003, p 84.

229. Petsev, D. N.; Chen, K.; Gliko, O.; Vekilov, P. G. *Proc Natl Acad Sci USA* **2003**, *100*, 792.
230. Yau, S.-T.; Thomas, B. R.; Vekilov, P. G. *Phys Rev Lett* **2000**, *85*, 353.
231. Yau, S.-T.; Petsev, D. N.; Thomas, B. R.; Vekilov, P. G. *J Mol Biol* **2000**, *303*, 667.
232. Derewenda, Z. *Structure (Camb)* **2004**, *12*, 529.
233. Derewenda, Z. S. *Acta crystallographica Section D, Biological crystallography* **2010**, *66*, 604.
234. Derewenda, Z. S.; Vekilov, P. G. *Acta Crystallogr Section D* **2006**, *62*, 116.
235. Reviakine, I.; Georgiou, D. K.; Vekilov, P. G. *J Am Chem Soc* **2003**, *125*, 11684.
236. Yau, S.-T.; Vekilov, P. G. *Nature* **2000**, *406*, 494.
237. Gibbs, J. W. *Trans Connect Acad Sci* **1876**, *3*, 108.
238. Gibbs, J. W. *Trans Connect Acad Sci* **1878**, *3*, 343.
239. Gibbs, J. W. *The scientific papers of J.W. Gibbs. Volume One Thermodynamics*; Oxbow Press: Woodbridge, Connecticut, 1993.
240. Pan, W.; Kolomeisky, A. B.; Vekilov, P. G. *J Chem Phys* **2005**, *122*, 174905.
241. Vekilov, P. G. *Nature Nanotechnology* **2011**, *6*, 82.
242. Yau, S.-T.; Vekilov, P. G. *J Am Chem Soc* **2001**, *123*, 1080.
243. Galkin, O.; Vekilov, P. G. *J Amer Chem Soc* **2000**, *122*, 156.
244. Harding, M. M.; Hodgkin, D. C.; Kennedy, A. F.; O'Connor, A.; Weitzmann, P. D. *J Mol Biol* **1966**, *16*, 212.
245. Schlichtkrull, J. *Acta Chemica Scandinavica* **1956**, *10*, 1455.
246. Schlichtkrull, J. *Acta Chemica Scandinavica* **1956**, *10*, 1459.

247. Bergeron, L.; Filobelo, L. F.; Galkin, O.; Vekilov, P. G. *Biophysical Journal* **2003**, *85*, 3935.
248. Savage, J. R.; Dinsmore, A. D. *Physical Review Letters* **2009**, *102*, 198302.
249. Harano, K.; Homma, T.; Niimi, Y.; Koshino, M.; Suenaga, K.; Leibler, L.; Nakamura, E. *Nature Materials* **2012**, *11*, 877.
250. Lomakin, A.; Chung, D. S.; Benedek, G. B.; Kirschner, D. A.; Teplow, D. B. *Proc Natl Acad Sci USA* **1996**, *93*, 1125.
251. Krishnan, R.; Lindquist, S. L. *Nature* **2005**, *435*, 765.
252. Galkin, O.; Chen, K.; Nagel, R. L.; Hirsch, R. E.; Vekilov, P. G. *Proceedings of the National Academy of Sciences* **2002**, *99*, 8479.
253. Vivares, D.; Kaler, E. W.; Lenhoff, A. M. *Acta Crystallogr D Biol Crystallogr* **2005**, *61*, 819.
254. Vekilov, P. G. *Soft Matter* **2010**, *6*, 5254.

