PROTEIN CONDENSATION IN CROWDED AQUEOUS SOLUTIONS

by

Ying Wang

Bachelor of Science in Chemistry, 2003
Tsinghua University
Beijing, China

Submitted to the Graduate Faculty of the
College of Science and Engineering
Texas Christian University
in partial fulfillment of the requirements
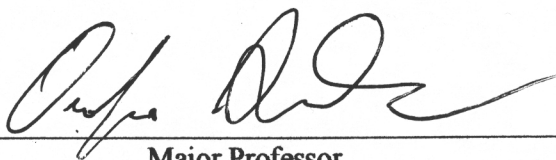for the degree of

Doctor of Philosophy

December, 2008
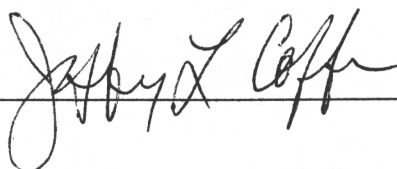
PROTEIN CONDENSATION IN CROWDED AQUEOUS SOLUTIONS

by

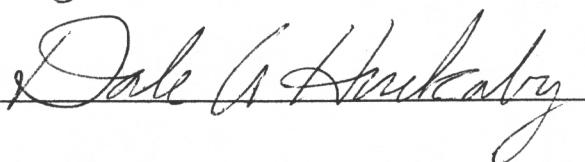Ying Wang

Dissertation approved:

_____
Major Professor

_____

_____

_____

_____
For The College of Science and Engineering

Acknowledgments

First and most, I wish to thank my advisor, Prof. Onofrio Annunziata. He has not only enlarged my knowledge of physical chemistry, but also kindled my enthusiasm at science and life. I could present this work only with his constant support and inspiring advices. I am also grateful for his patience and intensive discussions on protein phase behaviors and his help on my writing skills and English.

I want to thank Prof. John G. Albright for the educating discussions on protein diffusion.

I wish to thank Professor Bruce N. Miller from TCU physics department for the discussion and advices on Monte Carlo simulation.

I appreciate the enthusiastic courses given by Prof. Dale A. Huckaby preparing me with the fundamental knowledge of thermodynamics and statistical mechanics.

I am in debt to Prof. Jeffery L. Coffer and his student Dr. Ji Wu for their patience in instructing me to use the electron microscopes.

I am grateful to Prof. Sergei V. Dzyuba, Prof. Jean-Luc Montchamp and his student Dr. Monika I. Antczak for helping me with the derivatitation of PEG.

I wish to acknowledge ACS Petroleum Research Fund and TCU Research and Creative Activity Funds for their generous supports for this research.

Finally, I wish to thank my wife for not only doing all the cooking and cleaning but also having the chemistry conversations with me. I want to thank my parents for their love and supports over years.

TABLE OF CONTENTS

List of Figures

Chapter 10

Chapter 11

12 °C. (D) Cross-linked amorphous aggregates obtained from 10 mg/mL albumin in aqueous sodium acetate buffer, 0.1 M, pH 5.2, c_{CL} = 0.1% at 25 °C.

Figure 2. Images taken with SEM. Cross-linked amorphous aggregates obtained from 10 mg/mL albumin in aqueous sodium acetate buffer, 0.1 M, pH 5.2, c_{CL} = 0.1% at 25 °C. The length of the horizontal bars is 10 μm.

Figure 3. (A) Average radius of BSA microspheres, R, as a function of glutaraldehyde concentration, c_{CL}. The experimental conditions are those described for Figure 1. (B) Inverse of R as a function of c_{CL}. The R^{-1} data were fitted to straight lines.

Figure 4. Images taken with a light microscope using phase contrast. LLPS induced by lysozyme cross-linking at 25 °C. Cross-linked lysozyme microspheres are reported for two representative cases: (A) 10 mg/mL lysozyme in aqueous sodium acetate buffer, 0.1 M, pH 4.5, NaCl 0.5 M, c_{CL} = 0.1%, T_{ph}^{0} = -12 °C and (B) 10 mg/mL lysozyme in sodium phosphate buffer, 0.1 M, pH 7.1, NaCl 1.0 M, c_{CL} = 0.1%, T_{ph}^{0} = -1.7 °C.

Figure 5. Turbidity, τ, as a function of time, t, after cross-linking has started. This turbidity profile was obtained at 25 °C for the representative case: 10 mg/mL lysozyme in aqueous sodium acetate buffer, 0.1 M, pH 4.5, NaCl 0.5 M, c_{CL} = 0.5%.

Figure 6. Inverse of the induction time, t_{ind}, as a function of c_{CL} for four representative cases at 25 °C: 10 mg/mL lysozyme in aqueous sodium acetate buffer, 0.1 M, pH 4.5, NaCl 0.5 M, T_{ph}^{0} = -12 °C; 10 mg/mL lysozyme in

List of Tables

Chapter 1

Background

1. Introduction

Protein condensation refers to the formation of protein-rich phases such as crystals, concentrated liquid droplets, amorphous aggregates, and gels from protein solutions.[1-8] These phenomena are of biological, medical and biotechnological importance.[9-14]

One interesting application of protein condensation is preparation of protein-based materials.[15,16] Since protein molecules have distinct properties, e.g. catalytic activity, ligand binding, biocompatibility, molecular recognition, environment sensitivity, and versatility to chemical modification, they are very valuable for applications in biotechnology and material science.[17-19] Due to currently fast emerging of the fields of proteomics and protein engineering, the structures and functions of more and more proteins have become available and could be adapted for specific applications.[20,21] In order to increase the operational stability of protein-based materials and facilitate their recovery and recycling, chemical cross-linkers have been used to prepare immobilized protein-rich phases,[22-28] e.g. cross-linked enzyme crystal[24] and aggregates[25] for biocatalytical applications and cross-linked albumin microspheres for medical diagnosis and drug delivery.[26] The morphology and size of protein-rich phases depend on the chemical composition of solution generating these condensed phases.[2,29-36] However, most of the current approaches to prepare protein-based materials are developed by trial and error experiments, and the rationale for controlling morphology and size is not available.

Protein condensation also finds applications in structural and preparatory biochemistry. Production of high quality protein crystals is the bottleneck for determining

protein structure by X-ray diffraction.[2, 15] Crystallization, liquid-liquid phase separation and aggregation are widely used to purify and recover protein in industry.[37-39] Therefore, thermodynamic and kinetic studies on protein condensation are important for both biochemistry and chemical engineering. Finally, protein condensed phases are known to be involved in some diseases such as cataracts, sickle cell anemia, Alzheimer's disease, Parkinson's disease, Huntington's disease and mad cow disease.[40-46]

In order to understand protein condensation in aqueous solution and optimize the above applications, a sound scientific basis of the thermodynamic behavior of protein aqueous solutions is required. Protein aqueous solutions may undergo crystallization, liquid-liquid phase separation (droplet nucleation and spinodal decomposition), aggregation and gelation depending on the physicochemical condition of the solution such as temperature, pH, and type and concentration of ionic and nonionic additives.[2,47-60] Currently, it is not well-understood as to how a given experimental condition favors a transformation relative to another. However, it is clear that these transformations are all driven by the presence of a net attractive interaction between protein molecules.[60,61] Because the size, shape, net charge and surface chemistry of protein molecules are highly diverse, various kinds of interactions contribute to the net attractive force. Typical interactions between proteins include electrostatic interaction, hydrophobic interaction, hydrogen bonding, and Van der Waals interactions.[63-67] Substantial experimental and theoretical efforts have been put forward to understand the connection between the basic interactions amongst the molecules and the resulting condensed phases.[9, 47-61, 68, 69]

One source of thermodynamic complexity in protein aqueous solutions and their phase transformations is the presence of additives such as salts and polymers. These

additives are crucial for inducing protein condensation, because they can modulate protein-protein interactions. [2,57,58] Thus protein systems are invariably multicomponent in nature. Clearly, understanding the effect of concentration and type of additive on the thermodynamic behavior of a protein solution is important for controlling the formation of protein condensed phases. Furthermore, understanding this effect is also important for assessing the roles of additives in enzymatic activity and protein conformational changes.[70]

The chemical potential of the protein component is changed by the concentration of the additive in two ways. First, due to protein-additive net interactions, the additive can modify the thermodynamic state of the individual protein molecules. This can be described by the first derivative of protein chemical potential with respect to the additive concentration. Second, the additive can modify protein-protein net interactions, i.e., the collective behavior of the protein molecules. This can be described by the effect of the additive on the first derivative of protein chemical potential with respect to protein concentration. These two independent thermodynamic properties may be related to each other only if a microscopic model is introduced. Thus, a microscopic model may be reliable if it accurately describes both of them.

These thermodynamic aspects of protein-additive aqueous solutions have been investigated using several experimental techniques. Liquid-liquid partitioning,[71] equilibrium dialysis,[72] and ternary diffusion[73] are examples of techniques that have been used to determine protein-additive interactions. On the other hand, light scattering,[74] X-ray scattering,[75] and self-interaction chromatography[76] are examples of techniques usually used to determine the effect of additive on protein-protein interactions, mainly

through second-virial-coefficient data. The corresponding experimental results have been interpreted by using microscopic models based on preferential hydration (or binding).[72] DLVO (Derjaguin-Landau-Verwey-Overbeek) interactions,[77] depletion interactions (a.k.a. macromolecular crowding),[78] and Donnan equilibrium.[79]

Among all additives, polyethylene glycol (PEG) is a hydrophilic nonionic polymer used in many biochemical and pharmaceutical applications. Due to its mild action on the biological activity of cell components, PEG is commonly used for liquid-liquid partitioning,[80] the precipitation of biomacromolecules,[2] and the preparation of biomaterials.[81] In protein crystallography, PEG is considered the most successful precipitating agent for the production of protein crystals.[2] Due to the extensive practical use of PEG, it is of fundamental importance to understand the thermodynamic behavior of protein-PEG aqueous solutions. It is generally believed that the main mechanism of action of PEG on proteins can be described through the influence of mutual volume exclusion on the entropy of the system.[75,78] This mechanism is usually denoted using the terms: "depletion interactions"[82] or "macromolecular crowding".[83] Models based on depletion interactions have been successful in describing the effect of polymers on model colloidal suspensions especially in relation to their phase transitions.[84-86] Recently, light[56] and X-ray[87] scattering measurements have been even used to characterize both protein-protein interactions and protein-PEG interactions. The corresponding results show that depletion interaction models cannot be used to describe all protein cases.[56] Nonetheless, they have been qualitatively successful in describing the effect of PEG molecular weight on the thermodynamic behavior of protein solutions.[4,55,87] To our knowledge, the internal consistency between the experimental results on protein-PEG interactions and those on

the effect of PEG concentration on protein-protein interactions has not been quantitatively examined yet. This would be a valuable tool for verifying the validity of existing microscopic models and for developing more accurate ones. In this dissertation, we report an experimental investigation of the liquid-liquid phase separation (LLPS) of aqueous bovine serum albumin (BSA) in the presence of relatively small amounts of PEG with an average molecular weight of 1450 g/mol (PEG1450) (Chapt 8). LLPS of initially stable protein-PEG-buffer mixtures can be induced when the temperature is lowered. Since LLPS is driven by the presence of net attractive interactions between protein molecules, the location of the corresponding phase boundary in the temperature-concentration phase diagram has been measured to characterize molecular interactions. Furthermore, LLPS is metastable with respect to other phase transformations.[52,88] Thus, this phase transition is interesting not only because it competes with crystallization, aggregation, and gelation but also because it provides a distinctive kinetic route for these other processes. We experimentally characterize two thermodynamically independent properties of the phase boundary: (1) the effect of PEG1450 concentration on the LLPS temperature; (2) BSA/PEG1450 partitioning in the two liquid coexisting phases. We then use thermodynamic perturbation theory to relate the first property to the effect of PEG concentration on protein-protein interactions and the second property to protein-PEG interactions. The reliability of a depletion-interaction model is then examined for this system by investigating its ability in describing both experimental properties. Due to the general importance of LLPS of protein solutions,[9,51,52,88-92] we will also show how phase separation can be induced when the temperature of a protein-PEG-buffer solution is

either lowered or increased; i.e., two liquid-liquid phase transitions can be observed for the same mixture.

Depletion interaction in a protein-PEG-buffer ternary system can be theoretically characterized with the knowledge of microscopic parameters such as the radius of protein and gyration radius of PEG.[93-96] Current theoretical estimations of the depletion interactions have been made using the ideal-coil polymer model.[95-97] In the case of colloid (with radius of ~100nm)-polymer-solvent systems, theoretical results were in good agreement with both experiments and computer simulations.[98,99] However, this approach has failed in the case of protein-polymer-buffer solutions to quantitatively describe experimental results.[55] We will show that a large contribution of the observed discrepancy can be attributed to the dependence of depletion interactions on the polymer structure (i.e., bond length and angle, and conformational properties). This polymer-structure dependence is negligible in the colloid case because the size of colloidal particle is much larger than the bond length in a polymer coil. In the case of proteins (with a radius less than 10nm), this dependence becomes significant. Here the bond length of a polymer may no longer be negligible compared to the protein radius. In Chapt 9, we examine the dependence of depletion interactions on polymer structure in a suspension of hard spheres by Monte Carlo simulations. Using this method, we also show that depletion interactions can be affected by temperature. This is examined in Chapt 10. These investigations are crucial for improving current estimations of depletion interactions in protein-polymer-buffer systems starting from microscopic parameters such as the size of the macromolecules.

Finally, the phase behavior of protein solutions becomes even more intriguing if proteins undergo chemical association in solution. It is well-established in the polymerization field that isothermal polymerization reactions of small organic molecules can bring about phase transitions from initially homogeneous liquid mixtures.[100-107] This phenomenon is quite common in industrial polymerization processes.[108] The observed phase separation is related to the poor monomer-solvent miscibility and to the decrease of solution mixing entropy caused by polymerization. The interplay between polymerization and phase transitions is responsible for various microscopic two-phase patterns. Hence, several types of polymer-rich phases such as crystals, amorphous aggregates, microspheres, or bicontinuous gel-like networks can be produced depending on the chosen experimental conditions.[100-108] According to the type of application, one polymer-rich phase is preferred with respect to another. Thus, polymerization-induced phase transitions have been extensively studied to understand and control phase morphologies. Polymerization reactions also involve large molecules such as proteins. For example, actin[29] and sickle-cell hemoglobin[30,31] are known to undergo polymerization in living organisms. Hence, protein association in aqueous solutions has been investigated for understanding both the behavior of living systems and the formation of protein aggregates associated with various diseases.[9,29-31,107,108] The association of proteins also can be intentionally induced by chemical cross-linking, using bifunctional agents such as glutaraldehyde.[15,22-28] Protein cross-linking has found applications in biochemistry for studying protein-protein interactions and crystallography for the chemical stabilization of good-quality protein crystals.[2] Furthermore, it is also very valuable for the preparation of protein-based materials relevant to applications in biotechnology and materials

science.[22,23] For instance, in enzymology, cross-linked enzyme crystals[24] and aggregates[25] find applications relevant to petroleum products, due to their superior stability and recyclability as compared to the free enzyme. In pharmaceutics, cross-linked albumin microspheres are used for loading drugs relevant to medical diagnosis and drug delivery.[26] These applications of proteins as materials are expected to steadily increase due to protein engineering. In the absence of cross-linking agents, protein aqueous solutions are generally subject to phase transformations such as crystallization, aggregation, and gelation, depending on the experimental conditions.[2, 29-36] Currently, it is not well-understood how a given experimental condition favors one transformation relative to another. In Chapt 11, we demonstrate that, under specific experimental conditions, the addition of glutaraldehyde to homogeneous aqueous solutions can isothermally induce LLPS and the consequent formation of cross-linked protein-rich droplets. We relate this phenomenon to protein oligomerization in solution. This behavior is analogous to the phase separation induced by the polymerization of small organic molecules. Although glutaraldehyde has been extensively used to cross-link protein molecules, to our knowledge, its use in homogeneous solutions to induce LLPS has not been previously described. Understanding how protein cross-linking induces LLPS and the design of the experimental conditions that favors this process is of fundamental importance for applications of proteins in chemistry, biology, and materials science.

2. Thermodynamics of Protein Aqueous Solution

The phase behavior of protein aqueous solution can be predicted using the system's Helmholtz free energy, F, or the Gibbs free energy, G for liquid and solid phases.[4,55,60,110] The driving force for phase transitions is the difference of chemical potentials μ_i of a component i in different phases under specific conditions of temperature and composition. The chemical potentials μ_i is defined as:

$$\mu_i = (\partial F / \partial n_i)_{T,V,n_{j \neq i}} \quad \text{or} \quad \mu_i = (\partial G / \partial n_i)_{T,P,n_{j \neq i}} \tag{1}$$

where T is the temperature, V is the volume of the system, P is the pressure of system and n_i is the molar number of component i. The equilibrium conditions for two coexisting phases I and II are $\mu_i(\text{I}) = \mu_i(\text{II})$ for all components, $T(\text{I}) = T(\text{II})$ and $P(\text{I}) = P(\text{II})$.

We now observe that the total volume of protein solutions is not significantly changing upon phase transitions. We can describe their thermodynamic behavior by assuming that these systems are incompressible. We will now show that an incompressible solution of three components (two solutes and one solvent) can be equivalently treated as a "gaseous" mixture of two components only. Specifically, we consider a protein-polymer-solvent ternary system composed of n_1 moles of protein molecules, n_2 moles of polymer molecules and n_0 moles of solvent molecules. Assuming the partial molar volumes of protein, $\bar{V}_1$, and polymer, $\bar{V}_2$, do not depend on the composition at constant temperature and pressure, the partial molar volume of solvent, $\bar{V}_0$ is also constant according to the Gibbs-Duhem relation. This implies that the total volume is not changing upon replacing one protein mole with $\bar{V}_1 / \bar{V}_0$ solvent moles or one

polymer mole with $\bar{V}_2/\bar{V}_0$ solvent moles, i.e. the volume change caused by mixing is zero. Within this assumption, the total volume is an invariant when phase transition occurs. The total volume change, dV, of an open system at constant T and P is given by,

$$dV = \bar{V}_1 dn_1 + \bar{V}_2 dn_2 + \bar{V}_0 dn_0 \qquad (2)$$

and the Helmholtz free energy change, dF, of an open system at constant T and P is,

$$dF = -pdV + \mu_1 dn_1 + \mu_2 dn_2 + \mu_0 dn_0 \qquad (3)$$

If we substitute dn_0 in eq 3 with the corresponding expression obtained from eq 2, we obtain:

$$dF = -(p - \mu_0/V_0)dV + (\mu_1 - \gamma_1\mu_0)dn_1 + (\mu_2 - \gamma_2\mu_0)dn_2 \qquad (4)$$

where $\gamma_1 \equiv \bar{V}_1/\bar{V}_0$ and $\gamma_2 \equiv \bar{V}_2/\bar{V}_0$.

We then consider a system made of pure solvent with the same total volume V, containing $n_0' = \gamma_1 n_1 + \gamma_2 n_2 + n_0$ solvent moles. The free energy change of this system is expressed by:

$$dF' = -pdV - \mu_0' dn_0' = -(p - \mu_0'/V_0)dV \qquad (5)$$

where μ_0' is the chemical potential of pure solvent.

If we take the difference between eq 4 and eq 5, we obtain:

$$d(F - F') = \frac{(\mu_0 - \mu_0')dV}{V_0} + (\mu_1 - \gamma_1\mu_0)dn_1 + (\mu_2 - \gamma_2\mu_0)dn_2 \qquad (6)$$

We notice that $F - F'$ is the change of free energy of replacing a given amount of solvent molecules with the same volumetric amount of protein and polymer molecules. Because the osmotic pressure of the ternary solution is defined by $\Pi \equiv -(\mu_0 - \mu_0')/V_0$, we then obtain:

$$d\tilde{F} = -\Pi dV + \tilde{\mu}_1 dn_1 + \tilde{\mu}_2 dn_2 \qquad (7)$$

where $\tilde{F} \equiv F - F'$, $\tilde{\mu}_1 \equiv \mu_1 - \gamma_1 \mu_0$ and $\tilde{\mu}_2 \equiv \mu_2 - \gamma_2 \mu_0$. Eq 7 describes the free energy change, $d\tilde{F}$, for an effective two component system, i.e. a "compressible binary fluid", where $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are the protein and polymer effective chemical potentials in this "gaseous" mixture. Note that the osmotic pressure becomes the actual pressure of this effective fluid system.

It is convenient to introduce the reduced free energy:

$$\hat{f} \equiv \frac{\tilde{F} - n_1 \tilde{\mu}_1^0 - n_2 \tilde{\mu}_2^0}{RTV} \tag{8}$$

where $\tilde{\mu}_1^0$ and $\tilde{\mu}_2^0$ are the standard chemical potentials and R is the ideal gas constant. Because the Gibbs free energy is given by $\tilde{G} = n_1 \tilde{\mu}_1 + n_2 \tilde{\mu}_2$, the Helmholtz free energy can be expressed by $\tilde{F} = n_1 \tilde{\mu}_1 + n_2 \tilde{\mu}_2 - \Pi V$. Then the reduced free energy is given by

$$\hat{f} = \frac{[c_1(\tilde{\mu}_1 - \tilde{\mu}_1^0) + c_2(\tilde{\mu}_2 - \tilde{\mu}_2^0) - \Pi]}{RT} \tag{9}$$

where $c_1 \equiv n_1/V$ and $c_2 \equiv n_2/V$ are the molar concentrations of the protein and polymer components respectively. If we define the reduced chemical potentials of protein and polymer and osmotic pressure as:

$$\hat{\mu}_1 \equiv \frac{(\tilde{\mu}_1 - \tilde{\mu}_1^0)}{RT}$$

$$\hat{\mu}_2 \equiv \frac{(\tilde{\mu}_2 - \tilde{\mu}_2^0)}{RT} \tag{10}$$

$$\hat{\Pi} \equiv \frac{\Pi}{RT}$$

then the reduced free energy of the compressible protein-polymer fluid is given by:

$$\hat{f} = c_1 \hat{\mu}_1 + c_2 \hat{\mu}_2 - \hat{\Pi} \tag{11}$$

We note that the reduced chemical potentials of protein and polymer can be directly related to the reduced free energy by:

$$\hat{\mu}_1 = \left(\partial \hat{f}/\partial c_1\right)_{T,V,c_2} \text{ and } \hat{\mu}_2 = \left(\partial \hat{f}/\partial c_2\right)_{T,V,c_1} \quad (12)$$

We now observe what happens to the thermodynamic properties in the limiting case of an ideal gas mixture. In this case, we have: $\hat{\mu}_1 = \ln(c_1)$, $\hat{\mu}_2 = \ln(c_2)$, and the ideal-gas equation of state: $\widehat{\Pi} = c_1 + c_2$. Correspondingly, the reduced free energy for ideal gas mixture is given by

$$\hat{f} = c_1 \ln\left(\frac{c_1}{e}\right) + c_2 \ln\left(\frac{c_2}{e}\right) \quad (13)$$

This thermodynamic formalism is usually employed for protein solutions and colloidal suspensions in general. We will use this formalism throughout this dissertation.

3. Phase Behavior of Protein Solution

Protein condensed phases can be formed in solution through liquid-solid phase transition (crystallization), and liquid-liquid phase transition.[90,110,111] The reference system for proteins, colloids and simple fluids (e.g., argon) is a hard-sphere suspension. The phase diagram of this system is shown in Fig 1a. This diagram describes the presence of temperature-independent solid-fluid phase transition. The typical phase diagrams of protein systems are analogous to those of colloid systems (Fig 1c) and different from those of atomic/small molecular systems (Fig 1b) due to the relatively *short-range* nature of attraction between protein molecules.[90, 110, 111]

Atomic/small molecular systems are usually modeled by hard cores with a relatively long-range attraction. The phase diagram shows the occurrence of the gas,

liquid, and solid phases. We can also observe the existence of a critical point and a triple point (Fig 1b). At temperatures above the critical point, there is no gas-liquid equilibrium in the whole concentration range and the system is in a state of supercritical fluid. At the triple point, gas, liquid, and solid phases coexist. For protein and colloid particles, the range of attractive interaction may become comparable to or shorter than their diameter. These particles are usually modeled by hard cores with a very short-range attraction. In this case, equilibrium between protein solution (described as a supercritical fluid) and the protein crystal is also obtained. However, it is important to remark that a coexistence domain of protein-poor solution and protein-rich solution (described as a gas-liquid coexistence domain) is located within the fluid-solid coexistence domain, and the triple point does not exist in the phase diagram (Fig 1c). Thus, liquid-liquid phase separation (LLPS) of protein solution is metastable with respect to the liquid-solid transition. LLPS of a protein solutions is not only a direct tool for probing protein-protein interaction, but, due to its metastability, may also have dramatic effects on the kinetics of crystallization because of critical fluctuations and wetting of crystal nuclei.[52,88]

Interestingly, due to the *short-range* nature of protein-protein attraction, macroscopic aggregation and gelation may also occur depending on the location in the phase diagram (Fig 2). These non-equilibrium phase transformations compete with LLPS and crystallization. This leads to a usually very narrow crystallization domain for protein systems. The knowledge of phase diagrams and supersaturation conditions is also important for describing non-equilibrium phase transformation and for rationally controlling protein-condensation processes.
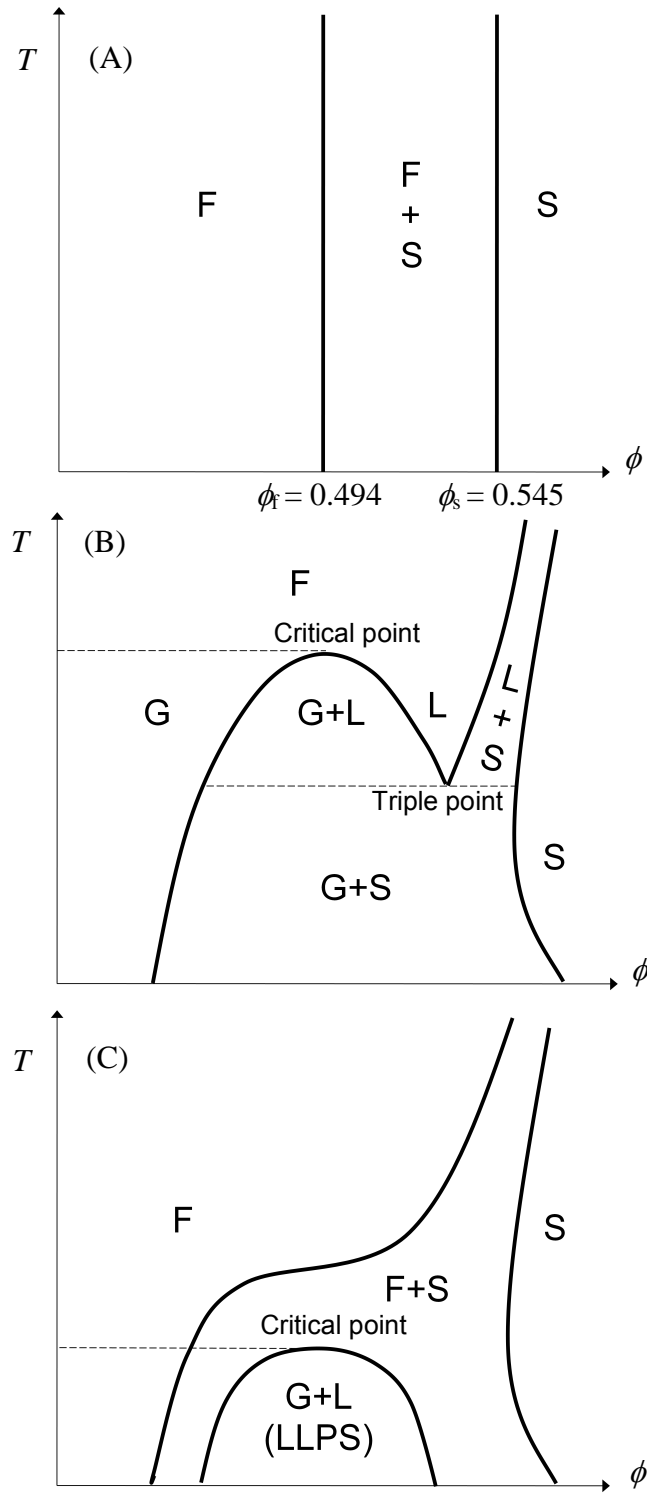
Figure 1. Phase diagram of one component system (adapted from ref 90), where T is temperature and ϕ is the particle volume fraction. (A), Hard spheres only have fluid (F) and crystal (S) phases; (B), atomic/small molecular systems modeled by hard sphere core with long range attraction may be in equilibrium between gas (G), liquid (L), and solid (S) phase; Their phase diagrams contain both critical point and triple point. (C), protein and colloid systems modeled by hard sphere core with short range attraction. In this chase, the triple point no longer exists, and the liquid-liquid coexistence described as a gas-liquid equilibrium becomes metastable with respect to liquid-crystal equilibrium.
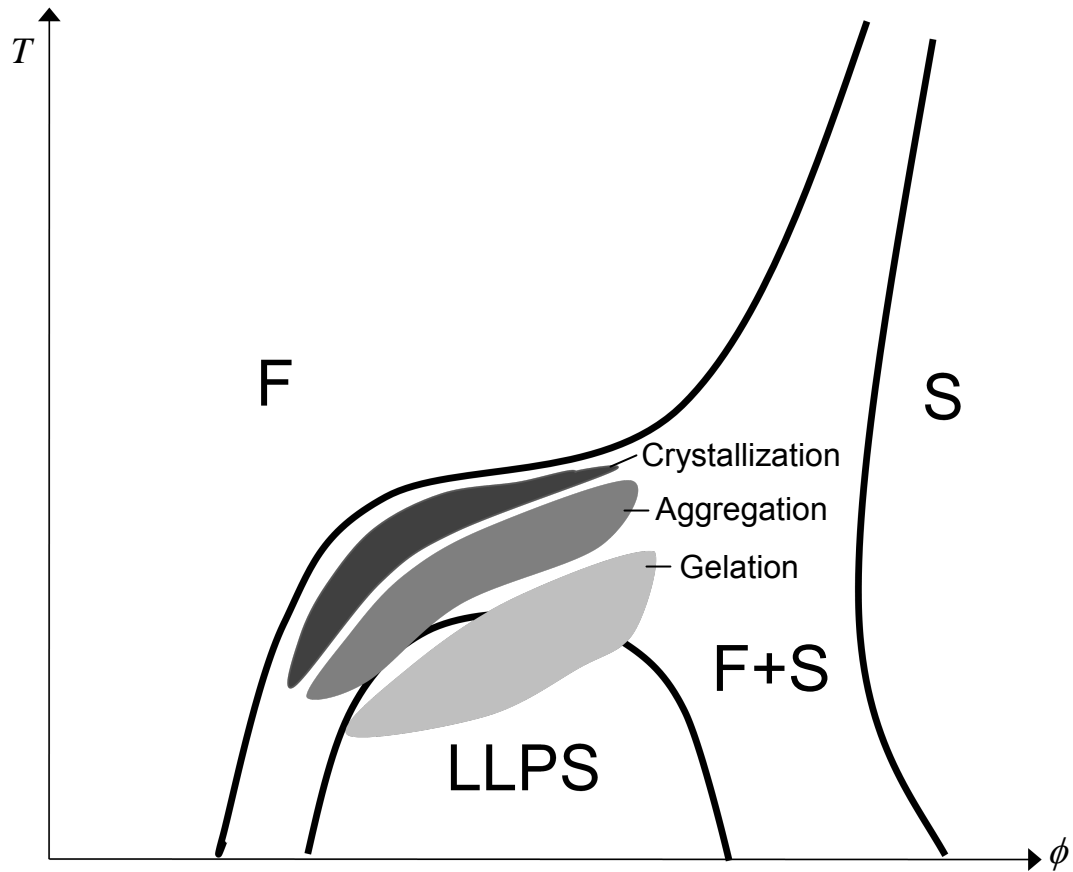
Figure 2. Phase diagram of protein solution (adapted from ref 90). Due to the short range attraction of proteins, non-equilibrium phase transformations such as aggregation and gelation may also occur in protein systems depending on the thermodynamic supersaturation. These non-equilibrium phase transformations compete with LLPS and crystallization. This leads to a usually very narrow crystallization domain for protein systems.

When there is strong *anisotropic* short-range attraction, proteins may also assemble into new ordered condensed phases.[112] For example, proteins, such as actin and sickle cell hemoglobin can form ordered condensed phases including fibrils and sheets.[5,113,114] Therefore, protein condensation is very complicated and highly diverse, and the type of observed condensed phase depends on the initial position of the system in the phase diagram.[90,111,115]

4. Protein-Protein Interactions in a Protein-Buffer system

Protein condensation is driven by attractive protein-protein interactions. From a thermodynamic point of view, protein-protein interactions are described by the protein effective gas component as described earlier (Section 2).[4, 55, 60,]

The theromodynamic of the one-component protein "fluid" can be described using the virial equation of state:

$$\widehat{\Pi} = c_1(1 + Bc_1 + Cc_1^2 + \cdots)\qquad(14)$$

where B and C are called second and third virial coefficients respectively.[116,117] If solvent is regarded as a continuum fluid, the second virial coefficient B describes two-body solvent-mediated interactions, while the third virial coefficient describes three-body solvent-mediated interactions. In statistical thermodynamics,[118] second virial coefficient for isotropic interactions is related to the mean-force potential energy $U(r)$ between two protein molecules:

$$B = 2\pi N_A \int_0^\infty \left(1 - e^{-U(r)/kT}\right) r^2 dr\qquad(15)$$

where r is the center-to-center distance between protein molecules, and N_A is the Avogadro number. If proteins are treated as hard spheres we obtain from eq 15:

$$B = 4\bar{V}_1\qquad(16)$$

where $\bar{V}_1$ is the protein partial molar volume, and σ is its diameter.

If we introduce a square-well potential (Section 7 of Chapt 8) with the range of $\lambda\sigma$ and the depth of ε, equation 16 becomes:

$$B = 4\bar{V}\left[\left(1 - e^{-\frac{\varepsilon}{kT}}\right)(\lambda^3 - 1) + 1\right]\qquad(17)$$

According to the Gibbs-Duhem equation at constant temperature, we obtain:

$$d\widehat{\Pi} = c_1 d\hat{\mu}_1 \tag{18}$$

We can therefore write:

$$c_1 \frac{\partial \hat{\mu}_1}{\partial c_1} = \frac{\partial \widehat{\Pi}}{\partial c_1} = 1 + 2Bc_1 + 3Cc_1^2 + \cdots \tag{19}$$

Thus, the protein chemical potential can then be expressed by:

$$\hat{\mu}_1 = \ln(c_1) + 2Bc_1 + 3Cc_1^2 + \cdots \tag{20}$$

5. Theoretical LLPS Phase Boundary

Protein-Protein net attractive interactions are responsible for LLPS. Protein chemical potential is used to theoretically determine the LLPS phase boundary.[110, 119] The equilibrium conditions for LLPS of protein-buffer binary system are:

$$\hat{\mu}_1(c_1^{\mathrm{I}}) = \hat{\mu}_1(c_1^{\mathrm{II}}) \tag{21}$$

$$\widehat{\Pi}(c_1^{\mathrm{I}}) = \widehat{\Pi}(c_1^{\mathrm{II}}) \tag{22}$$

Using the following thermodynamic relations:

$$\widehat{\Pi} = c_1 \hat{\mu}_1 - \hat{f} \tag{23}$$

$$\hat{\mu}_1 = \left(\frac{\partial \hat{f}}{\partial c_1}\right)_{T,V} \tag{24}$$

we obtain the following equivalent expression for the equilibrium condition expressed by eq 22:

$$(c_1^{\mathrm{II}} - c_1^{\mathrm{I}})\hat{\mu}_1(c_1^{\mathrm{I}}) = \int_{c_1^{\mathrm{I}}}^{c_1^{\mathrm{II}}} \hat{\mu}_1(c_1)dc_1 \tag{25}$$

where c_1^{I} and c_1^{II} stand for protein concentrations in phase I and II. Eq 18 is equivalent to the Maxwell equal-area rule.[110]

Eqs 21 and 25 can be used to reconstruct the LLPS phase boundary from an expression of $\hat{\mu}_1$ obtained by either fitting chemical potential data obtained applying the Monte Carlo method or by thermodynamic perturbation theory.[110] A typical plot of $\hat{\mu}_1$ as a function of c_1 is shown in Fig. 3 at different temperatures. The values of c_1 and T at critical point, c_1^c and T^c, can be obtained using the conditions:

$$\partial\hat{\mu}_1/\partial c_1 = 0 \quad \text{and} \quad \partial^2\hat{\mu}_1/\partial c_1{}^2 = 0 \tag{26}$$

Below the critical temperature, the $\hat{\mu}_1$-c_1 plots have a region with negative slope, $\partial\hat{\mu}_1/\partial c_1 < 0$ (Fig 3). This region is a thermodynamically unstable domain which is delimited by the spinodal boundary, where $\partial\hat{\mu}_1/\partial c_1 = 0$.
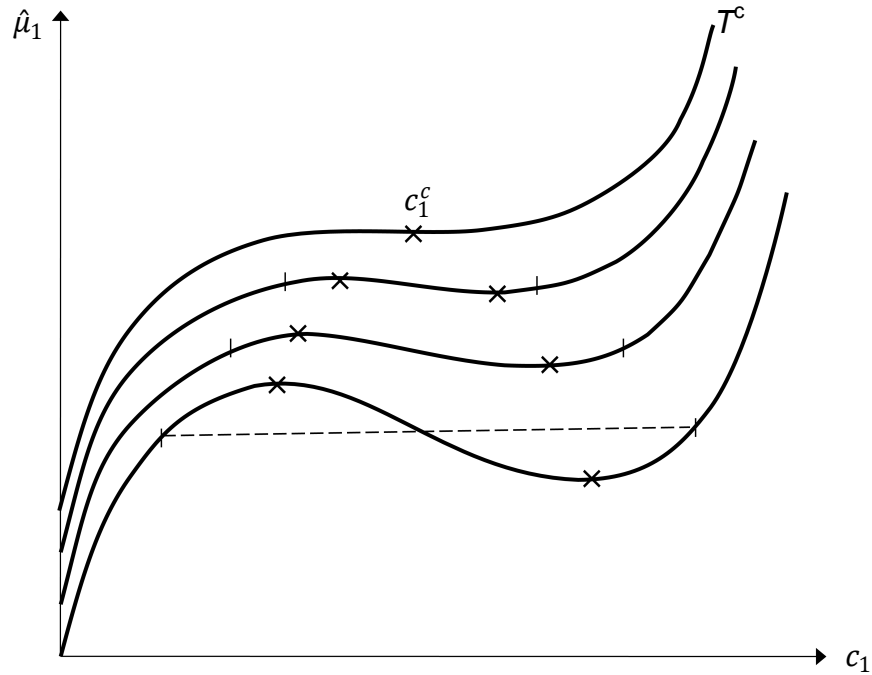
Figure 3. Typical isothermal plots of $\hat{\mu}_1$ as a function of c_1 (adapted from ref 110). Temperature values for the four curves decrease from top to bottom. The crosses denote the points on the spinodal boundary, and the vertical bar denotes the points on the LLPS boundary. The horizontal dashed line illustrates the Maxwell equal-area rule.

6. Macromolecular Crowding

We now remark that LLPS of protein-buffer solutions has been observed only for few protein cases (e.g., lysozyme and eye-lens crystallins).[52,40,44,66,120] There are two reasons which may explain why this phase transition has not been observed in many protein cases. First, the LLPS boundary may be theoretically located at temperature values outside the accessible temperature domain, where the solution freezes. Second, LLPS may require relatively high protein concentrations. At these concentrations, the kinetics of other phase transformations may be too fast and LLPS cannot be observed. However, LLPS may be observed for a wider range of protein cases by increasing the isotropic attraction between the protein molecules.[4,55] One known example of isotropic attraction is the depletion attraction, which has been investigated in the case of colloid-polymer suspensions.[90,93,94,111,121-123] It is believed that depletion attractions can be also brought about by the addition of polymers to the protein-buffer system.[4,55] The use of polymers to increase protein-protein attraction thereby inducing protein condensation is also known as macromolecular crowding.[83] One widely used polymer for inducing protein condensation is indeed PEG.[2,4,55,82]

$$HO \left[\quad \quad O \right]_n H$$
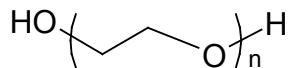
A depletion interaction is an entropic force. This is generated by the steric exclusion between two different types of particles. [93, 94] In Fig 4, we can appreciate that the mass centers of polymers are excluded from a shell region called depletion layer (gray layers) surrounding the two spherical particles (black circles) representing the protein molecules.

In order to maximize the entropy of system, the depletion layers tend to overlap to release more free volume for polymer.
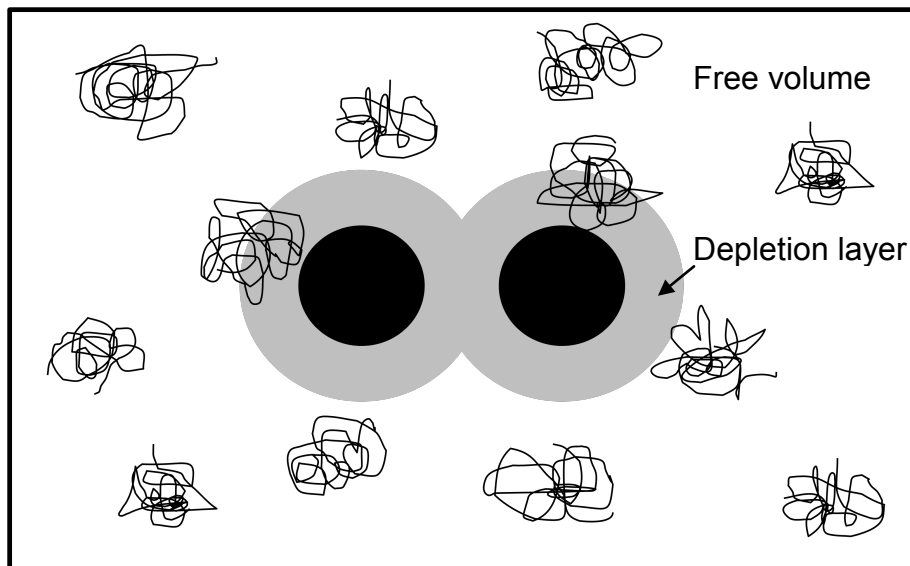
Figure 4. Depletion attraction between spherical particles induced by polymers. The solid black circles represent two globular protein molecules. The coils represent polymers. The gray shell areas represent the depletion layers.

The condensation of compact globular particles in the presence of polymer coils was first described by Asakura and Oosawa in 1954.[84] Asakura and Oosawa[93] first and Vrij[94] later have treated the compact particles as hard spheres and the coil polymers as interpenetrable spheres with radius, R_{AO}. We shall call these model coils AO spheres. We note that their theory assumes that steric interactions occur between one hard sphere and one AO sphere, however no steric interaction occurs between two AO spheres. This mimics the behavior of ideal polymer chains. We will now discuss the depletion potential energy between two particles in the presence of AO spheres. We first consider the simplest case for which the particles are two parallel plates located inside a box filled of AO spheres (now represented as black circles in Fig 5). If the AO spheres are small, they

can move into the inner region between two plates freely (Fig 5a). In this case, the osmotic pressures on the inner sides and outer sides of the plates are equal to each other. However, if the AO spheres are large, they cannot access the inner region between the plates when the distance between the plates is smaller than the diameter of the AO spheres (Fig 5b). Therefore, the osmotic pressure on the outer sides of the plates is higher and cannot be balanced. Thus, the plates are subjected to a force:

$$F(r) = \begin{cases} -\Pi A & 0 \le r < 2R_{AO} \\ 0 & r > 2R_{AO} \end{cases} \tag{27}$$

where Π is the osmotic pressure, A is the area of the plate, and r is the inter-plate distance. The negative sign describes the attractive nature of the interaction. It is important to observe that the AO-sphere concentration, which is proportional to the osmotic pressure, affects the magnitude of the force. On the other hand, the size of the AO spheres affects the range of the force.

If the plates are replaced by two hard spheres, these two particles will be subjected to a similar force (Fig 6). If we take into account the geometric properties of the depletion layer, this force can be calculated:

$$F(r) = \begin{cases} -\left(\frac{\pi}{4}\right)\Pi[4(R + R_{AO})^2 - r^2] & 2R \le r < 2(R + R_{AO}) \\ 0 & r > 2(R + R_{AO}) \end{cases} \tag{28}$$

where R is the diameter of the hard spheres.

According to the definition of potential energy $U(r) \equiv -\int_0^\infty F(r)\mathrm{d}r$, the depletion interaction potential between two spherical particles can be obtained by integrating eq 28:

$$U(r) = \begin{cases} \infty & r < 2R \tag{29} \\ -\left(\frac{\pi}{12}\right)\Pi[16(R + R_{AO})^3 - 12(R + R_{AO})^2 r + r^3] & 2R \le r < 2(R + R_{AO}) \\ 0 & r > 2(R + R_{AO}) \end{cases}$$
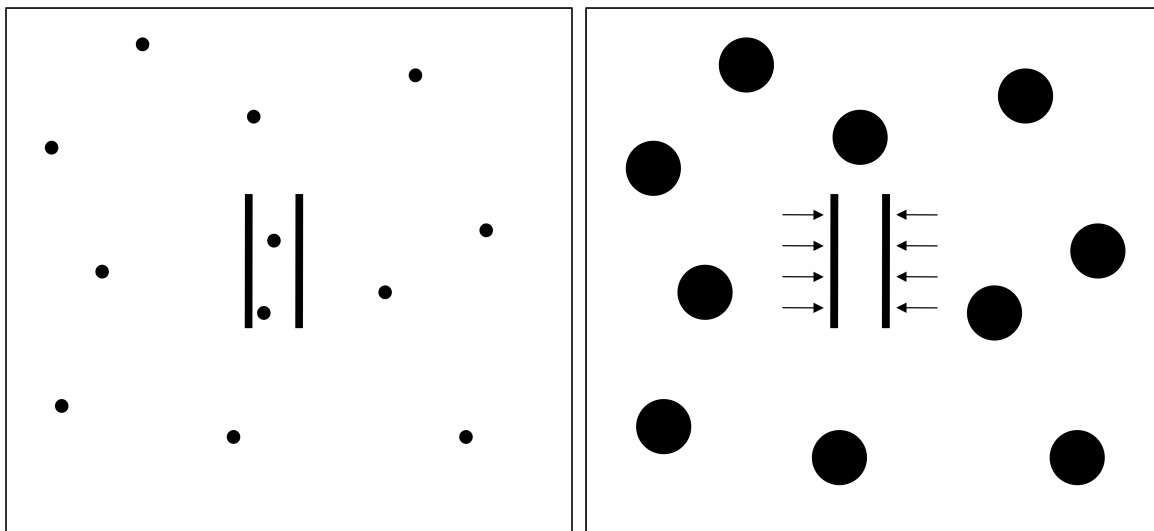
Figure 5. Depletion attraction between two parallel plates in the presence of AO spheres (black circles). (A) If the AO spheres are small, there is no significant depletion effect. (B) If the AO spheres are large, there is an effective attraction between the plates.
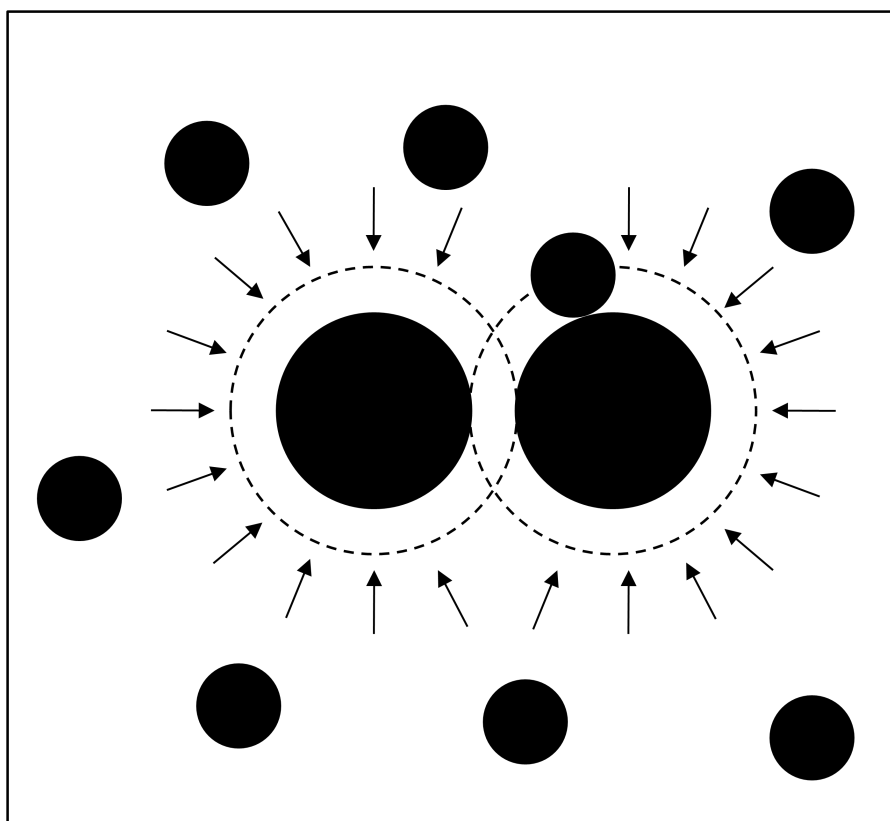
Figure 6. Depletion attraction between two hard spheres in the presence of AO spheres. The dashed circles identify the depletion layers.

In Fig 6, we can appreciate that the depletion interaction arises from the unbalanced osmotic pressure due to the overlapping of the depletion layers, which have a thickness equal to the AO sphere radius, R_{AO}. We can rewrite eq 28 by introducing the overlapping volume of depletion layer, $V_{overlap}$ and the normalized thickness of the depletion layer, $q = R_{AO}/R$:[95]

$$U(r) = \begin{cases} \infty & r < 2R \\ -\Pi V_{overlap} & 2R \leq r < 2(1+q)R \\ 0 & r > 2(1+q)R \end{cases} \quad (29)$$

$$V_{overlap} = \frac{4\pi(1+q)^3 R^3}{3}\left(1 - \frac{3r}{4(1+q)R} + \frac{r^3}{16(1+q)^3 R^3}\right)$$

The osmotic pressure generated by AO spheres is given by $\Pi = c^e RT$, where c^e is the effective molar concentration of AO-spheres in the free volume. This concentration is $c^e = c/\alpha$, where c is the overall concentration of AO-spheres and α is the free-volume fraction for AO spheres in the presence of hard spheres. We note that the depletion-interaction potential is proportional to temperature. This is consistent with the depletion potential being entropic in nature.

It is important to observe that actual polymer chains can be only approximately described as AO spheres, because we cannot neglect their conformational entropy.[94,96,124,125] Thus the depletion layer around hard spheres defined for AO spheres has a non negligible coil concentration.

The density of segments of the polymer coil, $\rho(x)$, where $x \equiv r/R$, at a distance r from the center of a hard sphere has been independently studied by Eisenriegler *et al.*,[124] and Taniguchi *et al.*[125] They have obtained the theoretical expression of segment density

profile $\rho(x)/\rho$ for an ideal chain (see Appendix B for details), where ρ is the segment density in the bulk phase. A representative density profile $\rho(x)/\rho$ is shown in Fig 7.
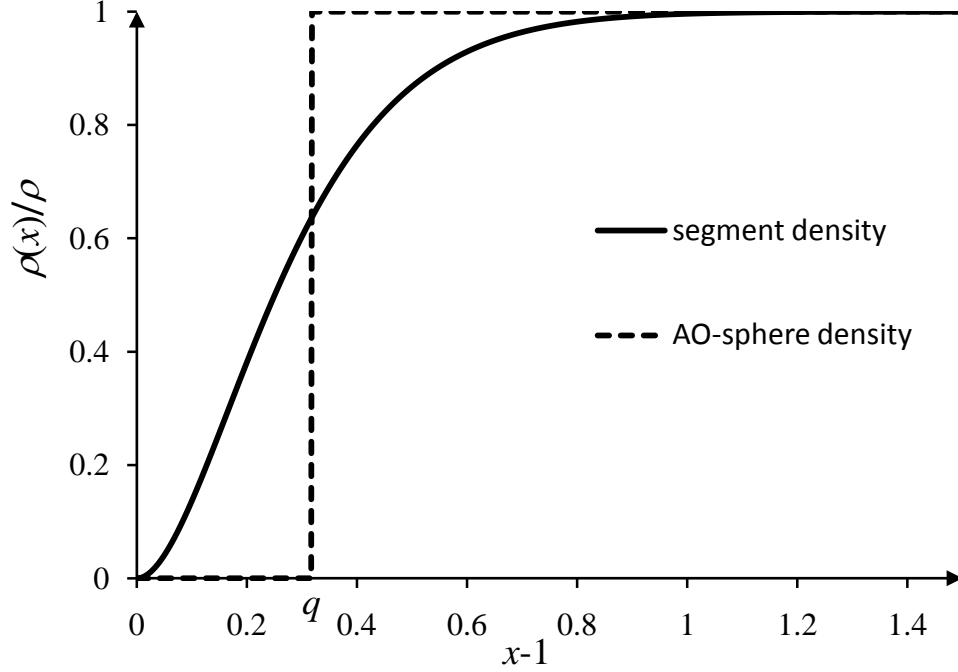
Figure 7. Density of segment of polymer coil as a function of the normalized distance $x \equiv r/R$ from the surface of a hard sphere. The dashed line indicates density profile of an equivalent AO-sphere.

Integration of the $\rho(x)/\rho$ profile can be used to define the thickness of depletion layer corresponding to an equivalent AO-sphere (Fig 7). The polymer coil can be consider as an equivalent AO-sphere with a radius of $R_{AO} = qR$, where q is the reduced thickness of depletion layer described above. The value of q for the effective AO-sphere is given by:

$$\int_{1}^{\infty} 4\pi x^2 [1 - \rho(x)/\rho]\mathrm{d}x = \int_{1}^{q} 4\pi x^2 \mathrm{d}x \tag{30}$$

Eq 30 is rewritten as

$$q = \sqrt[3]{1 + 3\int_{1}^{\infty} x^2 [1 - \rho(x)/\rho]\mathrm{d}x} - 1 \tag{31}$$

For an ideal chain, the spatial coordinates of its segments are changed in the presence of hard spheres. This is thermodynamically unfavorable because it decreases the conformational entropy of the coil. To maximize the conformational entropy, the spherical particles will get close to each other. This is equivalent to an attractive interaction between spherical particles.

7. Phase Behavior of Protein Oligomers

For small organic molecules, phase transitions have been commonly observed during the isothermal polymerization reactions in poor solvents.[100-107] This phenomenon can be attributed to the corresponding decrease of mixing entropy.

In order to illustrate the oligomerization effect on ΔS, we consider a simple model system consisting of two indistinguishable spheres (monomer) in a box, which consists of four cells. Each cell can only be filled with one sphere (Fig 8). The box is first considered as one phase, and the concentration is 0.5 spheres per cell. The number of microscopic states of n spheres, i.e. the number of ways to fill n spheres into m cells, is equal to the combinatorial $C_m^n = \binom{m}{n}$. In our case, $n=2$ and $m=4$. Hence, we obtain $C_4^2 = 6$ (see left side of Fig 9). We then consider the case in which the system separates into two phases: box 1 and 2 on the right side of Fig 8. We assume that box 1 is the dilute phase with zero concentration of spheres per cell, and box 2 is the concentrated phase with concentration of one sphere per cell. In this condition, the number of microscopic states of the two
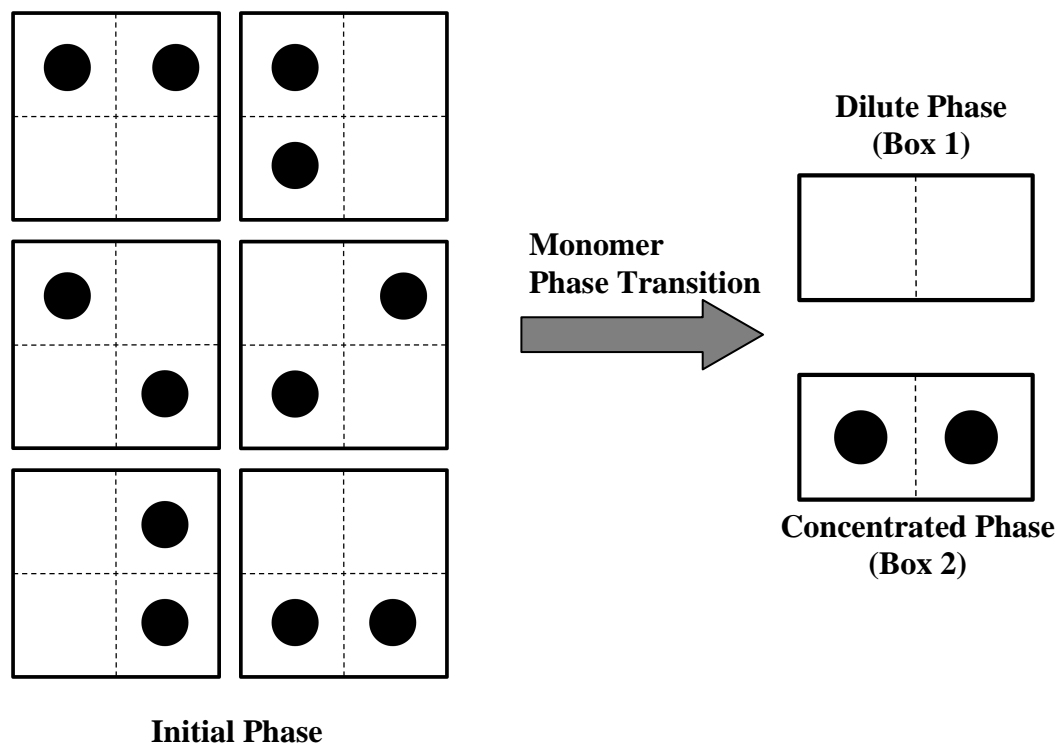
Figure 8. Number of microscopic states for a model system consisting of two monomers (black spheres) in a volume of four cells before (left) and after phase separation (right). The solid circles represent the monomers.
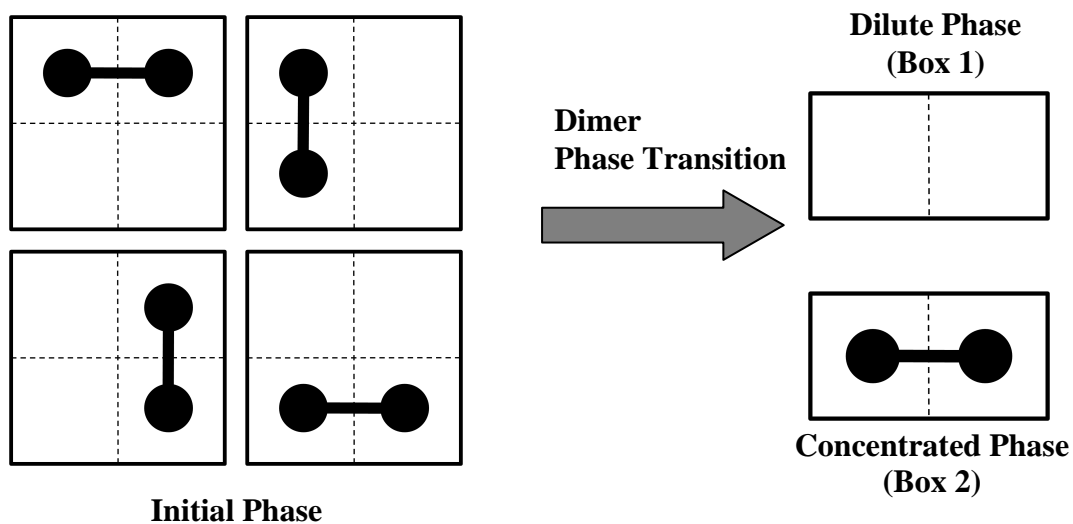
Figure 9. Number of microscopic states for a model system consisting of one dimer (two black spheres connected by a solid line) in a volume of four cells before (left) and after (right) separation.

spheres in the two phases is equals to one (see right side of Fig 8). Thus, for the sphere

(monomer), the change of entropy upon phase separation of this system is:

$$\Delta S = k\ln(1) - k\ln(6) = -k\ln(6) \tag{32}$$

We now consider the system after dimerization has occurred. We assume that only

the spheres in neighboring cells can bind, i.e. the dimer cannot exist at the diagonal

positions. Before phase separation, the number of microstates for the dimer of spheres is

equal to four (see the left side of Fig 9). After phase separation, the number of

microscopic states is equal to one (see the right side of Fig 9). Thus, the change of

entropy upon phase separation for dimer is:

$$\Delta S = k\ln(1) - k\ln(4) = -k\ln(4) \tag{33}$$

Therefore, for this model system, the decrease of entropy upon phase transition in the

case of dimer is smaller than that for the monomer.

Theoretical studies on phase equilibria for chain molecules have been conducted by

Chapman et al.[126] based on Wertheim's first-order thermodynamic perturbation

theory.[127-131] In their work, a n-mer chain molecules was modeled by n connected hard

spheres. A mean-field attractive interaction, which is responsible for phase transition, was

considered to occur between the hard-sphere monomers. In this approach, the

compressibility of the n-mer chain fluid was expressed as

$$Z_n \equiv \frac{\Pi}{c_n RT} = nZ_1 + Z_{chain} \tag{34}$$

where c_n is the molar concentration of the n-chain, Z_1 is the monomer compressibility, i.e.

the reference system. In their theory, Z_1 consists of two parts:

$$Z_1 = Z_{HS} + Z_{attr} \tag{35}$$

where Z_{HS} is hard-sphere compressibility and Z_{attr} is related to the presence of an attractive interaction energy.

In eq 34, Z_{chain} is the additional contribution to compressibility associated with the chain formation. It can be also written as a sum of two terms: the contribution, Z_{HS}^{bond}, of hard sphere connectivity and the deviation, Z_{attr}^{bond}, of monomer-monomer attractive energy due to sphere connectivity:

$$Z_{chain} = Z_{HS}^{bond} + Z_{attr}^{bond} \tag{36}$$

Chapman et al.[126] calculated Z_{HS} using the Carnahan-Starling equation of state for hard spheres:

$$Z_{HS} = \frac{1 + \phi + \phi^2 - \phi^3}{(1-\phi)^3} \tag{37}$$

and approximated Z_{attr} using a mean-field attraction energy: $Z_{attr} = -\varepsilon_{MF}\phi/kT$ where ϕ is the volume fraction of the chains, and ε_{MF} is the strength of the mean-field potential energy. The contribution Z_{chain} was calculated using first-order thermodynamic perturbation theory as described in Appendix F. The corresponding Helmholtz free energy of the n-mer chain fluid is obtained using:

$$\hat{f}_n = c_n \ln(c_n/e) + c_n \int_0^{c_n} \frac{Z_n - 1}{c_n} \, dc_n \tag{38}$$

Using this free energy expression, Chapman et al.[126] have investigated the phase coexistence curve with several values of n (Fig 10). As we can see in Fig 10, the phase separation temperature at constant volume fraction increases with n.

Based on Wertheim and Chapman's theoretical framework, Banaszak *et al.*[132] has extended this study to the chain fluids with sticky-sphere and square-well potentials. Similar dependence of phase transition temperature on n was observed. However, the

critical volume fraction is observed to increase as the range of interaction decreases. In conclusion, all these theoretical studies show that the phase-transition temperature increases with the degree of oligomerization.
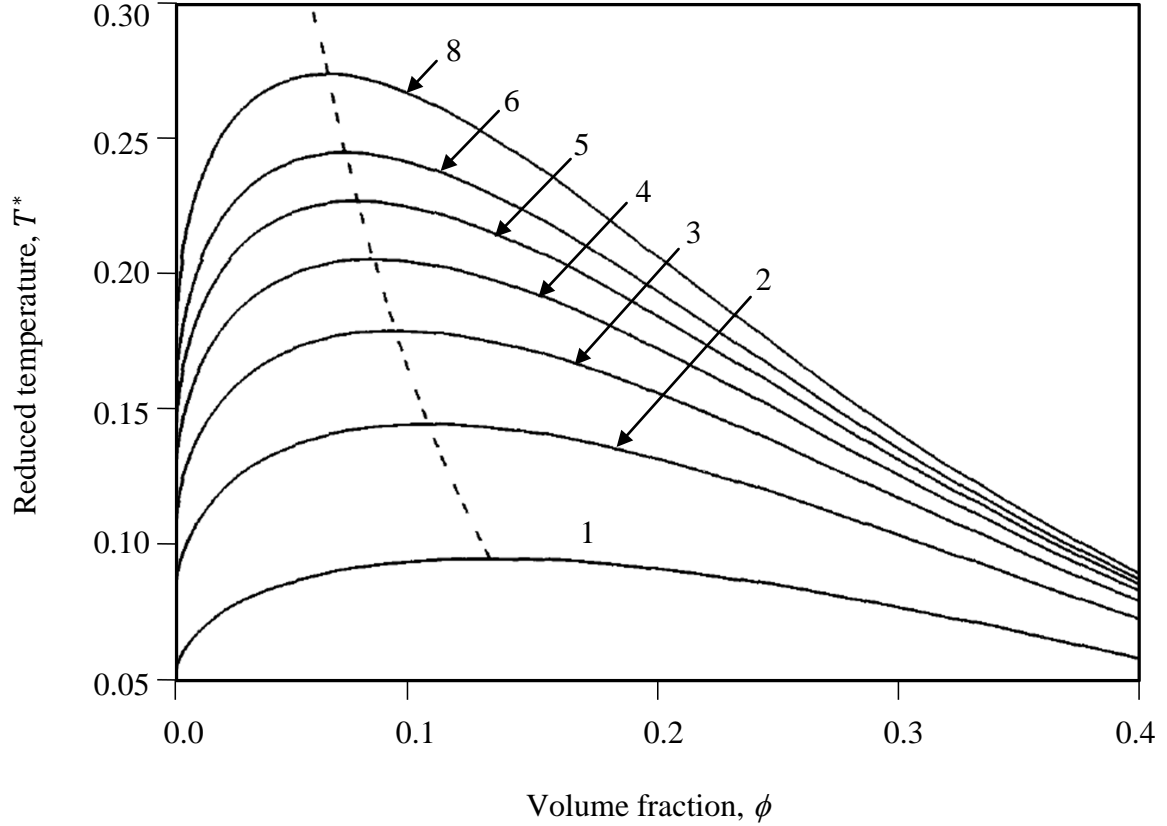
Figure 10. The reduced liquid-vapor transition temperatures, $T^* \equiv kT/\varepsilon_{MF}$, for n-mer chain molecular fluid as a function of the chain volume fraction, ϕ. ε_{MF} is strength of the mean-field potential energy. The curves are labeled with the number of spheres n in the chain. The dashed curve indicates the location of critical temperatures, T_c as a function of n. This figure is adapted from Chapman et al., Molecular physics, **1988**, *65*, 5, 1057-1079.

The increase of LLPS temperature with the degree of oligomerization has been experimentally observed for protein solution. Pande et al.[133] have found that LLPS temperature of γB-crystallin under oxidative conditions increases with time. This is attributed to the formation of loosely associated oligomers through intermolecular disulfide bonds. Later, Asherie et al.[134] have systematically investigated the LLPS of γD-crystallin and their oligomers. They have used bismaleimidohexane to link the thiol group on the surface of protein and produce well-defined dimers, trimers and mixture of higher *n*-mers. These oligomers were then separated and purified using size exclusion chromatography. They found that the LLPS temperatures for these well-defined oligomers increase with the degree of oligomerization (Fig 11). Finally, Annunziata *et. al.*[109] have shown that native βB1-crystallin undergoes reversible oligomerization in aqueous solution but not LLPS. Interestingly, if βB1-crystallin is truncated to its N-terminal (to mimic age-related degradation), reversible oligomerization is enhanced and gelation driven by LLPS is observed. All these studies provide the experimental evidence for the increase of phase separation temperature for protein solutions upon protein reversible and irreversible oligomerization.
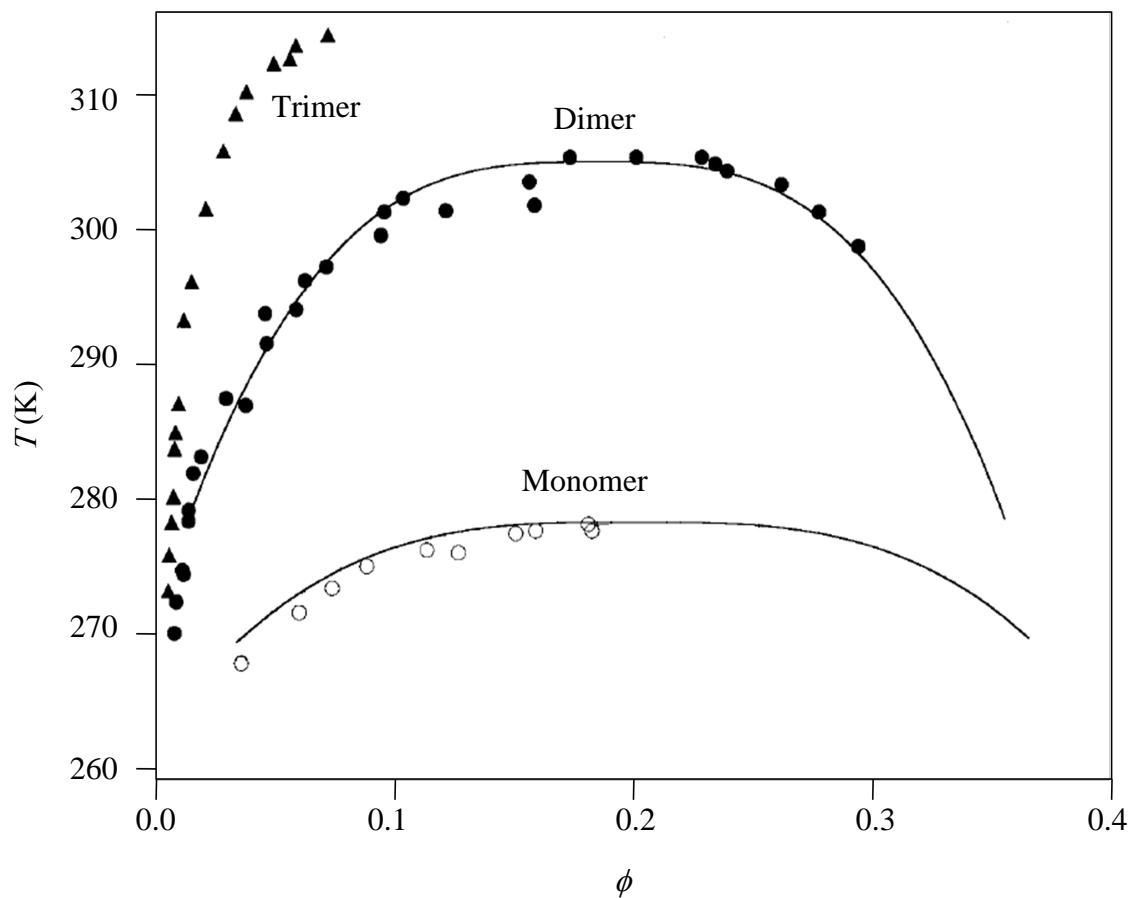
Figure 11. LLPS boundaries for γD-crystallin in 0.1M sodium phosphate (pH 7.1). The open circles denote the monomer data, the solid circles denote the dimer data, and the solid triangles denote the trimer data. The solid curves are the fits to the experimental data using the equation $|(\phi - \phi_c)/\phi_c| = A(1 - T/T_c)^\beta$, where A is fitting parameter , ϕ_c and T_c are critical volume fraction and temperature, β =0.325 is the critical exponent for the three-dimensional Ising model. This figure is adapted from Asherie et al., Biophysical Chemistry, **1998**, *75*, 213-227.

8. Protein Cross-linking

Protein oligomerization may naturally occur through chemical bonds between amino-acid residues such as disulfide bonding of thiol groups in cysteines.[135-138] It can also be induced by adding chemical cross-linkers.[15,22-28] There is a large commercial selection of protein cross-linkers. They react with specific functional groups (primary amine group, thiol group, carboxylic group, etc.) on protein surface.[140, 141] Several commonly used protein cross-linking approaches are shown as below.

_N_-Hydroxysuccinimide-esters (NHS-esters)[142]

| Primary amine | NHS ester | | Amide bond | NHS |

Imidoesters[143]

| Primary amine | Imidoester | | Amidine linkage | methanol |

Maleimide[144]

Thiol   Maleimide   Thioether bond

Carbodiimide[145]

Carboxylate   Carbodiimide   O-acylisourea

Step 1

O-acylisourea   Primary amine   Amide bond   Urea derivative
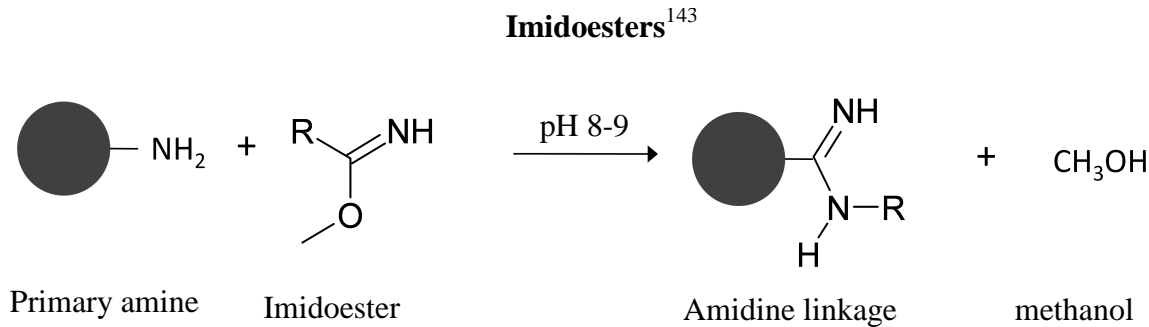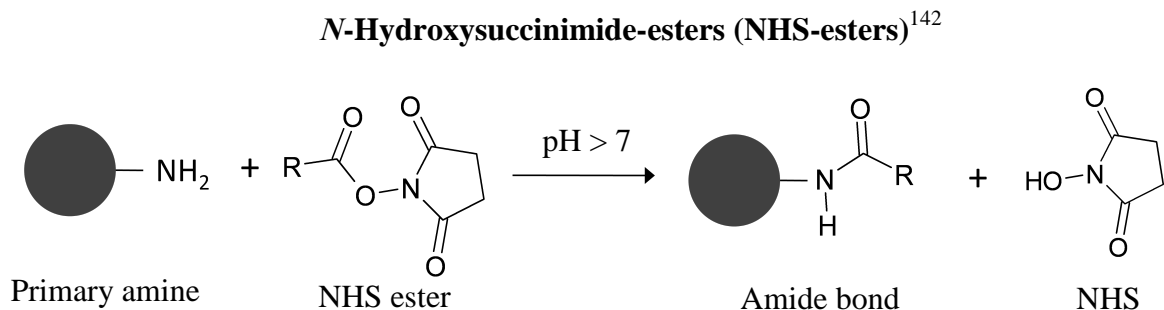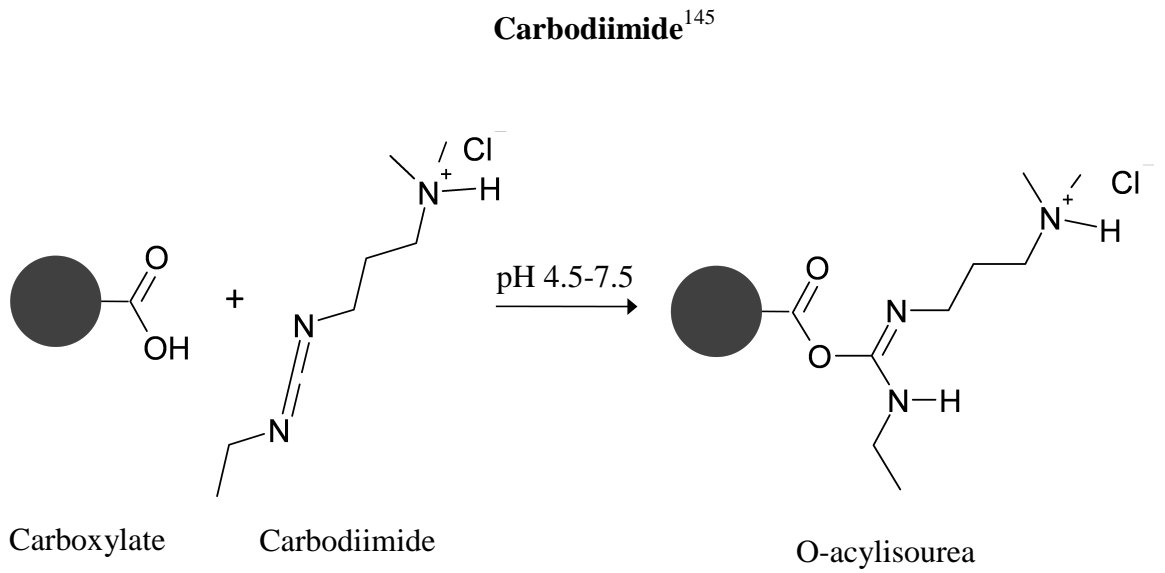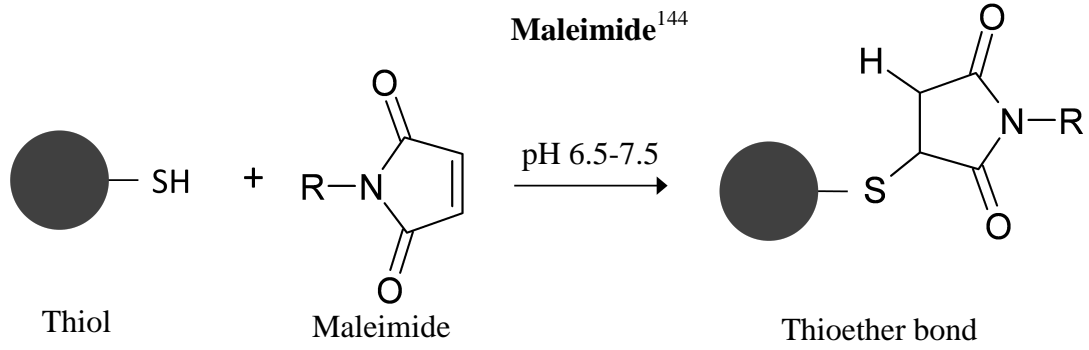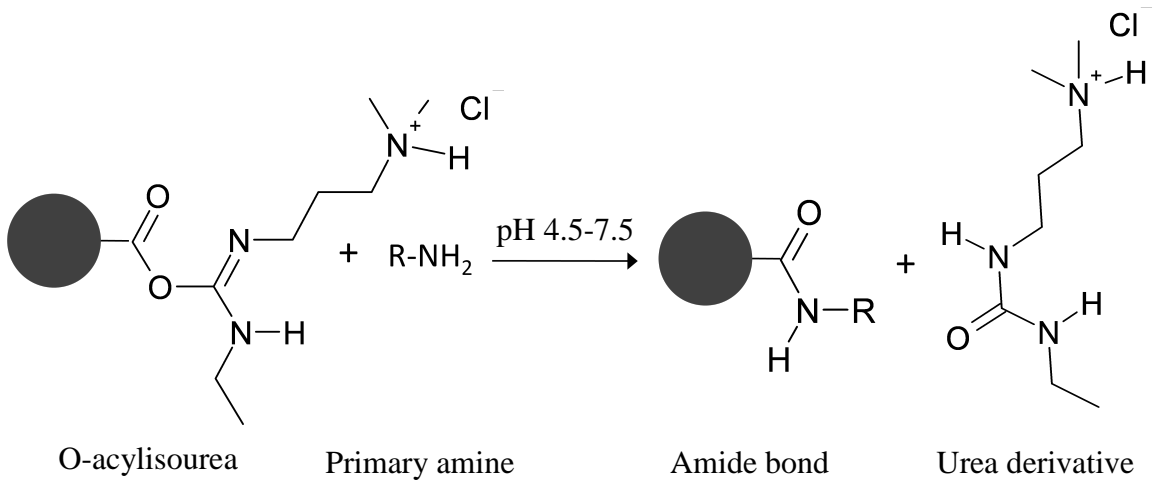
Step 2

The currently most widely used protein cross-linker is glutaraldehyde: [15, 22-28, 146]
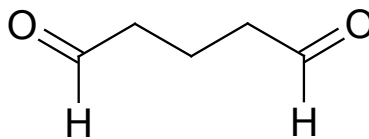
Glutaraldehyde effectively cross-links many cases of free protein molecules in aqueous solution as well as protein crystals and aggregates over a wide pH range (≥ 3).[146] X-ray crystallographic studies showed that cross-linking by glutaraldehyde has minor effect on protein structures.[147-150]

Glutaraldehyde is soluble in water, and reacts with the ε-amino group of lysine on protein. At high pH, it may also react with some other amino group.[146] The reactivity of glutaraldehyde increases with pH up to pH = 9, and the reactivity ranking for different amino acids is as follows:

ε-amino > α-amino >> guanidinyl > secondary amino > hydroxyl group

Practically in protein aqueous solution, gluraldehyde does not react with guanidinyl groups (arginine) and those with lower activity. The linkages between glutaraldehyde and lysine residues on proteins are complicated involving several mechanisms instead of simple Schiff base between aldehyde and amino group. The protein cross-linking by glutaraldehyde is irreversible even under the acidic condition, while typical Schiff bases are hydrolyzed in the acidic solution.[146] Glutaraldehyde in aqueous solution undergoes cyclization and polymerization (Fig 12).[146] The reaction between α,β-unsaturated aldehydes and amino groups on protein form robust cross-linking linkages.[146] Monsan et al.[151] proposed a mechanism of addition on the aldehydic part of the α,β-unsaturated aldehydes, which gives a Schiff base stabilized by conjugation (Fig 13(1)). Richards and Knowles[152] has proposed a slightly different mechanism of addition to ethylenic double

bonds (Fig 13(2)). However, Hardy et al.[153, 154] and Lubig et al.[155] have proposed a different mechanism. They showed that two glutaraldehyde monomers react with two protein molecules to form a pyridinium compound (Fig 14). Therefore, the mechanism of protein cross-linking by glutaraldhyde is complicated and still not well understood. Reaction between glutaraldehyde and amine groups generates yellow chromophores. The resulting solutions also have a strong UV absorbance at 260-280 nm.
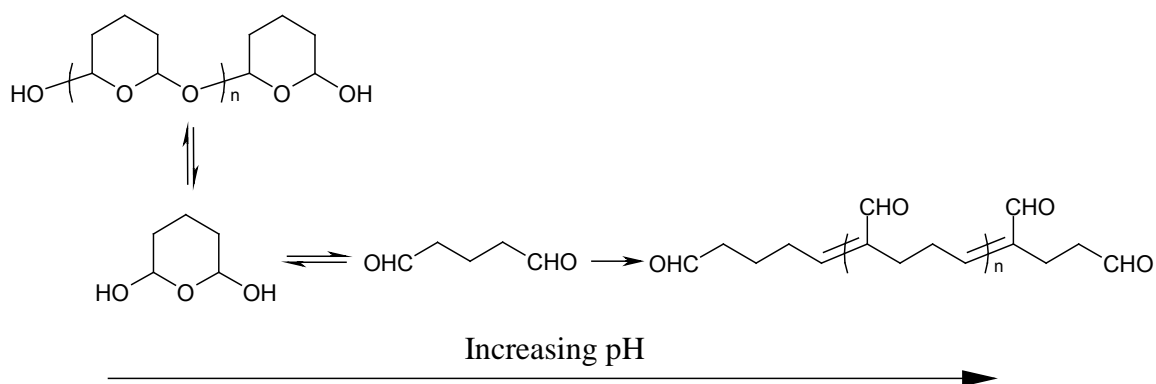
Figure 12. Glutaraldehyde in aqueous solution.
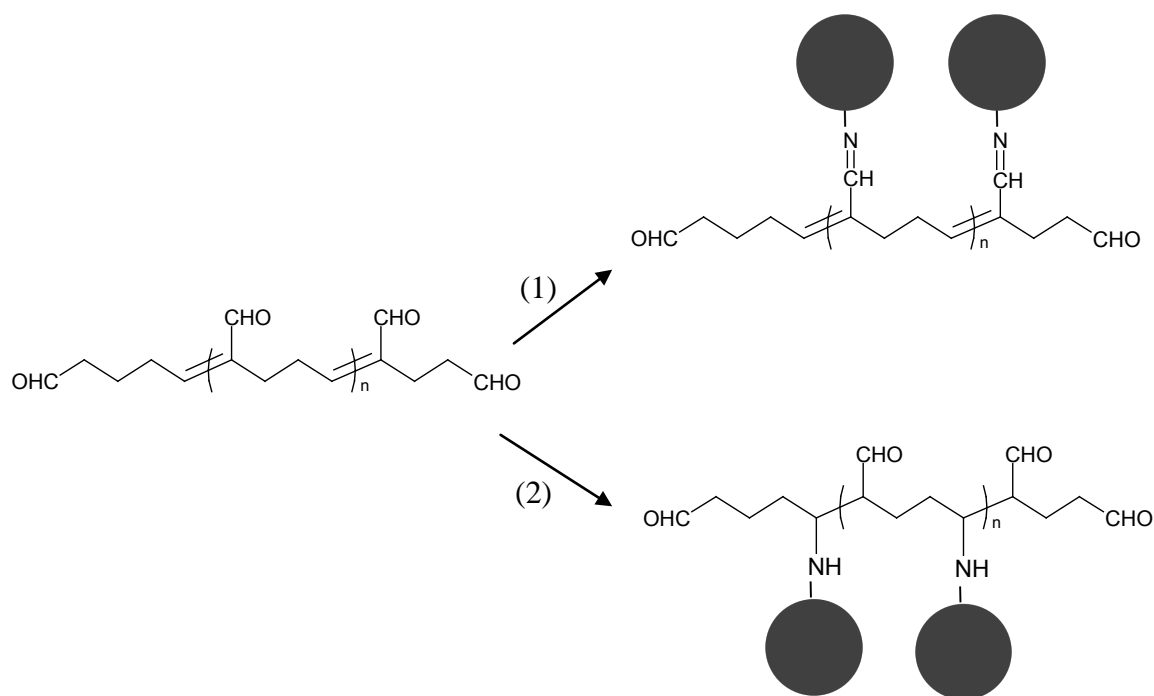
Figure 13. Proposed mechanism of protein cross-linking by glutaraldehyde.
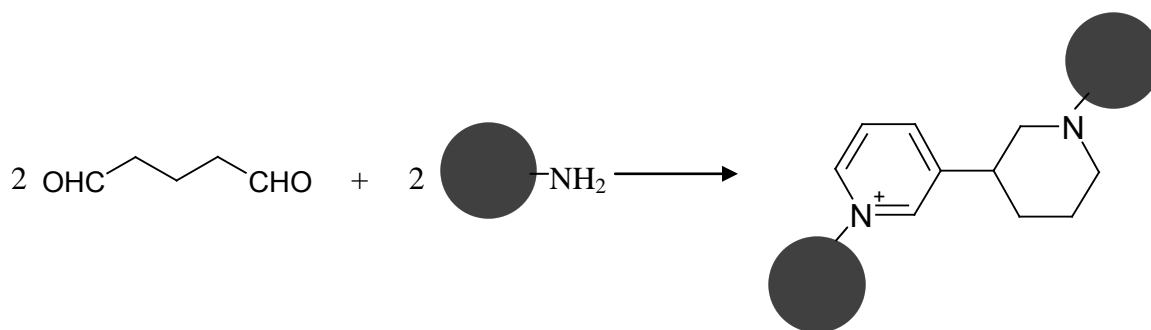
Figure 14. Proposed mechanism of protein cross-linking by glutaraldehyde

Chemical cross-linking of proteins have been extensively used to prepare protein-based materials. In most cases, protein cross-linking not only leaves the tertiary and secondary structure of protein unchanged, but also enhances protein's mechanical and thermal stability.[147-150] Furthermore, if the cross-linking reactive groups are not located at the binding or catalytic moiety of enzyme, cross-linking will not significantly perturb the enzymatic activity of proteins.[24, 25]

9. Protein Properties

In this dissertation, we have investigated the physiochemical properties of protein aqueous solutions for three globular proteins: lysozyme from chicken egg white, Bovine Serum Albumin (BSA), and laccase from *Trametes Versicolor*.

Lysozyme Lysozyme (EC 3.2.1.17) is a family of enzymes destroying the cell wall of certain microorganisms. It catalyzes hydrolysis of β–1,4 glucosidic linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan and

between N-acetyl-D-glucosamine residues in chitodextrins.[156] Lysozyme is abundant in a number of secretions, e.g. tears, saliva, and mucus. It can also be found in chicken egg white. The physicochemical and biological properties of chicken egg-white lysozyme have been extensively studied.[157] It also crystallizes easily. This makes lysozyme a widely used protein model for the characterization of phase transitions.[56,58,120,158,159]

Chicken egg-white lysozyme is composed of a single polypeptide chain of 129 amino acids (Table 1). The molecular weight of lysozyme is 14,307g/mol. The isoelectric point is at pH=10.7. The secondary and tertiary structures of chicken egg-white lysozyme were determined by X-ray diffraction (Fig 15).[160] The details of secondary structure are summarized in Fig 16. The polypeptide chain consists of 40% helical (7 helices; 52 residues), 10% beta sheet (9 strands; 14 residues) and four intra molecular disulfide bridges. The unit cell of lysozyme crystal has the coordinates (27.07 Å, 31.25 Å, 33.76Å) and the angles between axises (87.98°, 108.00°, 112.11°).

Table 1. Amino acids sequence of lysozyme

1-10	K	V	F	G	R	C	E	L	A	A
11-20	A	M	K	R	H	G	L	D	N	Y
21-30	R	G	Y	S	L	G	N	W	V	C
31-40	A	A	K	F	E	S	N	F	N	T
41-50	Q	A	T	N	R	N	T	D	G	S
51-60	T	D	Y	G	I	L	Q	I	N	S
61-70	R	W	W	C	N	D	G	R	T	P
71-80	G	S	R	N	L	C	N	I	P	C
81-90	S	A	L	L	S	S	D	I	T	A
91-100	S	V	N	C	A	K	K	I	V	S
101-110	D	G	N	G	M	N	A	W	V	A
111-120	W	R	N	R	C	K	G	T	D	V
121-129	Q	A	W	I	R	G	C	R	L	

Figure 15. Tertiary structure of chicken egg-white lysozyme determined by X-ray diffraction with resolution 0.65 Å. Image is rendered by vmd based on structure data from protein data bank (2VB1)

Figure 16. Secondary structure of chicken egg-white lysozyme. The black helix represents α-helix; the wide arrow represents β-sheet; the gray curve represents turn; the dot line represents disulfide bond.

Bovine Serum Albumin (BSA)   BSA is a serum albumin protein which finds numerous biochemical applications. BSA is relatively stable, inert in many biochemical reactions, and present in large amount in bovine blood. It is used in western blot and Enzyme-Linked Immunosorbent Assay (ELISAs) to block the nonspecific binding of protein to the membrane and plates.[161] It is also used in restriction digest of DNA to prevent adhesion of restriction enzymes to the vessel's wall. In colorimetric methods of measuring total protein concentration, BSA is used as the standard protein.[162] In medical applications, Human Serum Albumin (HSA), which is similar to BSA, is often used as a biocompatible material for drug-delivery carriers.[163]

BSA is usually present in a form of a dimer consisting of two identical polypeptide chains bridged by disulfide bonds. Each chain is composed of 525 amino acids (table 2). The molecular weight of BSA monomer is 66,400 g/mol. The isoelectric point of BSA is at pH = 5.0.

The X-ray diffraction data for BSA is not available. However, it is available for HSA, which resembles BSA for 90% of its primary structure. The tertiary and secondary structure of BSA was determined by computer modeling based on amino acid composition of BSA and the X-ray diffraction data for HAS (PDB 2BXH) (Fig 17).[164, 165] HSA is rich in helices (29 helices; 399 residues), has no β sheet, and has 16 intramolecular disulfide bridges. The unit cell of HSA crystal has the coordinates (54.71 Å, 55.20 Å, 119.88Å) and the angles between axises (81.27°, 91.09°, 65.29°).

Table 2. Amino acids sequence of BSA

1-10	S	E	I	A	H	R	F	K	D	L
11-20	G	E	H	F	K	G	L	V	L	I
21-30	A	F	S	Q	Y	L	Q	C	P	F
31-40	D	E	H	V	K	L	V	N	E	L
41-50	T	E	F	A	K	T	C	V	A	D
51-60	E	S	H	A	G	C	E	K	S	L
61-70	H	T	L	F	G	D	E	L	C	K
71-80	V	A	S	L	R	E	T	Y	G	D
81-90	M	A	D	C	E	K	Q	E	P	E
91-100	R	N	E	C	F	L	S	H	K	D
101-110	S	P	D	L	P	K	L	K	P	D
111-120	P	N	T	L	C	D	E	F	K	A
121-130	D	E	K	F	W	G	K	Y	L	Y
131-140	E	I	A	R	H	P	Y	F	Y	A
141-150	P	E	L	Y	A	N	K	Y	N	G
151-160	V	F	Q	E	C	Q	A	E	D	K
161-170	G	A	C	L	P	K	I	E	T	M
171-180	R	E	K	V	L	A	S	A	R	Q
181-190	R	L	R	C	A	S	I	Q	K	F
191-200	G	E	R	A	L	K	A	W	S	V
201-210	A	R	L	S	Q	K	F	P	K	A
211-220	E	F	V	E	V	T	K	L	V	T
221-230	D	L	T	K	V	H	K	E	C	H
231-240	G	D	L	E	C	A	D	R	A	D
241-250	L	A	K	Y	I	C	D	N	Q	D
251-260	T	I	S	K	L	K	E	C	D	K
261-270	P	L	E	K	S	H	C	I	A	E
271-280	V	E	K	D	A	I	P	E	N	L
281-290	P	L	T	A	D	F	A	E	D	K
291-300	D	V	C	K	N	Y	Q	E	A	K
301-310	D	A	F	L	G	S	F	L	Y	E
311-320	Y	S	R	H	P	E	Y	A	V	S
321-330	V	L	R	L	A	K	E	Y	E	A
331-340	T	L	E	C	A	K	D	P	H	A
341-350	C	Y	S	T	V	F	D	K	L	K
351-360	H	L	V	D	E	P	Q	N	L	I
361-370	K	Q	N	C	D	Q	F	E	K	L
371-380	G	E	Y	G	F	Q	N	A	L	I
381-390	V	R	Y	T	R	K	V	P	Q	V

391-400	S	T	P	T	L	V	E	V	S	R
401-410	S	L	G	K	V	G	T	R	C	T
411-420	K	P	E	S	E	R	M	P	C	T
421-430	E	D	Y	L	S	L	I	L	N	R
431-440	L	C	V	L	H	E	K	T	P	V
441-450	S	E	K	V	T	K	C	T	E	S
451-460	L	V	N	R	P	C	F	S	A	L
461-470	T	P	D	E	T	Y	V	P	K	A
471-480	F	D	E	K	L	F	T	F	H	A
481-490	D	I	C	T	L	P	D	T	E	K
491-500	Q	I	K	Q	T	A	L	V	E	L
501-510	K	H	K	P	K	A	T	E	Q	L
511-520	K	T	V	M	E	N	F	V	A	F
521-525	V	D	K	C	A					
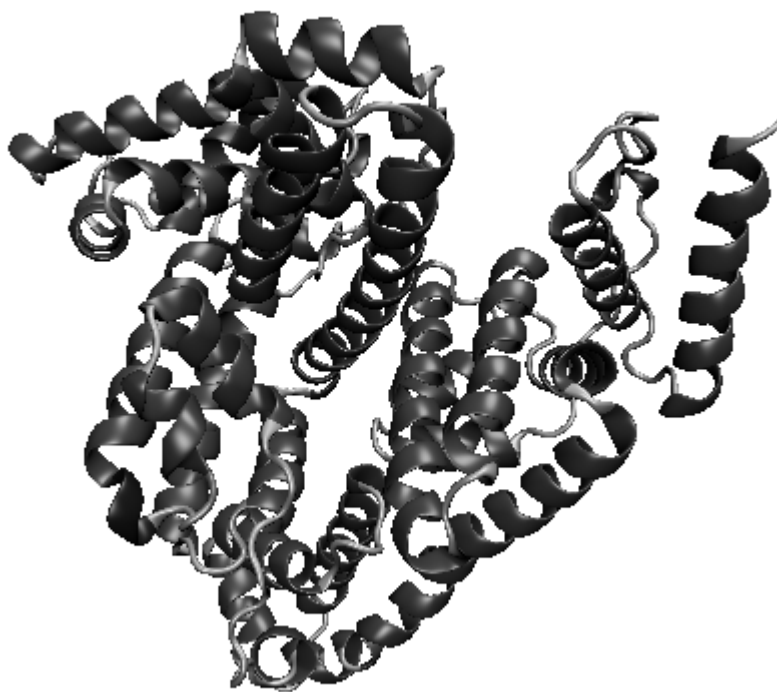
Figure 17. Tertiary structure of BSA determined by computer modeling.
The image is rendered by vmd using the structure data from PDB 2BXH.

Laccase Laccase (EC 1.10.3.2) is a family of multi-copper-containing oxidases. In acidic conditions, it catalyzes the one-electron oxidation of phenols, anilines and similar compounds to the corresponding reactive radicals (Fig 18), and the cofactors are copper ions.[166-168] Laccase finds its applications in industrial catalysts for the petroleum industry, textile dyeing and finishing, teeth whitening, environmental, diagnostic, and synthetic uses. [23, 167-170] It has also been used as cathode in an enzyme catalyzed fuel cell.[171] The activity of laccase can be monitored by a spectrophotometrical method.[172] Some commonly used substrates in this method are 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS), syringaldazine, 2,6-dimethoxyphenol, and dimethyl-p-phenylenediamine.[173-179] Oxygen sensors are also used for detecting laccase activity through monitoring the reduction of oxygen in water due to oxidation of substrates.[180]
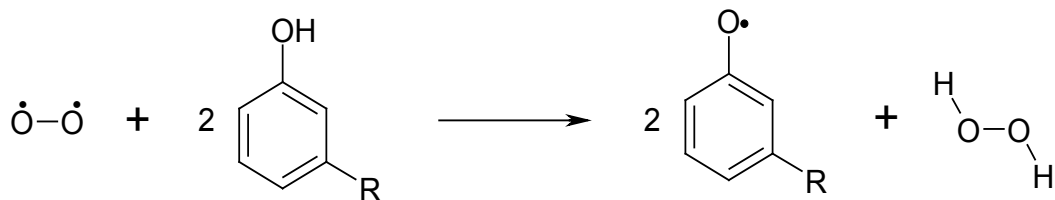
Figure 18. Monoelectronic oxidation of phenol derivative catalyzed by laccase.

Laccase can be found in various plants, fungi, and microorganisms. In our work, we use a laccase isolated from *Tranmetes versicolor* (a white-rot basidiomycete fungus). This type of laccase is composed of one polypeptide chain, but it also occurs in oligomer forms. [166-168] The 449 amino-acid sequence of the monomer is reported in Table 3.

Tertiary and secondary structures of laccase from *Tranmetes versicolor* has been determined by X-ray diffraction (PDB 1KYA) (Fig 19). The details of the secondary structure are shown in Fig 20. This polypeptide chain consists of 11% helical (13 helices; 57 residues), 37% beta sheet (44 strands; 189 residues) and two intramolecular disulfide bridges. The crystallography shows that the catalytic core of laccase from *Tranmetes versicolor* contains four copper ions. One copper ion (Cu1) is located in a mononuclear type I (T1) Cu site. The other three (Cu2a, Cu2b and Cu3) are in a trinuclear typeII and III (T2/T3) Cu site. Because of the presence of intense Cu-Cu linkage, the laccase has a blue color and an absorbance maximum at 600 nm. The substrate molecule is oxidized by transferring one electron to Cu1. Then, the electron is transferred to the trinuclear copper center and finally captured by oxygen molecules (Fig 21).[173-176]

Table 3. Amino acids sequence of Laccase from *Tranmetes versicolor*

Range										
1-10	A	I	G	P	A	A	S	L	V	V
11-20	A	N	A	P	V	S	P	D	G	F
21-30	L	R	D	A	I	V	V	N	G	V
31-40	F	P	S	P	L	I	T	G	K	K
41-50	G	D	R	F	Q	L	N	V	V	D
51-60	T	L	T	N	H	T	M	L	K	S
61-70	T	S	I	H	W	H	G	F	F	Q
71-80	A	G	T	N	W	A	D	G	P	A
81-90	F	V	N	Q	C	P	I	A	S	G
91-100	H	S	F	L	Y	D	F	H	V	P
101-110	D	Q	A	G	T	F	W	Y	H	S
111-120	H	L	S	T	Q	Y	C	D	G	L
121-130	R	G	P	F	V	V	Y	D	P	K
131-140	D	P	H	A	S	R	Y	D	V	D
141-150	N	E	S	T	V	I	T	L	T	D
151-160	W	Y	H	T	A	A	R	L	G	P
161-170	R	F	P	L	G	A	D	A	T	L
171-180	I	N	G	L	G	R	S	A	S	T
181-190	P	T	A	A	L	A	V	I	N	V
191-200	Q	H	G	K	R	Y	R	F	R	L
201-210	V	S	I	S	C	D	P	N	Y	T
211-220	F	S	I	D	G	H	N	L	T	V
221-230	I	E	V	D	G	I	N	S	Q	P
231-240	L	L	V	D	S	I	Q	I	F	A
241-250	A	Q	R	Y	S	F	V	L	N	A
251-260	N	Q	T	V	G	N	Y	W	I	R
261-270	A	N	P	N	F	G	T	V	G	F
271-280	A	G	G	I	N	S	A	I	L	R
281-290	Y	Q	G	A	P	V	A	E	P	T
291-300	T	T	Q	T	T	S	V	I	P	L
301-310	I	E	T	N	L	H	P	L	A	R
311-320	M	P	V	P	G	S	P	T	P	G
321-330	G	V	D	K	A	L	N	L	A	F
331-340	N	F	N	G	T	N	F	F	I	N
341-350	N	A	S	F	T	P	P	T	V	P
351-360	V	L	L	Q	I	L	S	G	A	Q
361-370	T	A	Q	D	L	L	P	A	G	S
371-380	V	Y	P	L	P	A	H	S	T	I
381-390	E	I	T	L	P	A	T	A	L	A

391-400	P	G	A	P	H	P	F	H	L	H
401-410	G	H	A	F	A	V	V	R	S	A
411-420	G	S	T	T	Y	N	Y	N	D	P
421-430	I	F	R	D	V	V	S	T	G	T
431-440	P	A	A	G	D	N	V	T	I	R
441-450	F	Q	T	D	N	P	G	P	W	F
451-460	L	H	C	H	I	D	F	H	L	E
461-470	A	G	F	A	I	V	F	A	E	D
471-480	V	A	D	V	K	A	A	N	P	V
481-490	P	K	A	W	S	D	L	C	P	I
491-499	Y	D	G	L	S	E	A	N	Q	

Figure 19. Tertiary structure of Laccase from *Tranmetes versicolor* determined by X-ray diffraction with resolution 1.9 Å. The image is rendered by vmd based on structure data from protein data bank 1KYA.
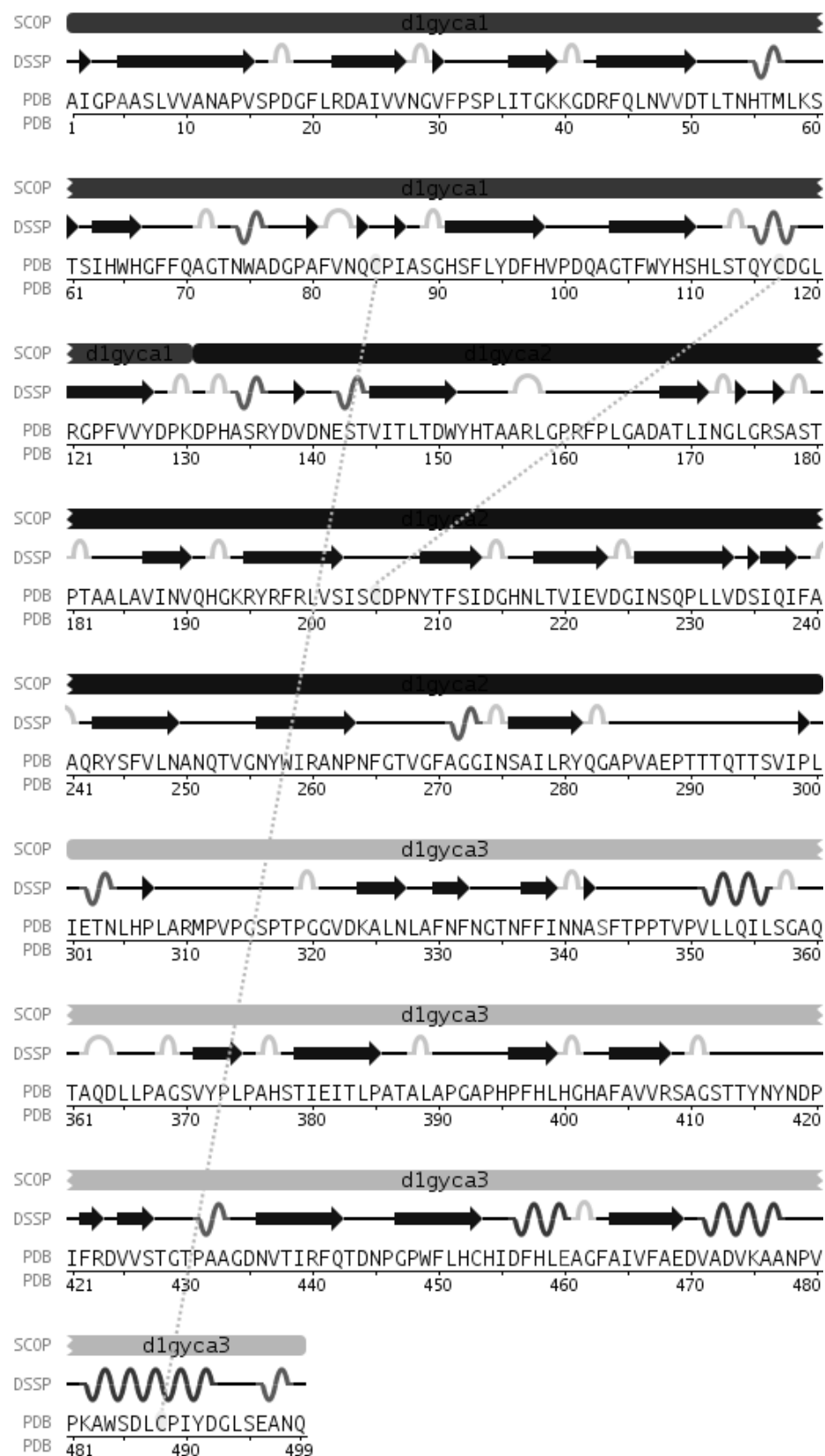
Figure 20. Secondary structure of Laccase from *Tranmetes versicolor*. The black helix represents α-helix; the wide arrow represents β-sheet; the gray curve represents turn; the dot line represents disulfide bond.
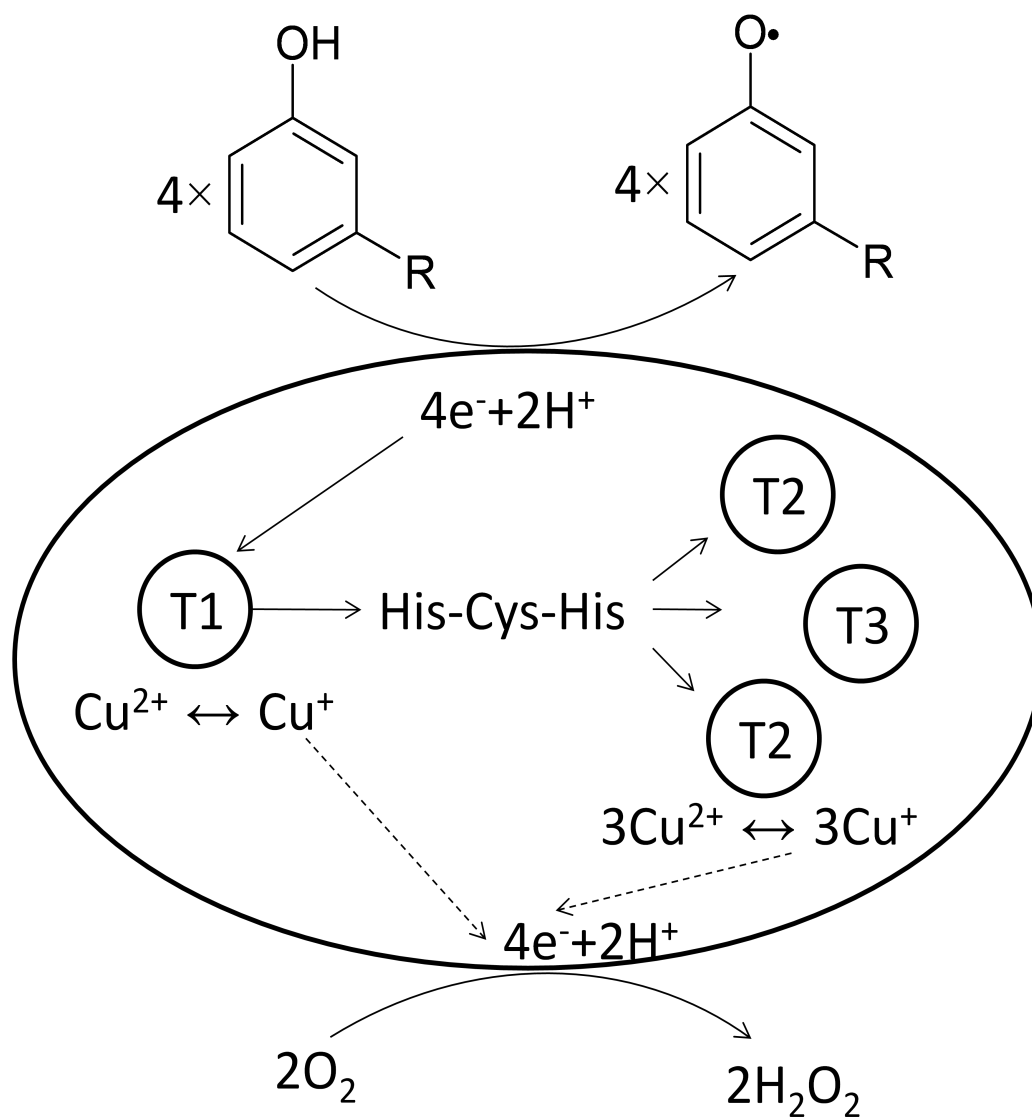
Figure 21. Scheme of oxidation catalyzed by laccase.

Part I:

Materials and Experimental Techniques

Chapter 2

Solution Preparation and Characterization

1.   Materials

Bovine serum albumin (BSA) (99% purity) and chicken egg-white lysozyme (recrystallized six times and lyophilized) were purchased from Sigma. The molecular weight was assumed to be 66.4 kg/mol for BSA and 14.3 kg/mol for lysozyme. HPLC (System Gold, Beckman Coulter) with a size-exclusion column (Biosep-SEC-S 2000, Phenomenex) on lysozyme showed the purity to be better than 99%. In the case of BSA, HPLC shows the presence of 20% oligomers (Fig 1a). Thus, further purification was performed for albumin using size-exclusion preparative chromatography described in the next section. Laccase from Trametes versicolor was purchased from Fluka. The molecular weight was assumed to be 56.0 kg/mol. The raw material contains a large amount of insoluble impurity. Furthermore size-exclusion HPLC shows that the soluble material contains 70% soluble impurity (Fig 2a). A purification procedure was designed for laccase based on ultrafiltration and ion exchange preparative chromatography. This procedure is described in the following section. All the protein raw materials were stored at -20°C in the freezer.

Polyethylene glycol (PEG) with average molecular weights of 1.45 kg/mol (PEG1450) and 8.0 kg/mol (PEG8000) were purchased from Acros Chemicals and used without further purification. Sodium chloride, ammonium sulfate, sodium sulfate, sodium perchlorate, sodium acetate, monobasic sodium phosphate, anhydrous dibasic sodium phosphate, and boric acid were purchased from Fischer Scientific and used without further purification. Glutaraldehyde (25% aqueous solution) was purchased from Acros and used without further purification.

2.   Protein Purification

Bovine serum albumin (BSA) raw material was dissolved into phosphate buffer (0.05M, pH 7.1) and purified using size-exclusion preparative chromatography. The column (Bio-Rad Econo-Colum 2.5cm×100cm) was packed using Sephacryl S-200 purchased from GE Healthcare. The mobile phase was a sodium phosphate buffer (0.05M, pH 7.1), and the flow rate was 1.5mL/min. Aliquots of 5mL were collected by a fraction collector (Amersham Biosciences, FRAC-100), and the corresponding protein concentrations were measured by UV absorption at 278nm (DU 800 spectrophotometer, Beckman Coulter). The monomer fraction was collected and stored at 4°C. Size-exclusion HPLC on the monomer fraction showed the purity to be better than 99% (Fig 1b).

Laccase raw material was dissolved into DI water and centrifuged to sediment the insoluble impurities. The supernatant was brown. The low molecular-weight fraction of soluble impurities was partially removed by ultrafiltration (YM30k, Amicon, Millipore) against DI water three times. The filtrate was yellow. However, aggregation was observed during the ultrafiltration procedure in the concentrated retentate. This retentate was centrifuged to sediment these aggregates. Then the supernatant was carefully transferred into a clean vial. The brown aggregates were washed with minimum amount of DI water. Anion-exchange preparative chromatography was then used to further purify the supernatant. The column (Bio-Rad Econo-Colum 2.5cm×30cm) was packed using DEAE Sephacel purchased from GE Healthcare. The mobile phase was a sodium acetate
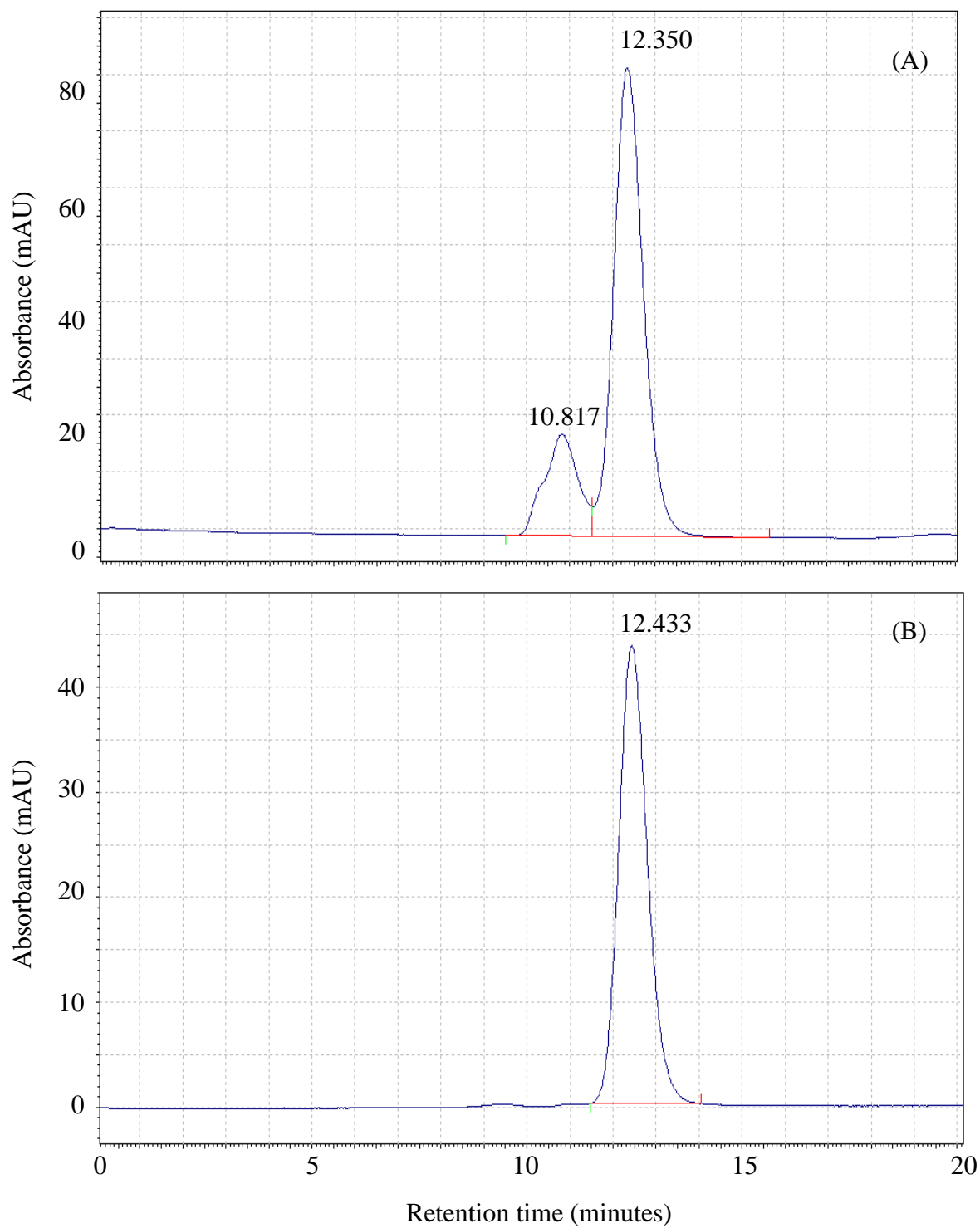
Figure 1. Size-exclusion Chromatography of BSA before (A) and after (B) purification. 0.1M phosphate buffer pH7.1, flow rate 0.6mg/mL, Superdex 75 Tricorn HPLC column (Amersham Biosciences)
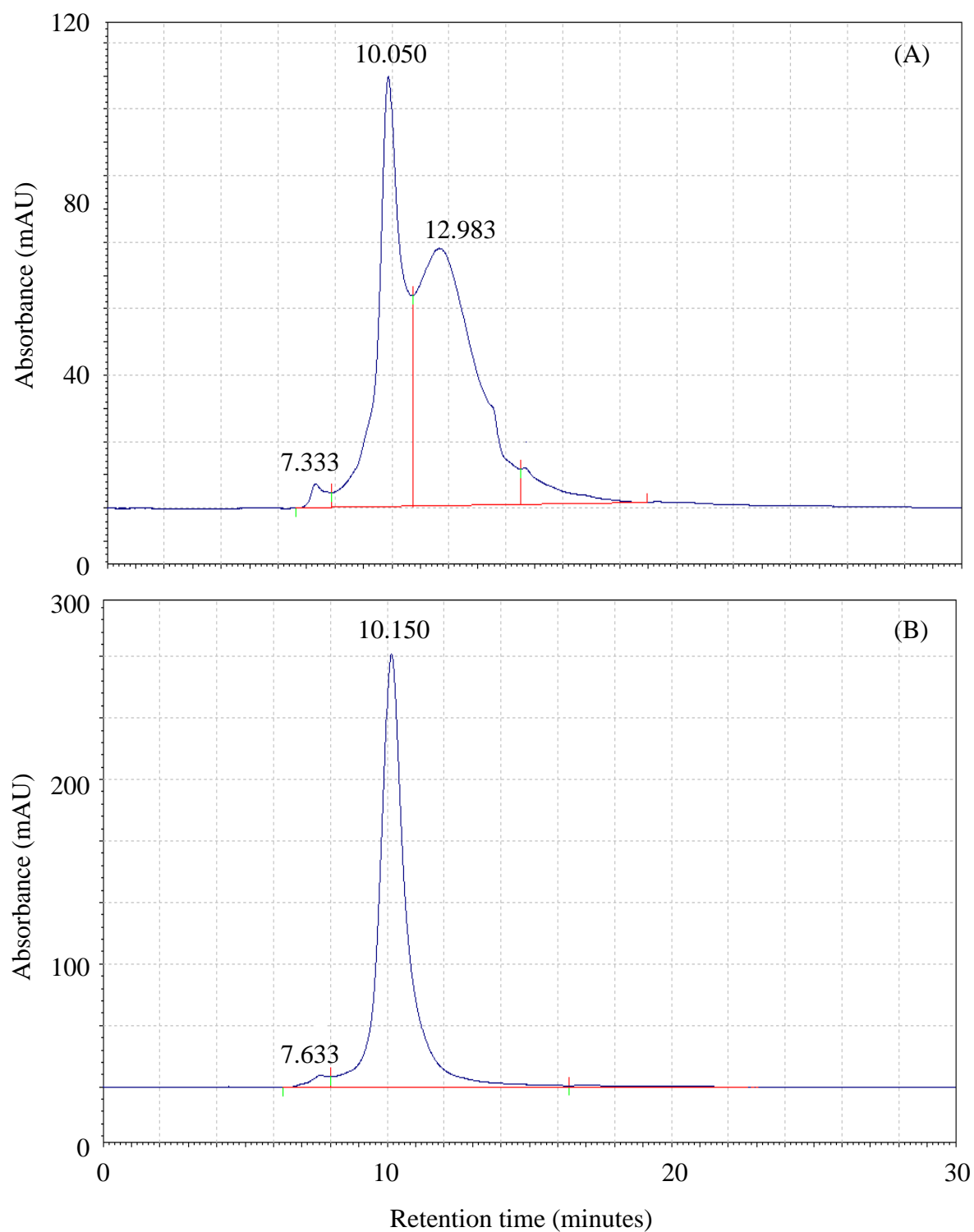
Figure 2. Size-exclusion Chromatography of laccase before (A) and after purification (B). 0.1M acetate buffer pH4.5, 0.2M NaCl, flow rate 1mg/mL, Biosep-SEC-S 2000 HPLC column (Phenomenex).

buffer (0.1 M, pH 4.5) with a sodium chloride linear gradient 0-5M in 300 minutes controlled by HPLC pumps, and the flow rate was 1.5 mL/min. Aliquots of 5mL were collected, and tested by the enzymatic activity assay described in Section 9. The fractions that showed laccase activity were collected and stored at 4 °C. The purified laccase solutions exhibit a green color and an absorbance band at 600nm. Size-exclusion HPLC on the purified laccase solution showed the purity to be better than 95% (Fig 2b).

3. pI Measurements

The isoelectric points of albumin and laccase were measured by isoelectric focusing electrophoresis (IEF) with Novex IEF gel pH 3-10. The cathode buffer was 20 mM lysine and 20 mM arginine. The anode buffer was 7 mM phosphoric acid. Sample buffer was 20 mM lysine, 20 mM arginine and 15% glycerol. IEF Marker 3-10 (SERVA liquid Mix) was used as standard. The electrophoresis apparatus was set up with high voltage DC power source (Power Designs Inc, Model 1570) and invitrogen XCell SureLock Mini-Cell. The voltage was increased from 0 V to 300 V in 30 minutes, and kept constant for 2.5 hours. Then the gel was removed from the cassette and fixed with a solution of 10% trichloroacetic acid (TCA) and 3% sulphosalicylic acid. At last, the gel was stained with SimplyBlue SafeStain from invitrogen for 3 hours and destained with distilled water for 3 hours. The results showed pI of BSA is 5.0 ± 0.2, and pI of laccase is 3.5 ± 0.5. The pI of hen egg white lysozyme is 11 from literature,[215] thus it cannot be determined by IEF gel pH 3-10.

4.   Solution Preparation

Protein solutions were dialyzed exhaustively (Amicon, Millipore) into the desired aqueous buffer and concentrated by centrifugation (3500g, Allegra 25R, Beckman Coulter) using ultrafiltration devices (Centricon YM-30, Milipore). Protein concentration in the concentrated protein stock solution was determined by UV absorption. PEG8000-buffer and NaCl-buffer stock solutions were prepared by weight with a Mettler Toledo AT400 electrobalance. Small amounts of these solutions were added to protein-buffer solutions. For protein cross-linking, glutaraldehyde-buffer stock solutions were prepared by weight. Small aliquots of glutaraldehyde stock solutions were added prior to the physicochemical characterizations. The final mass-volume concentrations of protein and PEG and the molarities of salt and glutaraldehyde were determined using the weights and specific volumes of the stock solutions. Measurements of pH were made using a Corning model 320 pH meter with an Orion model 8102 combination ROSS pH electrode. The meter was calibrated with standard pH 7.00 and pH 4.00.

5.   Density Measurements

All solution densities were measured with a Mettler-Paar DMA40 density meter with an RS-232 output to a Apple II+. By time averaging the output, a precision of ± 0.00001 g/mL or better could be achieved. Accuracy is somewhat lower because it is related to the composition uncertainty of the solutions. The temperature of the vibrating

tube in the density meter was controlled with water from a large well-regulated water bath whose temperature was 25.00±0.01°C. The solution density ρ_x is related to the vibration period T_x of cell tube by the following relation:

$$\rho_x = A + B\,T_x^2$$

where *A, B* are two instrumental constants. Two reference periods must be measured for the determination of the instrumental parameters. We have chosen the water (ρ_{water} = 0.997045 g/mL) and the air ($\rho_{air} \approx 0.0012$ g/mL) as reference systems. An accurate value of the air density was estimated by a state equation that shows explicit dependence on the pressure, temperature (25.00 °C), and humidity:

$$\rho_{air} = \frac{348.844p + (0.70892 - 0.00252T)h}{T} \quad (g/L)$$

where p = pressure (bar), h= % relative humidity, T=absolute temperature (K)

6.   Solution Volumetric Properties

The partial specific volume of component *i* in a solution, the volume occupied by unit mass, is the defined by

$$V_i^{sp} = \left(\frac{\partial V}{\partial m_i}\right)_{m_{j \neq i}}$$

where V is the volume of solution, and m_i is the mass of component *i*. By assuming that the volume change of mixing is negligible (this is a good approximation for aqueous solutions containing proteins and polyethylene glycol) the total volume can be calculated with the mass of each component by

$$V = \sum m_i V_i^{sp}.$$

The specific volumes used are 0.735mL/g for proteins[181] and 0.84mL/g for polyethylene glycol.[56] Specific volumes of buffer and salt solution were determined by density measurements. Therefore, concentrations in g/L can be calculated from the corresponding measured weight fractions.

7.   Measurements of Protein Concentration

The concentration of BSA in solution was determined by UV absorption at 278 nm (DU 800 spectrophotometer, Beckman Coulter), using an extinction coefficient value of 0.667 mg^{-1} mL cm^{-1}.[182] The concentration of lysozyme in solution was determined by UV absorption at 280 nm, using an extinction coefficient value of 2.64 mg^{-1} mL cm^{-1}.[183]

The concentration of laccase in the purified solution was determined by UV absorbance at 280 nm, using an extinction coefficient obtained using the total protein colorimetric (biuret) method. The biuret reagent was prepared by dissolving 1.50 g of pentahydrate copper (II) sulfate, 4.5 g of sodium potassium tartrate and 2.5 g of potassium iodide in 250 mL of distilled water, adding 50 mL of 6.0-M NaOH and diluting to a total volume of 500 mL with distilled water. The reagent was stored in a parafilmed bottle at 4°C and was stable for approximately six months. The standard BSA solutions were prepared by weight, and their concentrations were determined by UV absorption at 278 nm. The assay was performed by mixing precisely 1 mL standard solution and 9 mL biuret reagent, incubating the solution at 25°C for 20 min, and then

measuring the absorbance at 550nm. The standard curve was obtained and reported in Fig. 3. Three purified laccase solutions were subjected to the biuret assay following the above procedures. Their concentrations were determined by the standard curve, and the corresponding UV absorbance at 280 nm was measured. Then, the extinction coefficient of laccase at 280nm was determined to be 0.84 mg^{-1} mL cm^{-1}.
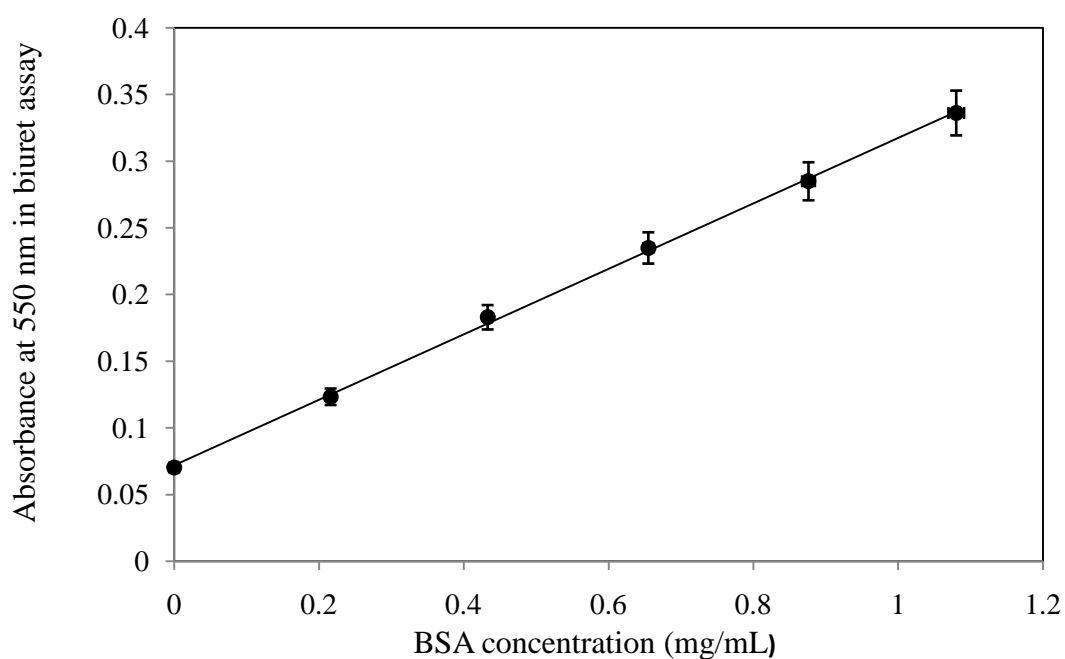
Figure 3. The standard curve of total protein concentration biuret method using BSA standards.

8.   PEG Concentration Measurements in Coexsisting Phases

In order to characterize the composition of coexisting phases for the phase-transition experiments, the concentration of PEG in the two coexisting phases was determined by size-exclusion HPLC with a Superdex 75 Tricorn HPLC column and a refractive index detector (RI-2031, Jasco), using a standard curve. This detector is a deflection differential refractometer and measures the difference between the refractive indices of the sample and the reference (solvent). The measurable refractive index (RI) range is 1.00-1.75 RIU (refractive index unit). The linear range of RI difference is 5.0×10^{-6} RIU in the "STD" mode, and 5.0×10^{-5} RIU in the "WIDE" mode for the concentrated samples. The static noise of this detector is 0.2×10^{-8} RIU with water as solvent. This detector can work with three response speed, i.e. "STD" for flow rate lower than10mL/min, "FAST" for flow rate 10-50 mL/min and "SLOW" for the digital filtering, depending on the tubing arrangement. All the measurements were performed in the "STD" RI range with a signal sensitivity of 6.4×10^{-5} RIU/100 mV(full scale). Flow rate was set in the "STD" mode and temperature was set at 25°C. The solvent passed through both the reference cell and the sample cell when the "purge" valve was open. This action replaced the solvent in the reference cell and removed bubbles. Then the valve was closed, and the solvent only passed through the sample cell. The standard curve was obtained by the HPLC-RI measurements on solutions with known PEG concentration prepared by weight.

9. Laccase Assay

The activity of laccase was measured using the 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) diammonium sulfate (ABTS) as a substrate (Fig 4). A 0.1-M acetate buffer at pH 4.0 was saturated with oxygen by bubbling this gas into the buffer for one minute. 1.0 mM ABTS stock solution was freshly prepared by disolving 2.2 mg ABTS into 4 mL acetate buffer. Then, 1 μL laccase solution (~0.1 mg/mL) was mixed with 399 μL ABTS stock solution and the absorbance of product $ABTS^{\cdot+}$ was measured at 420 nm ($\varepsilon_{420} = 3.6 \times 10^4$ $M^{-1}cm^{-1}$) every 15 seconds for 5 minutes. The initial rates of enzymatic reactions were calculated from the time dependence of absorbance.
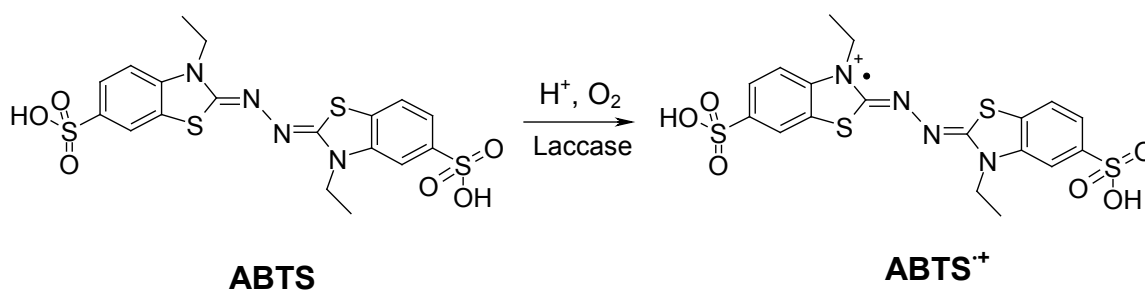
Figure 4. One electron oxidation of ABTS catalyzed by laccase.

Chapter 3

Characterization of Phase Diagram by

Isothermal Partitioning Measurements

1. Introduction

The phase diagram can be characterized by measuring the protein and polymer concentrations in the two coexisting phases at several fixed temperatures within the two-phase domain of the phase diagram (Chapter 8). A cartoon of the two coexisting phases associated with liquid-liquid phase separation (LLPS) is shown in Fig 1.
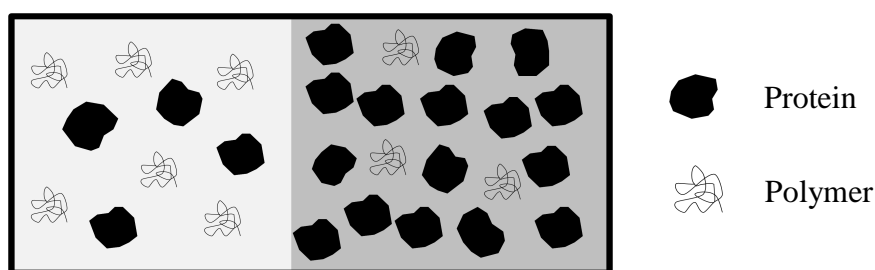
Figure 1. Protein-polymer partitioning between the two coexisting phases.

Partitioning measurements are performed one day after LLPS is induced so that equilibrium could be reached. Their experimental errors in the determination of concentration are often of the order of 10% due to the incomplete separation of the two phases and cross-contamination caused by the operation of sampling the two phases from small test tubes (400 μL). Nonetheless partitioning measurements only can be used for the determination of tie-lines associated with the phase boundary; i.e. those lines that connect pairs of composition-coordinates corresponding to the coexisting phases.

2. LLPS phase diagram of Protein/Polymer Solution

The LLPS boundary of BSA-PEG1450-buffer solutions has been characterized by isothermal partitioning measurements. The two coexisting phases were obtained by quenching a sample to a temperature below the LLPS temperature as described in previous work. The samples were left at the established temperature for approximately one day. The opaque samples were then inserted in Teflon test tube holders thermally equilibrated at exactly the same temperature and promptly located in a centrifuge thermally equilibrated at approximately the same temperature (± 1 °C). After centrifugation ($1000g$, 5 min), two liquid phases separated by a meniscus were obtained (Fig 2).
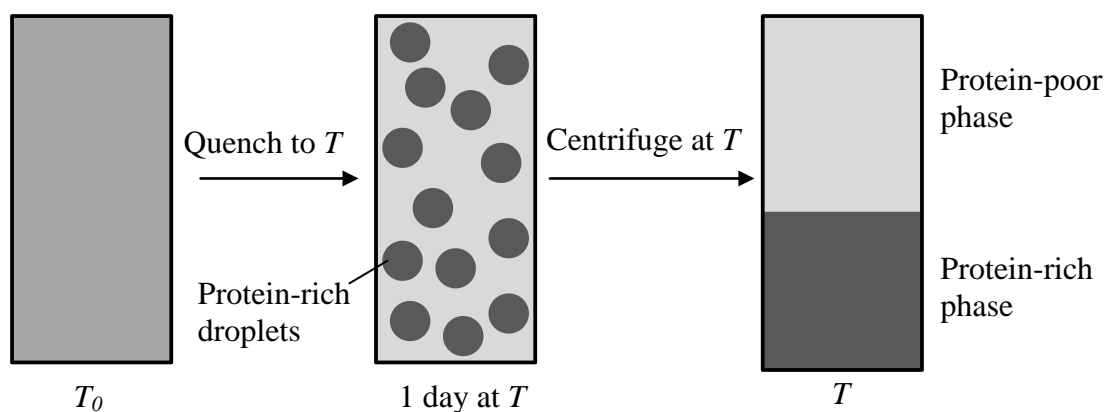
Figure 2. Scheme of partitioning measurements on LLPS of protein/PEG solution T_0 is the initial temperature at which the solution is homogenous, and T is the temperature, at which the LLPS occurs.

A fraction of the top solution (far from the liquid-liquid interface) was taken from the test tube. Then the remaining part of the top solution and a fraction of the bottom solution close to the meniscus were carefully removed minimizing mixing of the two phases. A fraction of the bottom solution was then taken. The protein concentration in each phase was determined by UV absorption. The concentration of PEG1450 in each aliquot was determined by using the standardized refractive-index detector (RI-2031, Jasco) as described in Chapt 2. Separation of PEG from BSA was achieved by isocratic elution of the mixture on a size-exclusion HPLC column (Superdex 75, Amersham Biosciences) (Fig 3). Sodium acetate buffer (0.01 M, pH 5.2) was used as the mobile phase with a flow rate of 0.75 mL/min. The procedure was verified with BSA-PEG aqueous solutions of known compositions showing that the error in PEG concentration was less than 5%. In all cases, the measured protein and PEG concentrations in the two coexisting phases were consistent with the protein and PEG concentrations in the original homogeneous samples.
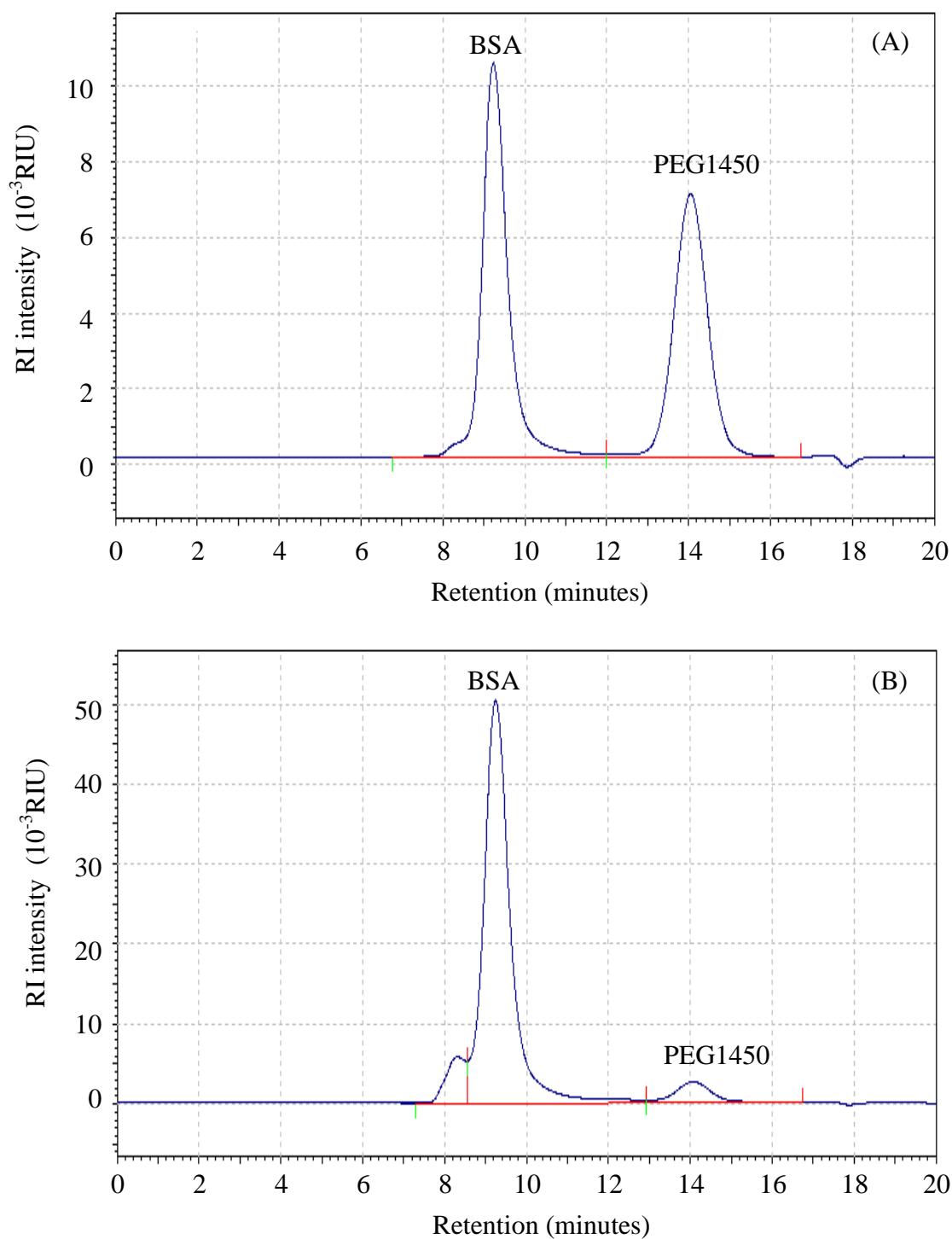
Figure 3. Size-exclusion Chromatography of the aliquots of the two coexisting phases. (A) is the protein-poor top solution, and (B) is the protein-rich bottom solution. Mobile phase is 0.1-M acetate buffer pH4.5, 0.2-M NaCl, flow rate is 1mg/mL. Superdex 75 SEC HPLC column (Amersham Biosciences) and refractive-index detector were used. RIU refers to the refractive index unit.

Chapter 4

Characterization of Phase Diagrams by

Turbidity Measurements

1. Introduction

The LLPS boundary was also characterized by turbidity measurements. When phase separation occurs, the turbidity of the initially transparent sample will dramatically increase due to the strong scattering of the incipient condensed phase. Therefore, turbidity measurements can be used to monitor the temperatures at which phase transformations occur, and the phase boundary can be determined by measuring phase separation temperature as a function of protein and PEG concentrations. This method provides results with a much smaller experimental error than that of isothermal partitioning measurements. However, contrary to partitioning, it cannot be used for the characterization of tie lines.

2. Apparatus

A turbidity meter was built by using a programmable circulating bath (1197P, VWR) connected to a homemade optical cell where the sample is located (Fig 1). Our device was optimized for low-temperature measurements. A 50% (v/v) methanol-water mixture was used as antifreeze in the circulating bath. The optical cell was subjected to continuous air flows to prevent condensation of moisture. The temperature at the sample location was measured by using a calibrated thermocouple (±0.1 °C). Light coming from a solid-state laser (633 nm, 5 mW, Coherent) goes through the sample, and the

transmitted intensity, I, is measured using a photodiode detector and collected by a computer-interfaced optical meter (1835-C, Newport).
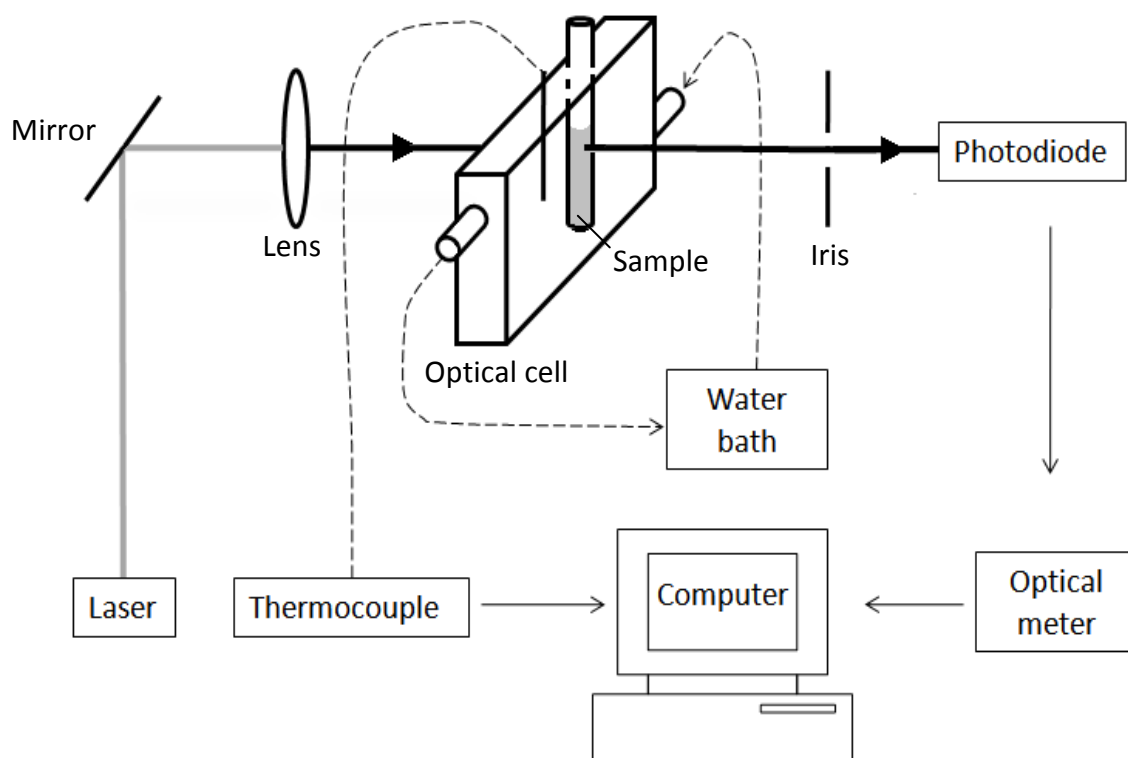
Figure 1. Scheme of "home-made" turbidity meter.

3. Turbidity Measurements on Phase Transformations

For a given transparent sample in the single-phase state, the transmitted intensity, I_0, was measured. The temperature was then slowly lowered to induce the phase transformation, and the corresponding transmitted intensity, I, was registered to the photodiode. To monitor opacification and phase separation as a function of temperature, the turbidity, τ, was employed. Turbidity is defined by,

$$\tau(t) = (1/L)\ln(I_0/I)$$

where $L = 0.4$ cm is the internal diameter of the test tube. There are two possible behaviors of τ as a function of temperature. If a sharp increase in τ occurs at a well-defined value of temperature, T_{ph}, then a temperature-induced phase transition such as LLPS has occurred. (Fig 2a). On the other hand, if τ steadily increases with temperature even at very slow cooling rate (1°C/hour), this suggests the presence of slow phase transformations such as aggregation or gelation (Fig 2b). This kinetic profile is typically associated with protein aggregation or gelation, which may or may not be induced by LLPS.

When phase transition occurred, a well-defined phase separation temperature was determined by turbidity measurements. The values of phase-separation temperatures, T_{ph}, for a given protein concentration were obtained by monitoring τ as a function of temperature. Different methods were employed to determine T_{ph}, depending on the kinetic evolution of opacification.
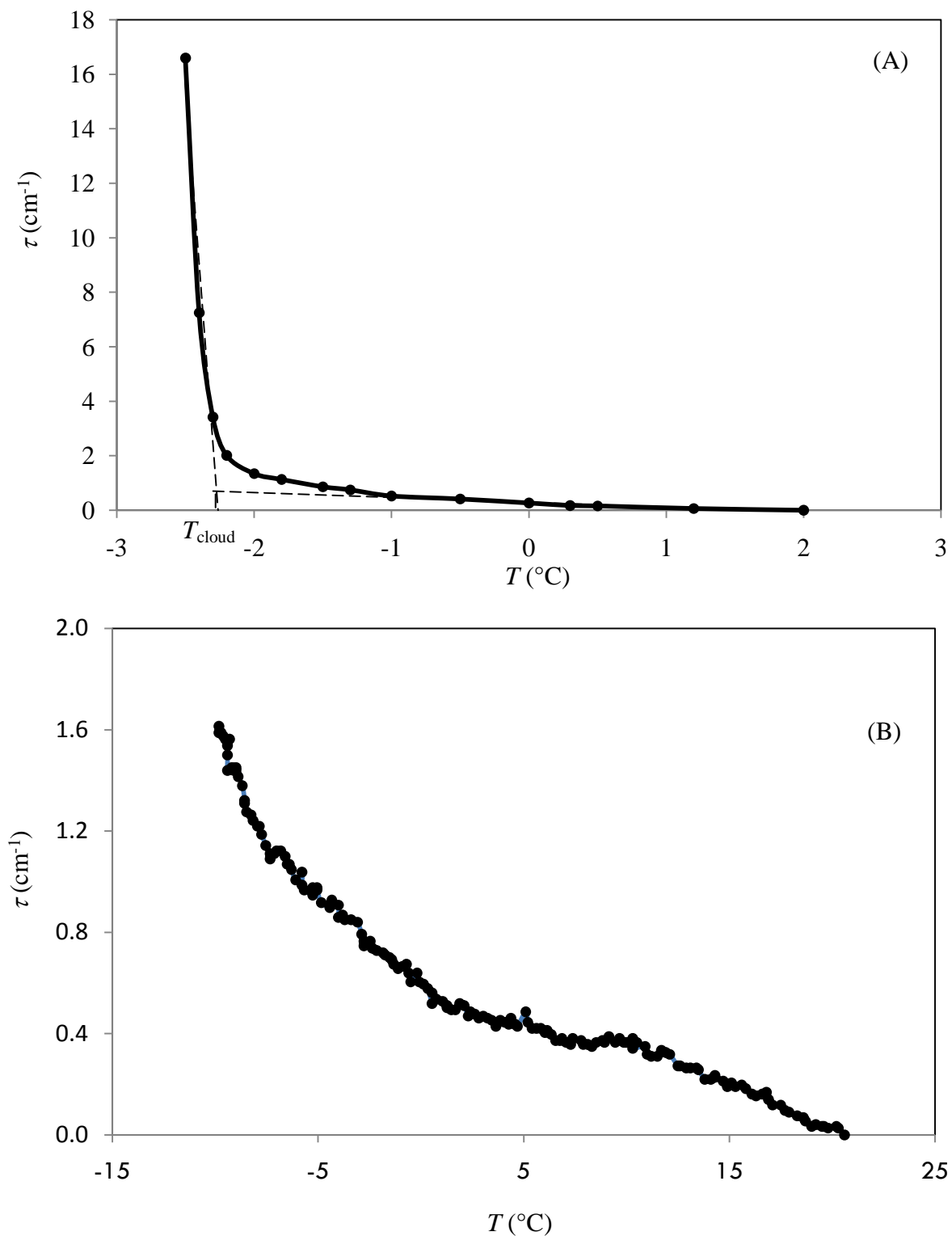
Figure 2. Turbidity-temperature profiles of phase separation induced by lowering the temperature. (A) LLPS of BSA in the presence of PEG 1450 at pH 5, where T_{ph} is the phase separation temperature. (B) Aggregation of laccase in the presence of PEG 8000 at pH 4.

4. Phase Separation Temperature

Method I: In the case of LLPS of BSA/PEG and lysozyme/NaCl, τ was recorded as a function of temperature, and increased rapidly after a well-defined temperature was reached. The temperature of the water bath was slowly decreased (0.5 °C/min), and the temperature at which the τ has a sharp increase was taken as the phase separation temperature, T_{cloud} When $I = 0$, the water bath temperature was increased at the same rate, and the temperature at which the τ has an inflection point was taken as the clearing temperature, T_{clear}. The phase separation temperature was calculated by $T_{ph} = (T_{cloud} + T_{clear})/2$ as recommended in previous work.[184] In some cases, the condensed phase sediment/coalescence was too fast (e.g. laccase/ammonium sulfate system) or the phase transition is irreversible (LLPS in the presence of cross-linker), thus the clouding temperature was chosen as phase separation temperature, $T_{ph} = T_{cloud}$.

Method II: In the case of laccase-PEG-buffer systems, τ weakly increases as temperature decreases. In this case, τ was recorded as a function of time at several given temperatures. Due to the generally slow rate of opacification, the temperature of the water bath was then lowered by 1 °C every 30 minutes. The temperature, at which the τ starts to steadily increase, was taken as the clouding temperature, T_{cloud}. The temperature of the water bath was then raised in 0.5 °C increments and the turbidity was recorded as a function of time. The temperature, at which the transmitted intensity would recover its initial value in less than 30 minutes, was taken as the clearing temperature, T_{clear}. The phase separation temperature was calculated by $T_{ph} = (T_{cloud} + T_{clear})/2$.

Therefore, the LLPS phase boundaries have been mapped by measuring phase separation temperature of solutions at different protein concentrations and polymer concentrations.

5. Kinetics of Phase Transformation and Measurements of Induction Time

To characterize the rates of isothermal opacification and kinetics of phase transformation, the turbidity of the sample undergoing phase transformation was recorded as a function of time at a given temperature. For the oligomerization induced liquid-liquid phase separation experiments at 25.0 ± 0.1 °C, the value of τ remained equal to zero after the addition of glutaraldehyde. However, after a well-defined and reproducible induction time was reached, a rapid increase of turbidity was observed. To characterize the rate of opacification, we determined the induction time, t_{ind}, by determining the time at which turbidity was observed to sharply increase.

Chapter 5

Investigation of Protein Behavior

in Solution by Light Scattering

1. Introduction

Light scattering is an optical method suitable to detect the solution properties of protein molecules.[185] Static light scattering (SLS) can be used to characterize the molecular weights of proteins and the second virial coefficient (related to protein-protein interaction) by measuring the temporal average scattered intensity at different protein concentrations.[56] Dynamic light scattering (DLS) can be used to determine diffusion coefficients of protein species by measuring the temporal fluctuations of scattered intensity.[185-187] For dilute solutions, the hydrodynamic radius of macromolecules can be calculated using the Stokes-Einstein equation (see Section 4).

DLS measurements are noninvasive, nondestructive, relatively fast, and very sensitive to protein aggregates. However, the drawbacks of this method are its fairly poor resolution, low sensitivity to chemical nature, and relatively complex theory.

2. Apparatus

Both static and dynamic light experiments were performed on a light-scattering apparatus (Fig 1) built using the following main components: He-Ne laser (35 mW, 632.8 nm, Coherent Radiation), manual goniometer and thermostat (Photocor Instruments), multi-tau correlator, APD detector and software (PD4042, Precision Detectors).

For DLS measurements, the laser needs to meet several criteria. It must have stable and sufficient intensity. It must generate a single fundamental transverse electromagnetic
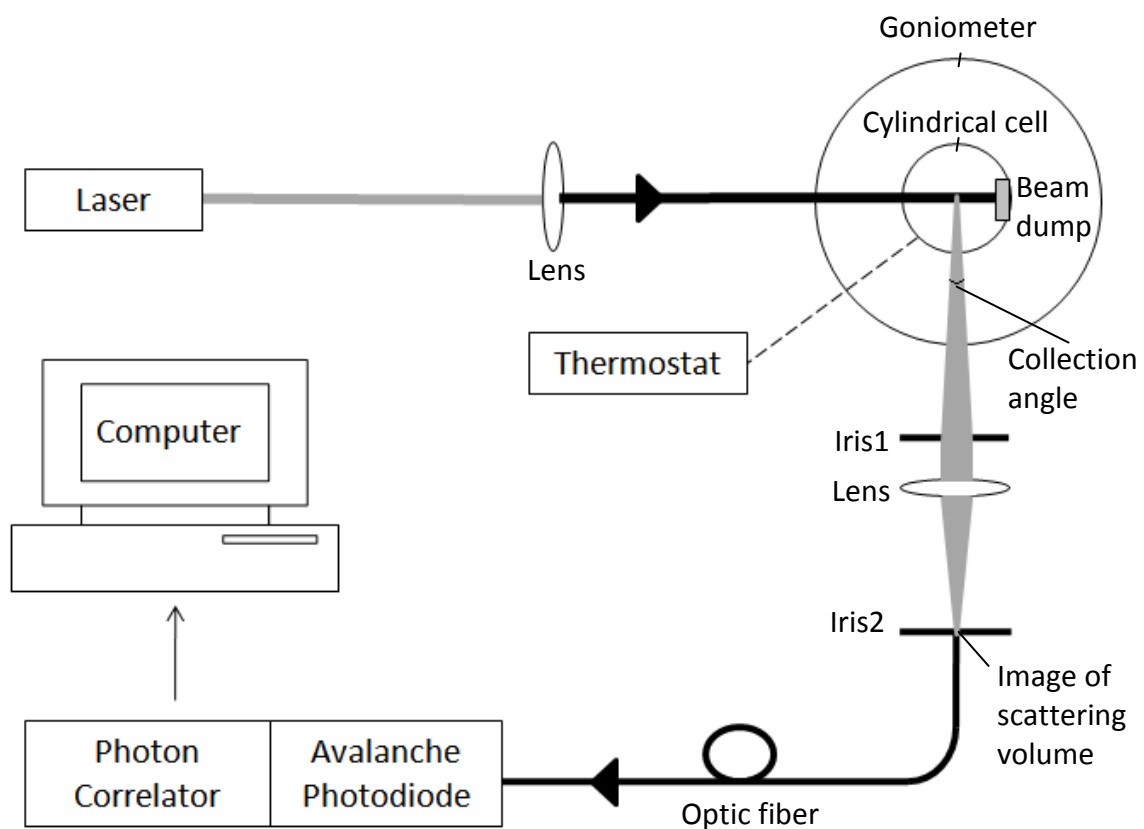
Figure 1. Scheme of custom-built light scattering system.

(A)   (B)
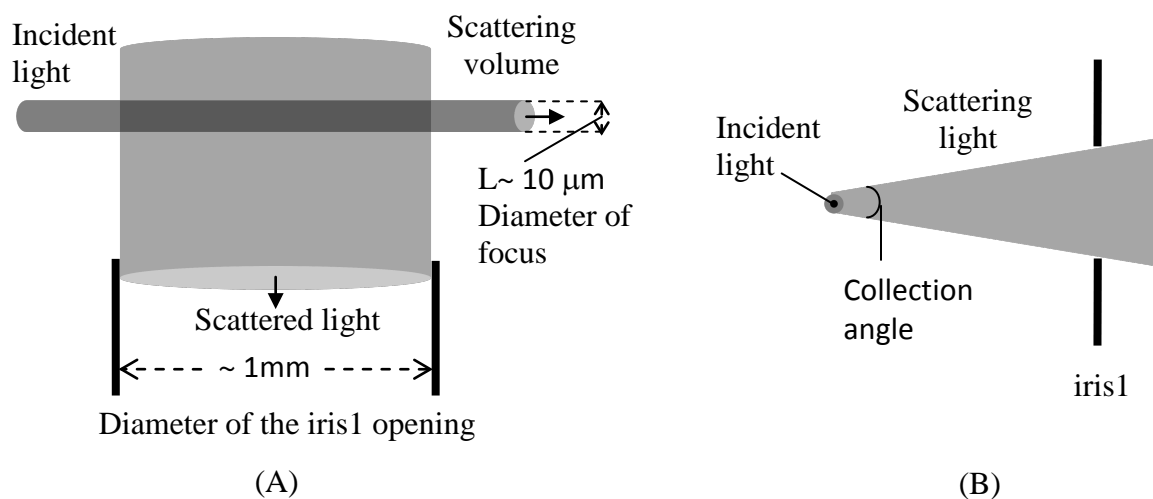
Figure 2. (A) Top view of the scattering volume; (B) View of scattering light in the direction of incident beam.

mode (TEM$_{00}$ mode), i.e. a true Gaussian beam. Multiple-mode laser introduces optical beating noise, i.e. the intensity fluctuation caused by the small frequency difference between two transverse modes. Moreover, the laser should be polarized perpendicular to the plane formed by light scource, sample, and photodetector, so that the scattering intensity is maximized.

The laser beam is well focused by a lens onto the center of the sample. The width of the beam at the focus point is about 10 μm. Generally, the narrower the beam diameter, the better signal-to-noise ratio is, due to the smaller coherent angle ($\sim\lambda/2\pi L$, where λ is the wavelength of laser, L is the cross-section length of scattered volume in the scattering direction shown in Fig 2). However, a narrowly focused incident beam may not be suitable for studying strongly scattering objects, because in this case the intensity fluctuation not only depends on concentration fluctuation inside the scattering volume but also on fluctuations of the concentration of the whole scattering volume.

The scattered light is collected by a condenser lens from the scattering volume within a particular solid angle controlled by an iris placed between the sample and the lens. The larger this angle the more scattered intensity is registered. However, if this angle becomes larger than the coherence angle, the light from more than one construction speckle in the interference pattern of the image of scattering volume will be registered. Since the temporal intensity fluctuations from two different speckles are independent and will be averaged out, increasing the collecting angle beyond the coherent angle does not lead to improvement in the signal-to-noise ratio. The scattered light is focused towards the tip of optical fiber with a pinhole in front of it minimizing the stray light. The avalanche photodiode detector receives the photons and converts them to electrical pulses,

which are processed and analyzed by the 256 channel correlator. The correlation functions are calculated every 2 seconds and averaged over 100-3200 accumulation times. The precision of correlation function increases with the square root of accumulation number. Therefore, a high accumulation number is preferred for the weakly scattering solute, while the low accumulation number is required for studying fast kinetic process. The diffusion coefficients, D, were obtained from the average correlation function using the software Precision Deconvolve32. The temperature of the cylindrical cell is controlled by a thermostat with the accuracy of $\pm 0.1°C$. All measurements were performed at a scattering angle of 90° at 25.0 ± 0.1 °C.

In the case of laccase, which absorbs at around 600 nm, the diffusion coefficient for the same sample was measured with different incident intensities to test the effect of the local temperature elevation due to the absorbance of laccase.

Dust and bubbles strongly scatter light and will distort the correlation function significantly. Thus, the solution must be optically clean. All samples were filtered through a 0.02 μm filter (Anotop 10, Whatman) and placed in a test tube. When protein aggregates are generated fast in the solution, the corresponding spikes of intensity were removed by the software dust filter.

3. The Correlator

In DLS measurements, the temporal intensity of the scattered light is registered by the photodetector. All informations are contained in the temporal correlations in the signal. The intensity correlation function $G^{(2)}(\tau)$ is defined as follows: [185-187]

$$G^{(2)}(\tau) \equiv\, <I(t)I(t+\tau)> = \lim_{T\to\infty} \frac{1}{T}\int_0^T I(t)I(t+\tau)dt \qquad (1)$$

This continuous intensity correlation function is approximated with discrete points obtained by a summation over the duration of the experiment. This summation is performed by the correlator, a circuit board composed of various logic chips and operational amplifiers which continuously multiplies and adds measured intensity values. An important parameter for a correlator is its speed, i.e. its ability to add up numbers fast, or inversely expressed, the smallest delay time α that can be handled. Using a time increment of α, a grid of experimental intensity data values can be obtained. At beginning of the experiment, $t_0 = 0$. The first time then is at $t_1 = \alpha$, the second time is at $t_2 = 2\alpha$, and so on. The stream of numbers can be displayed in an array ending with $t_n = n\alpha$.

The correlator then evaluates how these intensities are related to each other by comparing each value to another corresponding to a defined later time. This comparison is carried throughout the duration of the experiment. The correlation function can be expressed by the following summation $G_k(\tau_k)$,

$$G_k(\tau_k) = \frac{1}{n}\sum_{i=1}^{n} I(t_i)I(t_i + \tau_k) \qquad (2)$$

The series of $G_k(\tau_k)$ values are then normalized respect to the square of the average intensity and generate the series of $g^{(2)}(\tau) = G^{(2)}(\tau)/\langle I \rangle^2$. For standard light (with Gaussian statistics $g^{(2)}(\infty) = 1$) and detectors collecting light from a significantly small angle, this normalization imposes a theoretical limit of $g^{(2)}(0) = 2$. However, even carefully optimized optical systems cannot achieve $g^{(2)}(0) = 2$. For a random signal and a large measurement time, all G values would add up to roughly the same number leading to $g^{(2)} = 1$ independent of time. The signal would then be described as 'uncorrelated', and the resultant intensity correlation curve would look like random fluctuation about the baseline. For light scattered from diffusing molecules, however, the intensity correlation curve usually exhibits an exponential decay (Fig 3), indicating that the signal is correlated.
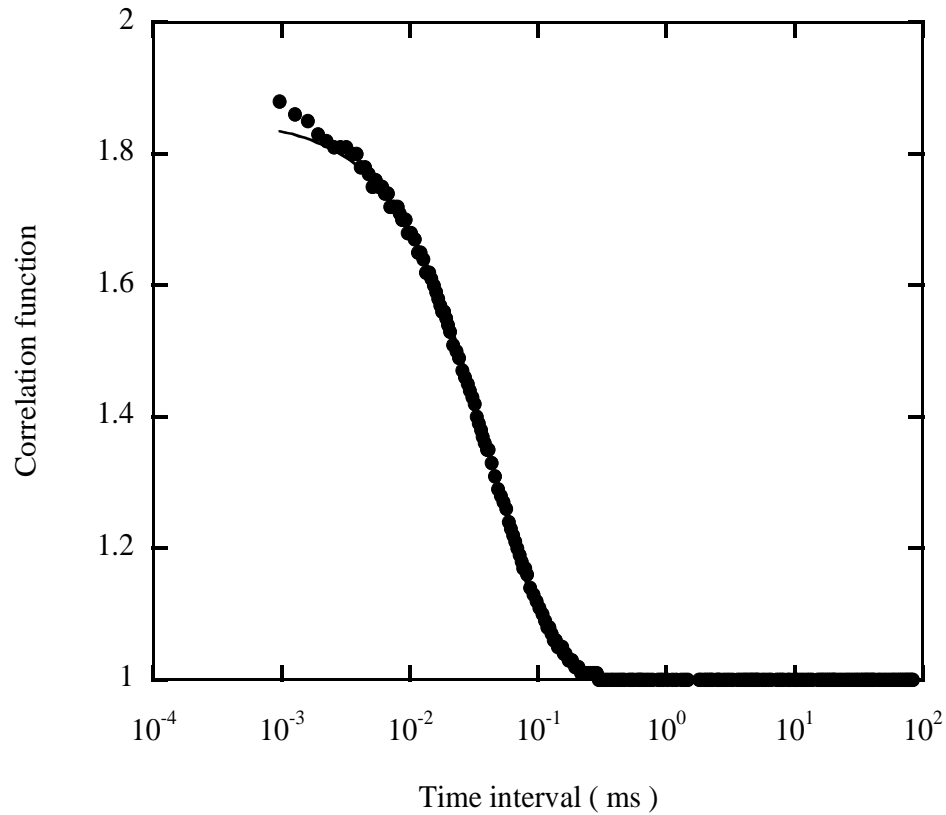
Figure 3. Normalized intensity correlation function of 10mg/mL BSA in 0.1M acetate buffer at pH 5.2.

In all the DLS experiments, the correlation functions were accumulated for many times and averaged. The experimental error decreased with the square root of the accumulation number. For samples in equilibrium, we used an accumulation number of 800. For kinetic study, we used accumulation number 100. For samples with very low scatterin, the maximum accumulation number 3200 was used.

4. Correlation Function Analysis

The correlation function for monodisperse and polydisperse samples was fitted to a multi-exponential decay. Specifically, the normalized-intensity correlation function is written as follows, [185-187]

$$g^{(2)}(\tau) = B + A(\sum_i W_i e^{-D_i q^2 \tau})^2 \qquad (3)$$

where $q = 4\pi n \sin(\theta/2)/\lambda$ is the scattering vector, n is the refractive index ($n=1.33$ for aqueous solutions), $\lambda = 632.8$nm and $\theta = 90°$ is the scattering angle, D_i is the diffusion coefficient of species i and W_i is the normalized contribution of species i to the scattering intensity, which is proportional to the molarity, c_i, and the square of molecular weight, M_i. The goal of the mathematical analysis of DLS data is to reconstruct $g^{(2)}(\tau)$ as precisely and accurately as possible the distribution function W_i from the experimentally measured correlation function. However, the corresponding mathematical minimization problem, an inverse Laplace transformation, is "ill-posed". This means that significantly different distributions W_i may lead to nearly identical correlation functions and therefore are

equally acceptable fits to the experimental data. There are three approaches to solve this ill-posed problem.

The simplest approach is the direct fit method. Here the functional form of W_i, i.e. $W(D)$, is assumed *a priori* (single modal, bimodal, Gaussian, etc). The parameters of the assumed function that lead to the best fit to the experimental data are then determined. This method is only as good as the original guess of the functional form of $W(D)$. Moreover, the method can be misleading because it tends to "confirm" any *a priori* assumption made. It is also important to note that the more parameters there are in the assumed functional form of $W(D)$, the better the experimental data can be fit but the less meaningful the values of the fitting parameters become. In practice, typical DLS data allow reliable determination of about three independent parameters of the size distribution of scattering particles.

An assumption-free approach is the cumulant method. In this approach, the focus is not on the shape of the distribution but instead on the moments of the distribution, or closely related quantities called cumulants. Cumulants, κ_n, are generated by a cumulant generating function defined as follows,

$$h(\tau) = \ln\left(\sum_i W_i e^{-D_i q^2 \tau}\right) \equiv \sum_{n=1}^{\infty} \kappa_n \frac{\tau^n}{n!} \qquad (4)$$

$$\kappa_n = g^{(n)}(0) = \left(\frac{\partial^n h(\tau)}{\partial \tau^n}\right)_{\tau=0}$$

From this definition, it is easy to know that κ_1 equals to the first raw moment of the $W(D)$, $\frac{\mu_1'}{q^2} = \sum_i W_i D_i = <D>_z$, i.e. the z average of D, and κ_2 equals to the second central moment of the $W(D)$, $\frac{\mu_2}{q^4} = \sum_i W_i (D_i - <D>_z)^2 = \frac{\mu_2' - (\mu_1')^2}{q^4} = <D^2>_z - <D>_z^2$, i.e.

the variance of D. Therefore, the average diffusion coefficient $< D >_z$ can be determined from the initial slope of the $h(\tau)$. The variance of D can be determined from the curvature of the initial part of $h(\tau)$. As in the direct fit method, the accuracy of the real DLS experiment allows determination of at most three moments of the distribution. The first moment, $< D >_z$, can be determined with less than 1% error. The second moment, the width of the distribution, can be determined with an error of 5 to 10%. The third moment, which characterizes the asymmetry of the distribution, can be determined with an error about 100%.

The third approach, which is used in the software Precision Deconvolve32, is based on the regulation algorithm. It combines the advantages of both of the previous methods. It assumes that the distribution $W(D)$ is an arbitrary, but smooth, non-negative function. The regularization requirement of smoothness precludes spikes in the distribution, allowing unique solutions to the minimization problem. In this algorithm, the distribution of D depends on the value of smoothness parameter. A small smoothness number gives more details of the distribution (Fig 4a), while a large smoothness number produces a more stable and reproducible distribution (Fig 4c). The rule of thumb is to choose as small a smoothness number as possible, which maintains the reproducibility of the D distribution (Fig 4b). In Fig. 4, the bimodal distribution functions of a lysozyme-PEG20k solution are presented. The distribution in Fig. 4a was barely smoothed. This spiky distribution is the best fit to the measured correlation function, but it is not stable and varies from one experiment to another. The same data with an appropriately chosen smoothing parameter is shown in Fig 4b. This distribution fits the measured data only marginally worse than that in Fig 4a, but is stable from one experiment to another. In Fig

4c, an excessively smoothed distribution was presented. It is stable, but lacks details and fits the measured data noticeably worse than the previous two. Also, we noted that the average diffusion coefficient and the overall width of the reconstructed distribution are affected little by the choice of smoothing parameter.
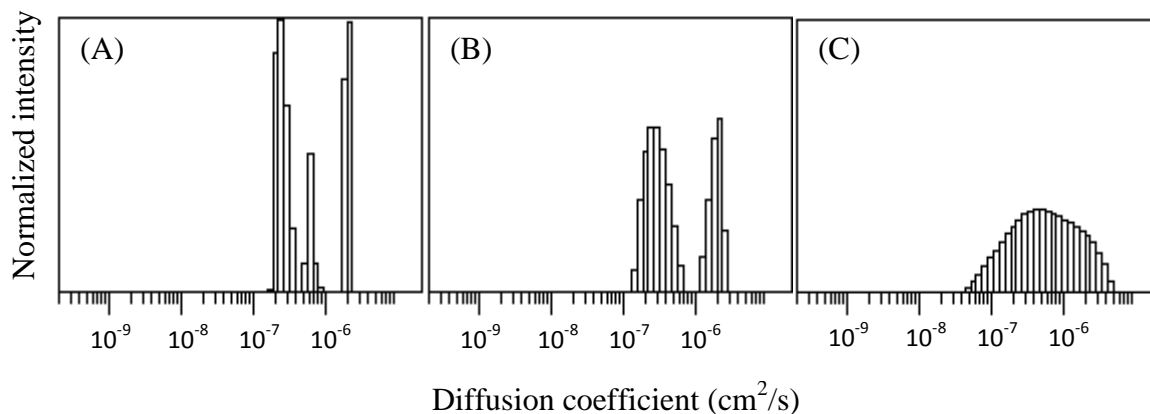
Figure 4. Scattering distribution of diffusion coefficients of lysozyme and PEG20K in 0.1M acetate buffer at pH 4.5. Distributions determined by regulation method with different smoothness. (A) smoothness = 1. (B) smoothness = 12. (C) smoothness = 25.

Using the regulation algorithm, the diffusion coefficients of proteins (BSA, lysozyme, and laccase), PEG and aggregates were determined. The hydrodynamic radii of a given species, R_h, were then calculated from the corresponding D value using Stokes-Einstein equation, [185-187]

$$R_h = \frac{kT}{6\pi\eta D}$$

(5)

where k is the Boltzmann constant, T is temperature (298K) and η is the viscosity of the aqueous buffer.

5. Effect of sample absorbance on diffusion coefficients determined by DLS

Our DLS experiments were performed using a He-Ne laser with a wavelength of 633 nm. However, laccase in aqueous solution has an absorbance band around 600 nm. This implies that laccase absorbance at the laser wavelength is appreciable. Sample absorbance during a DLS experiment may lead to an increase of local temperature in the illuminated volume. This, in turn, increases the value of the measured diffusion coefficients. To assess the role of sample absorbance, we have measured the diffusion coefficient of laccase as a function of the intensity of the incident laser (Fig 5). Due to sample absorbance, we have indeed observed that the scattering intensity for laccase samples is significantly lower than that observed in the case of non-absorbing BSA samples in the same experimental condition. Nonetheless, as shown in Fig 5, no correlation was observed between the extracted value of D and the observed scattered intensity. We have therefore concluded that laccase absorbance has a negligible effect on the accuracy of the determined diffusion coefficients.
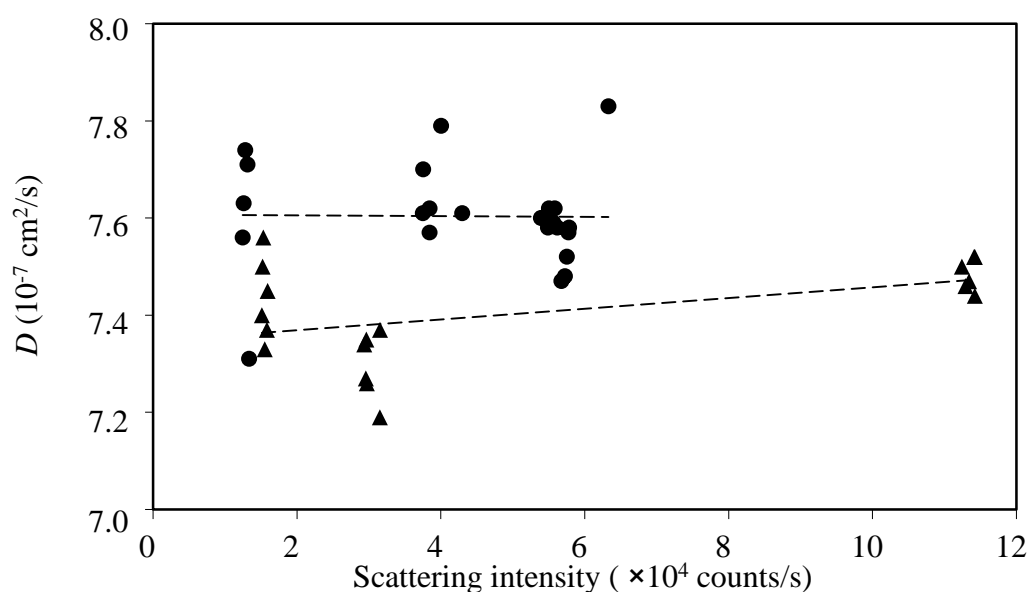
Figure 5. Diffusion coefficients, D, of laccase in 0.1 M acetate buffer at pH4.5. The circles and triangles denote diffusion data obtained at a laccase concentration of 22 mg/mL and 15mg/mL respectively. The dashed lines are guides for the eye.

6. Static Light Scattering

The average scattered intensities of dilute protein solutions with different concentrations were measured at scattering angle $\theta = 90°$ and temperature 25.0 ± 0.1 °C. The average intensities were plotted against protein concentrations. The second virial coefficient was calculated from the initial slope of this plot using the following equation,[56, 116]

$$\frac{kc}{R_{90°}} = \frac{1}{M}(1 + 2BMc + \cdots) \tag{6}$$

where B is the second virial coefficient, c is the mass/volume concentration of protein, M is the molecular weight of protein, $k = 4\pi^2 n_0^2 (\mathrm{d}n/\mathrm{d}c)^2/(N_A\lambda^4)$, n_0 is the refractive index of buffer, $\mathrm{d}n/\mathrm{d}c$ is the refractive index increment (for proteins, $\mathrm{d}n/\mathrm{d}c=0.185$ ml/g), N_A is the Avogadro's number, λ is the laser wavelength in vacuum, $R_{90°}$ is the excess Rayleigh ration at 90°.The value $R_{90°}$ were obtained from,

$$R_{90°} = \frac{(I_s - I_{s,0})}{I_{s,R}}\left(\frac{n_0^2}{n_R^2}\right) R_{90°,R} \tag{7}$$

where I_s is the scattered intensity of the solution, $I_{s,0}$ is the scattered intensity of the buffer, $I_{s,R}$ is the scattering intensity of toluene (the standard), and n_R and $R_{90°,R}$ are, respectively, the toluene refractive index (1.492) and its Rayleigh ratio (10.31×10^{-6} cm^{-1} at 633 nm).

Chapter 6

Characterization of Protein

Condensed Phases by Microscopy

1. Introduction

Microscopy (light and electron) is a direct way to visualize small objects. It is suitable for the characterization of protein condensed phases, since the size of protein condensed phases (> 100nm) are usually larger than the resolution limit of microscopes. The size and morphology of protein condensed phases larger than 1μm can be investigated with light microscopy, whose resolution is around 500 nm limited by the diffraction of visible light.[188] The size and morphology of a smaller protein condensed phase needs to be studied with electron microscopy, whose resolution may become lower than one nanometer. Light microscopy has the advantage over electron microscopy of being non-invasive, because it requires neither sample preparation nor solvent removal. The advantages of electron microscopy are higher resolution and the ability to perform microchemical analysis on the sample with the help of energy dispersive X-ray spectroscopy.

2. Light Microscopy

After phase separation, around 1μL suspension containing protein condensed phase was promptly applied on a microscopic slide, covered by a cover slip and observed under a light microscope (Axioskop 40, Zeiss). The transmitted-light bright field was set according to Köhler illumination. The 2× eyepieces and the 10×, 20×, 40× and 63× objectives were used. Since many of the protein condensed phases lack sufficient contrast,

they were studied using transmitted phase-contrast technique with a phase stop in the front focal plane of condenser and a corresponding phase ring in the back focal plane of the objective. This setup increased the contrast significantly and resulted in clear images. However, this technique was not suitable for thick samples because a halo was formed around the objects obscuring the detail of shape. Protein crystals were carefully picked up by a 25μL ultra precise pipette with pistol. Birefringence of protein crystals was detected by the transmitted-light polarization techniques with a pair of crossed polarizer (polarizer ⊥ analyzer). In these conditions, the crystals were brightened while the surroundings were dark. Images were taken using a digital camera (Axiocam MRc, Zeiss) interfaced by a computer with software (Axiovision AC 4.5, Zeiss). The dimensions of protein condensed phases were determined using the ImageJ software.

3. Electron Microscopy

Images of protein condensed phases were also taken using the scanning electron microscopy (JSM-6100, JEOL). The electron microscope works with a high vacuum, and the image will be obscure when the sample contains even a small amount of water. Thus, a drop of sample was spread on the carbon tape, which was stuck to the top of the aluminum specimen mount, and dried in the vacuum box (LAB-LINE Instruments, Inc) overnight. For the characterization of the size and morphology of protein condensed phases, the dried sample was coated with an ultrathin (10 nm) coating of gold. The gold is conductive and prevents charging of the sample when scanned by the electron beam,

which is the main reason for the poor resolution of the images of nonconductive samples, especially in the secondary electron mode. The working distance of the microscope was set at 15mm. A representative SEM image of lysozyme microspheres is shown in Fig 1.

The microchemical analysis was also done by detecting the characteristic X-ray of the elements in the sample struck by the electron beam. For this application, the sample was studied without coating gold in order to avoid the disturbance of strong gold peak. The working distance of the microscope was set at 30 mm for element analysis. Energy dispersive X-ray spectra (EDS) were obtained with the Ge detector, Oxford Inst. An example of microchemical analysis on the sulfur content (EDX mapping of Kα line of S) in lysozyme microspheres is shown in Fig 2.
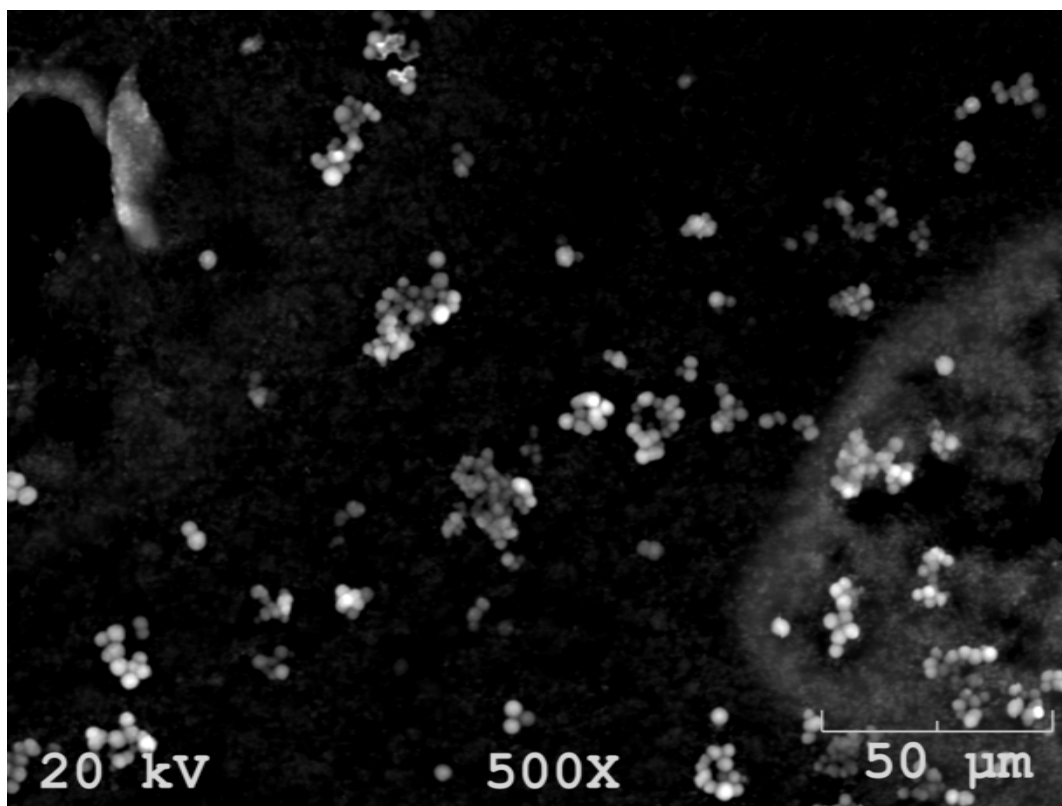
Figure 1. SEM image of Cross-linked lysozyme microspheres.

(A) SEM   (B) Sulfur Distribution
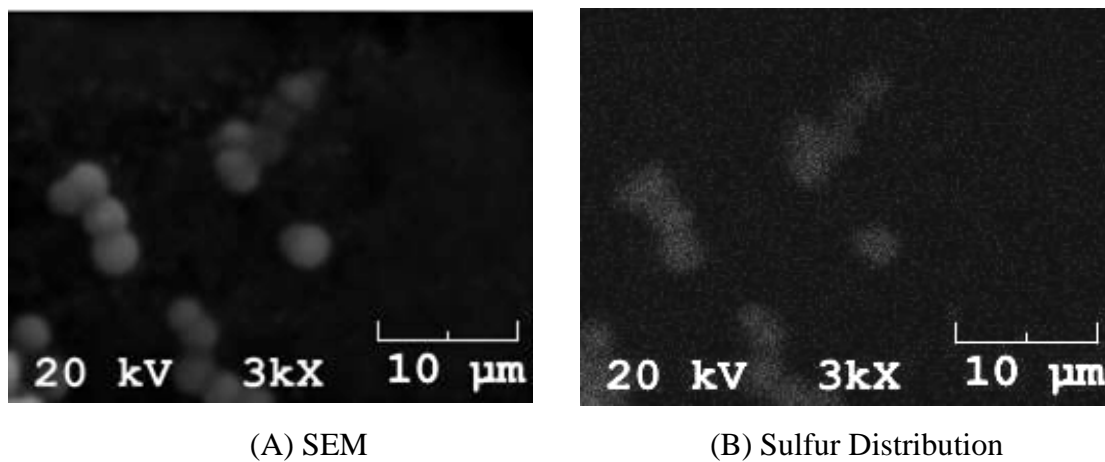
Figure 2. Cross-linked lysozyme microspheres. (A) SEM image.
(B) Sulfur distribution determined by EDS mapping.

Chapter 7

Monte Carlo Simulation on

Model Sphere-Coil Systems

1. Introduction

To understand the effect of non-absorbing polymer on protein condensations, we have used excluded volume models.[189-192] The key quantity of these models is the volume fraction available to the solute particles inside the system (i.e. free volume fraction).[189-196] Solute chemical potentials can be quantitatively described by expressions containing the free-volume fraction. We have used Monte Carlo simulations to understand the free-volume of polymer in the protein solution and to study the effects of temperature and the polymer structure.

2. Monte Carlo Simulation

To design the model system for simulation, we consider a suspension of N spheres with diameter σ inside a cubic box with volume V. The interactions between hard spheres are described by the potential energy $u(r)$ for a pair of spheres whose centers are separated by a distance r:

$$u(r) = \begin{cases} +\infty, & \text{for } r < \sigma \\ 0, & \text{for } r \geq \sigma \end{cases} \qquad (1)$$

When it was required to describe attractive interactions, we have used the following square-well potential:

$$u(r) = \begin{cases} +\infty, & \text{for } r < \sigma \\ -\epsilon, & \text{for } \sigma \leq r < \lambda\sigma \\ 0, & \text{for } r \geq \lambda\sigma \end{cases} \qquad (2)$$

where λ is the reduced range of the potential well and ϵ is its depth. With this potential, we calculate the internal energy E, as a function of protein volume fraction and temperature. This is achieved by counting the number of sphere-sphere contacts, N_{con}; i.e., the number of sphere pairs whose distance between the centers is within the range $\sigma \leq r < \lambda\sigma$. The internal energy is calculated as $E = -N_{con}\epsilon$. The main objective of our simulations was to determine free-volume inside a suspension of spherical particles (representing the protein molecules) upon the insertion of a test particle (representing a polymer molecule).

Monte Carlo simulations were performed in Fedora7 terminal on the HP xw8400 Workstation with Dual-Core Intel® Xeon® processor 5130 2.00GHz 4 MB L2 cache 1333 and 1GB 667 MHz DDR2 SDRAM. The programs was compiled using the g++ 4.1 (GNU project C and C++ compiler).

To perform a Monte Carlo simulation, we initially placed N spheres in the FCC close-packing configuration inside a cube with unit volume and periodic boundary conditions. The simulations have been performed at constant volume fraction ϕ and temperature T. The corresponding hard core diameters, σ, of the spheres were calculated from the equation, $\phi = \pi\sigma^3 N/6$. In order to reach the thermodynamic equilibrium, a randomly chosen sphere was displaced using the well-established NVT Metropolis algorithm. In this algorithm, the displacement of a sphere was accepted unconditionally if the change in the total internal energy of the system ΔE, due to the displacement, is negative. The displacement was accepted with a probability exp(-ΔE/kT) if ΔE is positive. The step length of the displacement was chosen as the value that leads to around 50% acceptance of the displacement. This corresponds to the fastest approach to equilibrium

(i.e. the thermodynamic quantities fluctuate around their expectation value). The displacement of spheres was repeated K_0 times until equilibrium was reached, and then continued $K_{\max}$ times to generate a statistic ensemble of sphere configurations. For each configuration of this ensemble, we inserted a particle into the hard sphere suspension. In our study on the temperature effect on free-volume fraction, the inserted particle was an AO-sphere with a diameter of $q\sigma$, where q is the ratio of inserted particle to sphere diameter. A particle is accepted if it does not overlap with any sphere. The probability of acceptance is the free-volume fraction α for the particle in the sphere suspension at volume fraction ϕ and temperature T.

In order to study the effect of polymer structure on the free-volume, a coil was modeled by linear chains built with parameters taking in account real structural factors to different extent (e.g. bond length, bond angle, atomic volume, reduced trans-gauche transition energy) (Chapter 9). In order to examine the free-volume fraction behavior of coils compared to spheres we set the intermolecular potential to zero. The free-volumes for the coils in the sphere suspension were determined in the same way as before. However, here we also calculate the radius of gyration of the coil, R_g, for each configuration of polymer chain, and the quadratic mean of R_g over all the accepted coil configurations, $R_g^{RMS} = \sqrt{\langle R_g^2 \rangle}$. This was taken as the ensemble average gyration radius of coil in the sphere suspension at volume fraction ϕ. We note that α and R_g^{RMS} are equal to the ensemble average when a statistical number of coils have been successfully inserted into a given hard-sphere configuration, and the insertion process has been repeated for a statistical number of hard sphere configuration. However, because inserting coil and changing sphere configuration are two statistically independent

processes, we can conveniently compute α and R_g^{RMS} by inserting only one coil for each generated sphere configuration. The simulation errors of α and R_g^{RMS} are estimated by three times of their standard deviation,

$$\text{error}(\alpha) = 3\sqrt{\frac{(v - v^2)}{K - K_0}} \tag{3}$$

$$\text{error}\left(R_g^{RMS}\right) = \frac{3}{2R_g^{RMS}}\sqrt{\left(\frac{R_g^4}{P} - \left(\frac{R_g^2}{P}\right)^2\right)/P} \tag{4}$$

where P is the total number of the successful attempts of adding polymer. The detailed flowchart Monte Carlo algorithm is given in Appendix G. The C program of this Monte Carlo algorithm for the self-avoiding polymer chain is given in Appendix H.

3. Consideration of Parameters for Simulation

We performed Monte Carlo simulation within the single phase domain. All the simulations have been performed at volume fractions below 0.49 in order to avoid the fluid-solid transition that occurs when the volume fraction of monodisperse hard sphere fluids are larger than this value. Our highest experimental volume fraction of protein is well within the volume fraction range of our simulations.

When considering the temperature effect for the spheres with the square-well potential, we performed simulations with ranges of interaction λ's larger than 1.05 since

"sticky" spheres with very short λ would take impracticable long time to reach equilibrium. For a given value of λ, the depth of potential well, ϵ, was chosen to be lower than that corresponding to critical conditions, so that no phase transition would occur.

The number of spheres, N, used for our simulation has to be large enough to virtually eliminate systematic errors associated with the periodic boundary conditions. To test the effect of N on α in our simulation, we have conducted simulations with different N values. For hard spheres, the results (Fig 1a) were independent of the number of spheres within the simulation error, when $N \geq 125$. When using the square-well potential, the results become independent of N when $N \geq 1000$. We therefor e choose $N = 125$ for simulation on hard spheres and $N = 1000$ for simulation on spheres with attractive potential.

The spheres were randomly selected, and moved with a specified step length according to Metropolis equilibration algorithm. The step length of sphere displacement was estimated by $\delta = \frac{1}{2}\left(\sqrt[3]{1/N} - \sigma\right)$ to get 50% successful attempts of moving spheres and therefore minimize the time of equilibration.[197] The cut-off number of attempts, K_0, for reaching equilibrium was set as 10^6 and the total number of attempts, K_{max}, of moving sphere was set as 10^8 (Fig 2).
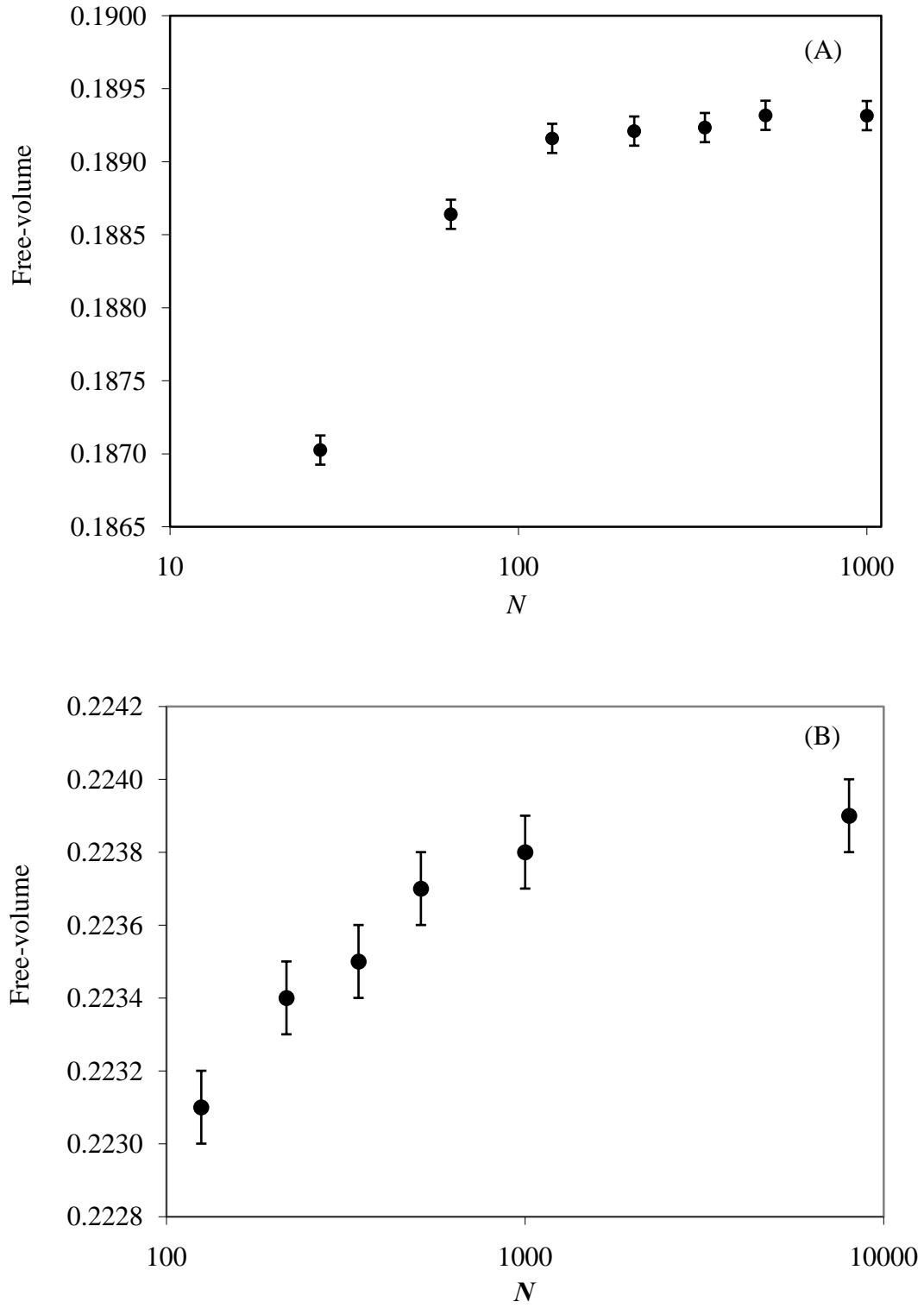
Figure 1. The free-volume correlation to number of sphere being tested. $\phi = 0.3$, $q = 0.5$, (A) Hard-spheres; (B) $\lambda = 1.05\sigma$, $\epsilon = 2kT$.
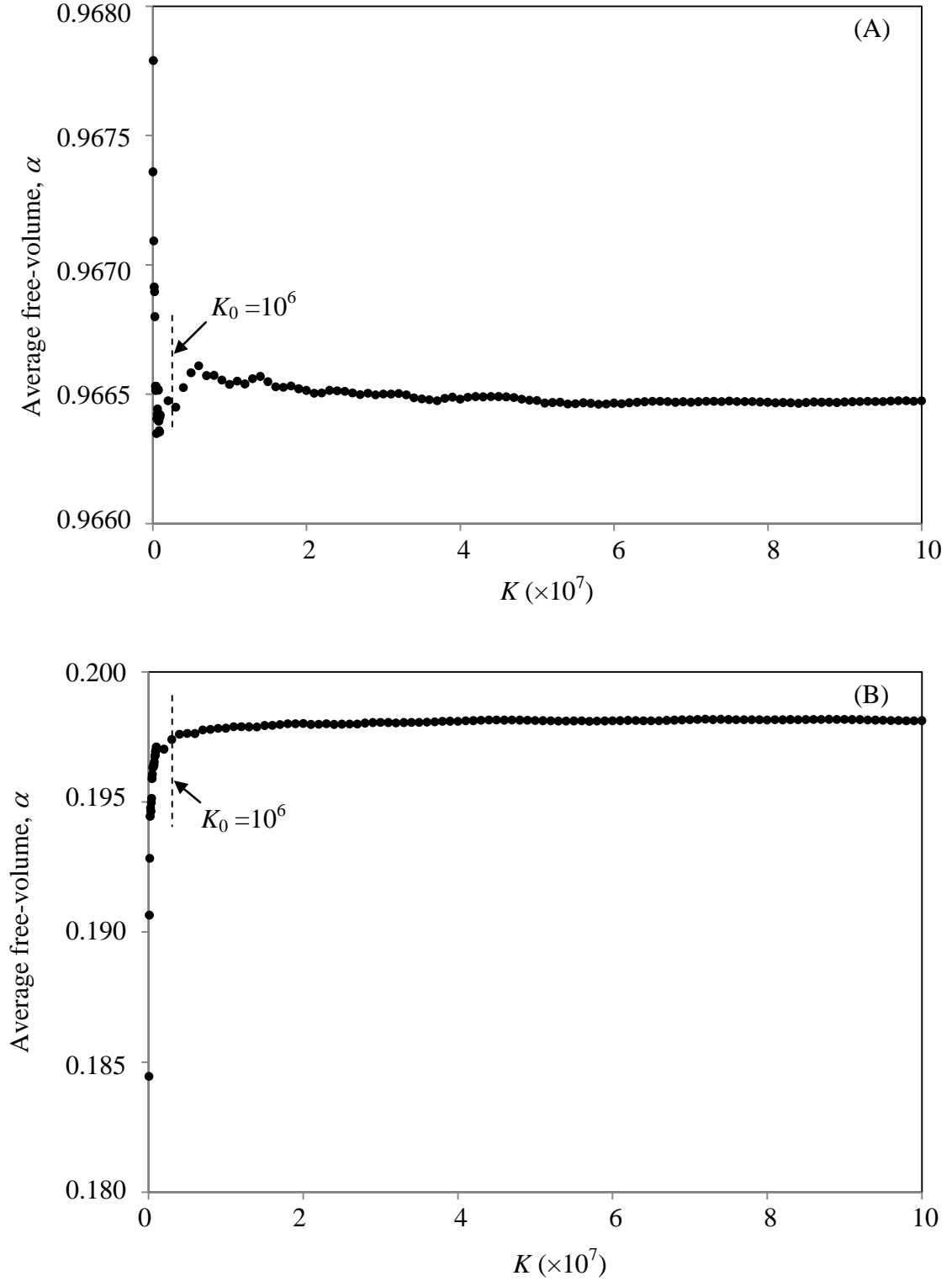
Figure 2. The average free-volume of the system as a function of number of attempts shows the system reaches equilibrium after 10^6 attempts. (A), $\phi = 0.01$, $\lambda = 1.05\sigma$, $\varepsilon = 2kT$, $\delta = 0.036635$; (B) $\phi = 0.3$, $\lambda = 1.5\sigma$, $\varepsilon = 0.5kT$, $\delta = 0.008472$.

Part II:

Results and Discussions

Chapter 8

LLPS of the BSA-PEG-buffer System

1. Introduction

Although LLPS for protein-buffer binary systems are rarely observed, this transition can be effectively induced by adding non-absorbing polymers. Adding polymers to protein solutions introduces protein-polymer interactions and affect protein-protein interactions. Thermodynamic perturbation theory can be used to relate the effect of polymer concentration on protein-protein interactions to the dependence of LLPS temperature on polymer concentration and protein-polymer interactions to isothermal protein-polymer partitioning in the two coexisting phases.

In this chapter, we report an experimental investigation of the liquid-liquid phase separation (LLPS) of aqueous bovine serum albumin (BSA) in the presence of relatively small amounts of PEG with an average molecular weight of 1450 g/mol (PEG1450). LLPS of initially stable BSA-PEG1450-buffer mixtures has been observed when temperature is lowered. We experimentally characterize two thermodynamically independent properties of the phase boundary: (1) the effect of PEG1450 concentration on the LLPS temperature (Section 4), (2) BSA-PEG1450 partitioning in the two liquid coexisting phases (Section 5). The reliability of a depletion-interaction model is then examined for this system by investigating its ability to describe both experimental properties (Sections 6-8). The internal consistency between the experimental results on two thermodynamically independent properties in complex systems would be a valuable tool for supporting or rejecting existing microscopic models and provide insight for developing more accurate ones. Our experimental results were also examined in view of microscopic structural considerations (Section 8).

Finally, due to the general importance of LLPS of protein solutions, we will also show how phase separation can be induced when the temperature of a protein-PEG-buffer solution is either lowered or increased; i.e., two liquid-liquid phase transitions can be observed for the same mixture (Section 9).

2. The BSA-Buffer Binary System

The chosen buffer for these studies was a 0.1 M sodium acetate aqueous solution at pH 5.2. At this pH, BSA is close to its isoelectric point. Because the protein net charge is nearly zero, Donnan effects with sodium acetate and protein-protein repulsive electrostatic interactions should not be significant. This also suggests that the amount of PEG1450 required to induce LLPS at this pH is close to the minimum.

To characterize protein-protein interactions in the protein-buffer binary system, we determine the value of its second virial coefficient, B_2, at 298.15 K. In Fig 1, we report static light-scattering measurements on the BSA-buffer binary system.

We find that $B_2 = -0.1 \pm 0.1 \times 10^{-4}$ mL mol g^{-1}. The value calculated for the corresponding hard-sphere system is 0.44×10^{-4} mL mol g^{-1}. Thus our results provide a strong indication that the protein-protein interaction energy is attractive in this binary system. However, protein-buffer systems do not undergo LLPS within the experimental temperature range, 260-310 K, and for protein concentrations as high as 400 mg/mL. This is consistent with the hypothesis that the binary system behaves as a "supercritical fluid".
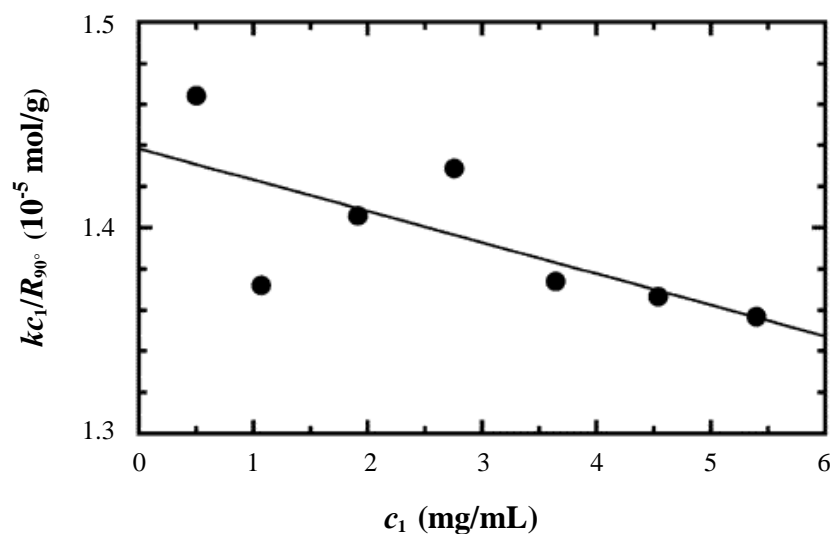
Figure 1. Static light-scattering data for the BSA-acetate buffer system at pH 5.2 and 298.15 K. The second-virial-coefficient value, B, is $-0.1 \pm 0.1 \times 10^{-4}$ mL mol g^{-1}, $kc_1/R_{90}{}^{\circ} = 1/M_1 + 2Bc_1$ (see Chapter 5 for details).

3. Thermodynamic Perturbation Theory

Adding non-absorbing polymers has shown to be an effective way to induce LLPS of protein solutions. We will now outline basic thermodynamic properties that will be used to describe the liquid-liquid phase transition for the protein-polymer-buffer system.[55,85] This provides the basis for the interpretation of protein-PEG interactions and the effect of PEG concentration on protein-protein interactions.

We describe the composition of this system by the protein molar concentration c_1 and the polymer molar concentration c_2. The buffer is assumed to be one pseudo-component. To analyze the physical factors that determine protein-polymer interactions and the effect of the polymer on protein-protein interactions, we consider the reduced

Helmoholtz free energy $\hat{f} = (\tilde{F} - c_1\tilde{\mu}_1^0 - c_2\tilde{\mu}_2^0)/RTV$ of the system as described in Chapt 1 and assume that the amount of added polymer is relatively small. To a first-order approximation with respect to c_2, we can write:

$$\hat{f}(c_1, c_2, T) = \hat{f}'(c_1, T) + c_2\ln\frac{c_2}{e} + c_2\left(\frac{\partial \hat{f}^{ex}}{\partial c_2}\right)_{T,c_1,c_2=0} \tag{1}$$

where $\hat{f}'(c_1, T) \equiv \hat{f}(c_1, 0, T)$ is the reduced free energy for the protein-buffer binary system. The quantities $c_2\ln(c_2/e)$ and $c_2\left(\partial\hat{f}^{ex}/\partial c_2\right)_{T,c_1,c_2=0}$ are the contributions to the reduced free energy associated with the replacement of solvent molecules by polymer molecules. The contribution of polymer to the translational entropy of the system is represented by $c_2\ln(c_2/e)$ (Chapter 1), whereas $c_2\left(\partial\hat{f}^{ex}/\partial c_2\right)_{T,c_1,c_2=0}$ is the first term of a series expansion describing the effect of polymer concentration on the excess free energy, $\hat{f}^{ex}$. For ideal polymer coils, the higher terms of the series disappear.

Although this theory makes no assumption on the nature of the protein-protein and protein-polymer interaction, we will consider the form of eq 1 in the case of the depletion-interaction model.[191,192,198] For proteins treated as spherical particles in the presence of non-adsorbing polymers, depletion interactions become the only source of thermodynamic nonideality. Due to steric hindrance, the center of mass of a polymer coil is not only excluded from the volume occupied by the protein molecules but also from a region surrounding them referred as the depletion layer.[86] The width of this layer is proportional to the gyration radius of the polymer coil. For pure depletion interactions: $\alpha(c_1) = \exp\left(-\left(\partial\hat{f}^{ex}/\partial c_2\right)_{T,c_1,c_2=0}\right)$,[55] where α is the volume fraction of the system available to the centers of mass of the polymer coils.

On a model-free basis, we define α as an *apparent* free-volume fraction. We therefore replace $\left(\partial \hat{f}^{ex}/\partial c_2\right)_{T,c_1,c_2=0}$ with $-\ln\alpha$ in eq 1. By differentiating eq 1 with respect to c_1 and c_2, we obtain the following expressions for the reduced chemical potentials and osmotic pressure (to first order in c_2)

$$\hat{\mu}_1 = \left(\frac{\partial \hat{f}}{\partial c_1}\right)_{T,c_2} = \hat{\mu}_1'(c_1,T) - \frac{c_2}{\alpha}\left(\frac{\partial\alpha}{\partial c_1}\right)_T \tag{2a}$$

$$\hat{\mu}_2 = \left(\frac{\partial \hat{f}}{\partial c_2}\right)_{T,c_1} = \ln\left(\frac{c_2}{\alpha}\right)_T \tag{2b}$$

$$\widehat{\Pi}(c_1,c_2,T) = c_1\hat{\mu}_1 + c_2\hat{\mu}_2 - \hat{f} = \widehat{\Pi}'(c_1,T) + c_2 - c_1\frac{c_2}{\alpha}\left(\frac{\partial\alpha}{\partial c_1}\right)_T \tag{2c}$$

where $\hat{\mu}_1'(c_1,T)$ and $\widehat{\Pi}'(c_1,T)$ are respectively, the protein reduced chemical potential and the reduced osmotic pressure of the protein-buffer binary system. Eq 2b shows that the thermodynamic activity of the polymer component is represented by the polymer concentration in the free volume: c_2/α.

Protein-polymer thermodynamic interactions can be described by the following cross-derivative:[55]

$$\left(\frac{\partial\hat{\mu}_1}{\partial c_2}\right)_{T,c_1} = \left(\frac{\partial\hat{\mu}_2}{\partial c_1}\right)_{T,c_2} = -\frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial c_1}\right)_T \tag{3}$$

For pure depletion interactions, $(\partial\alpha/\partial c_1)_T$ is negative because the free volume decreases as the protein concentration increases. This implies that $(\partial\hat{\mu}_1/\partial c_2)_{T,c2} > 0$, consistent with protein-polymer repulsive interactions.

Protein-protein thermodynamic interactions can be described by the following chemical-potential derivative

$$\left(\frac{\partial \hat{\mu}_1}{\partial c_1}\right)_{T,\hat{\mu}_2} = \left(\frac{\partial \hat{\mu}_1'}{\partial c_1}\right)_T - \frac{c_2}{\alpha}\left(\frac{\partial^2 \alpha}{\partial^2 c_1}\right)_T \tag{4}$$

Eq 4 provides also the basis for the thermodynamic definition of the protein second virial coefficient. For pure depletion interactions, $(\partial^2 \alpha/\partial^2 c_1)_T$ is positive because the overlap of depletion layers produces a free-volume excess with respect to that of isolated layers. This overlapping increases with protein concentration. This implies that, as the polymer concentration increases, $(\partial \hat{\mu}_1/\partial c_1)_{T,\hat{\mu}_2}$ decreases, consistent with a corresponding increase of protein-protein attractive interactions.

In summary, the first derivative $(\partial \alpha/\partial c_1)_T$ is related to protein-polymer interactions while the second derivative $(\partial^2 \alpha/\partial^2 c_1)_T$ is related to the effect of polymer concentration on protein-protein interactions. Since α is, in general, not necessarily a free-volume fraction, no *a priori* assumption on the sign of $(\partial \alpha/\partial c_1)_T$ and $(\partial^2 \alpha/\partial^2 c_1)_T$ can or should be made. Thus protein-polymer repulsive interactions can in principle occur together with a corresponding increase of protein-protein repulsive interaction. This has been recently shown to be the case for lysozyme-PEG mixtures.[56,89,199]

We note that the protein-polymer-buffer system has been often described as an effective protein one-component system.[75, 87] According to this description, the polymer effect can be expressed in the form of an effective pair potential in the coordinates of the protein particles (potential of the mean force). For proteins treated as hard spheres in the presence of non-adsorbing polymers, an attractive depletion potential is generated when the depletion layers of two protein particles overlap (Chapter 1). However, the pair approximation of this depletion potential is reasonably accurate only at low protein

concentrations[98,200] and, consequently, becomes an acceptable approximation for describing second-virial-coefficient data. At the concentrations relevant to LLPS, it has been shown that the pair-potential approximation can become a significant source of error because overlapping of depletion layers can involve three particles at the same time.[98,200] The two-component approach avoids this problem. Furthermore, the latter approach can be also used to describe the presence of protein-polymer partitioning in the two coexisting liquid phases.

We now consider the liquid-liquid phase transition of the protein-polymer-buffer system. This is described by the LLPS phase boundary $T_{\text{ph}}(c_1,c_2)$. We will show that protein-polymer partitioning between the two liquid phases can be related to protein-polymer interactions while the dependence of T_{ph} on polymer concentration can be related to the effect of polymer concentration on protein-protein interactions.

Protein/polymer partitioning in the two coexisting liquid phases I and II $(c_1^{\text{I}}, c_2^{\text{I}}, c_1^{\text{II}}, c_2^{\text{II}})$ can be examined by considering eq 2b and applying the condition of chemical equilibrium for the polymer component rewritten as:

$$\frac{c_2^{\text{I}}}{\alpha^{\text{I}}} = \frac{c_2^{\text{II}}}{\alpha^{\text{II}}} \tag{5}$$

To qualitatively show that the protein/polymer partitioning is related to $(\partial\alpha/\partial c_1)_T$, we perform a power series expansion of eq 5 about the protein critical concentration c_1^{c}. This yields:

$$\ln\frac{c_2^{\text{II}}}{c_2^{\text{I}}} = \frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial c_1}\right)_{T,c_1=c_1^{\text{c}}}(c_1^{\text{II}} - c_1^{\text{I}}) + \cdots \tag{6}$$

Eq 6 shows the relation of protein/polymer partitioning to protein-polymer interactions.

The effect of polymer concentration on T_{ph} can be examined by numerically solving the three equilibrium conditions between two liquid phases I and II: $\hat{\mu}_1(\text{I}) = \hat{\mu}_1(\text{II})$, $\hat{\mu}_2(\text{I}) = \hat{\mu}_2(\text{II})$ and $\widehat{\Pi}(\text{I}) = \widehat{\Pi}(\text{II})$. However an analytical expression relating T_{ph} to $(\partial^2\alpha/\partial^2 c_1)_T$ cannot be determined. To show qualitatively that the dependence of T_{ph} on c_2 is related to $(\partial^2\alpha/\partial^2 c_1)_T$, we will consider the analytically accessible spinodal condition, $(\partial\hat{\mu}_1/\partial c_1)_{T=T_{sp},\hat{\mu}_2} = 0$. This defines the boundary $T_{sp}(c_1,c_2)$ between the stable domain and the unstable domain of the system. The effects of polymer concentration on T_{ph} and T_{sp} are closely related because the LLPS boundary is tangent to the spinodal boundary.[91]

For pure excluded-volume interactions, eq 4 predicts that, as the polymer concentration increases, $(\partial\hat{\mu}_1/\partial c_1)_{T,\hat{\mu}_2}$ decreases and can be made to reach zero. Thus, sufficient polymer can be used to bring the protein-polymer-buffer system to the spinodal boundary. This is a sufficient condition for LLPS to occur.

Protein-protein net attraction energy drives LLPS. We however observe that LLPS of the protein-buffer binary system rarely has been reported in the experimentally accessible temperature domain (~260-320 K). One reasonable hypothesis is that this attraction is usually weak for water-soluble proteins. This implies that protein-buffer systems (with attractive energy) would usually behave as "supercritical fluids" in the accessible temperature domain. This observation allows us to conclude that the following high-temperature series expansions can be written for $\hat{f}'(c_1,T)$.[55, 201-203]

$$\hat{f}'(c_1,T) = \hat{f}^{(0)}(c_1) + \hat{f}^{(1)}(c_1)\frac{1}{RT} + \hat{f}^{(2)}(c_1)\left(\frac{1}{RT}\right)^2 + \cdots \qquad (7)$$

In eq 7, $\hat{f}^{(1)}(c_1)$ represents the internal energy (per unit volume) of the protein-buffer binary system to the first order in $1/RT$. This quantity is negative in the presence of protein-protein weak attraction energy. We comment that eq 7 may fail to describe the LLPS boundary of the binary protein-buffer system, which would be located at relatively low temperatures, where higher-order terms may not be negligible.

If we insert eq 7 into eq 4 and apply the spinodal condition, then we obtain to the first order:

$$\frac{1}{T_{sp}} = -R \frac{\left(\mathrm{d}^2 \hat{f}^{(0)}/\mathrm{d}c_1^2\right)}{\left(\mathrm{d}^2 \hat{f}^{(1)}/\mathrm{d}c_1^2\right)}\left[1 - \frac{(\mathrm{d}^2 \alpha/\mathrm{d}c_1^2)/\alpha}{\mathrm{d}^2 \hat{f}^{(0)}/\mathrm{d}c_1^2} c_2\right] + \cdots \tag{8}$$

Eq 8 shows how the dependence of T_{sp} on polymer concentration is related to the effect of polymer concentration on protein-protein interactions. For our protein-buffer system, $\left(\mathrm{d}^2 \hat{f}^{(0)}/\mathrm{d}c_1^2\right) > 0$ due to protein-protein hard-core repulsion,[85] and $\left(\mathrm{d}^2 \hat{f}^{(1)}/\mathrm{d}c_1^2\right) < 0$ due to the protein-protein attraction energy.[55] Thus, T_{sp} is expected to increase with polymer concentration for pure depletion interactions.

Although this thermodynamic perturbation theory makes no assumption on the nature of microscopic interactions, there are two important limitations to take into account. These are: (1) the buffer is regarded as one pseudo-component; (2) polymer-polymer interactions are not taken into account. We will discuss these two limitations in section 5 and 7 respectively.

4.   The Effect of PEG1450 Concentration on the LLPS Temperature

Upon the addition of PEG1450 (20-80 mg/mL), LLPS is observed within the experimental temperature domain. In Fig 2, we report our measurements of LLPS temperature, T_{ph}, as a function of PEG1450 concentration, c_2, at five constant BSA concentrations, c_1 around the protein critical concentration (here in mg/mL), $c_1^c \approx 240$ mg/mL. This value of c_1^c was obtained from BSA/PEG1450 partitioning measurements shown later (Section 5). In all five cases shown in Fig 2, the LLPS temperature increases with PEG1450 concentration. According to eq 8, $(d^2\alpha/dc_1^2)/(d^2\hat{f}^{(1)}/dc_1^2) < 0$ from our experimental results. This result is consistent with (1) the protein-protein interaction energy being attractive in the protein-buffer binary system and (2) the PEG1450 concentration increasing protein-protein attractive interactions. The first effect is in agreement with our B value while the second effect is expected in the presence of depletion interactions. We also observe that T_{ph} increases with c_1 for all of our PEG1450 concentrations (see Fig 2). This implies that $(\partial T_{ph}/\partial c_1)_{c2} > 0$. Since $(\partial T_{ph}/\partial c_1)_{c2} = -(\partial T_{ph}/\partial c_2)_{c1}(\partial c_2/\partial c_1)_{T_{ph}}$ (a mathematic condition for a ternary system), we conclude that $(\partial c_2/\partial c_1)_{T_{ph}} < 0$ around the critical protein concentration. This is consistent with the presence of significant BSA/PEG1450 partitioning. Similar results were also obtained for γ-crystallin–PEG aqueous mixtures.[4, 55]
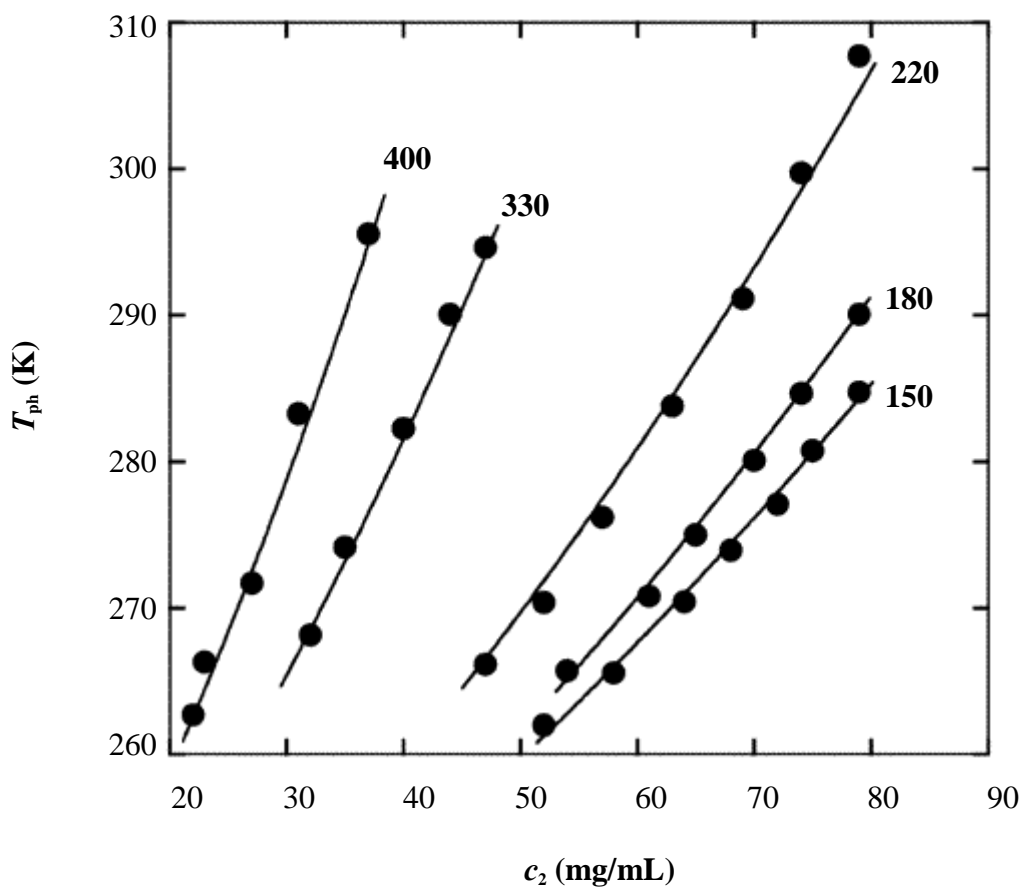
Figure 2. LLPS temperature, T_{ph}, as a function of PEG1450 concentration, c_2, at five constant BSA concentrations, c_1. The solid curves represent our calculated values obtained by applying equilibrium conditions to equations 2a-c as described in the Discussion section. The numbers associated with each curve identify the corresponding value of c_1 in mg/mL.

5. Isothermal Partitioning in the BSA-PEG1450-Buffer Ternary System

In Fig 3, we show measurements of BSA/PEG1450 partitioning (c_1^{I}, c_2^{I}, c_1^{II}, c_2^{II}) performed at three different temperatures. The straight lines (tie lines), which connect the pairs of points representing the coexisting phases, show that there is a large difference in polymer concentration between the two phases. We also find that this difference is still significant if we report the PEG concentrations with respect to the buffer volume (Table 1), calculated by removing the volumetric contribution of the protein component, c_2' $=c_2/(1-\phi)$, where $\phi = V_1^{sp} c_1$ is the volume fraction of protein, V_1^{sp} is the specific volume of protein. This is consistent with the presence of BSA-PEG net repulsive interactions as expected for depletion interactions. We have used our measurements of BSA/PEG1450 partitioning to estimate the protein critical concentration, c_1^{c}, by simply averaging the two concentrations c_1^{I} and c_1^{II} (Table 2). We estimate the critical concentration to be 240 ± 20 mg/mL, corresponding to a volume fraction of $\phi_{\mathrm{c}} = 0.18 \pm 0.01$.

In Fig 3, we also show values of (c_1, c_2) extracted by interpolation of our $T_{\mathrm{ph}}(c_1, c_2)$ results of Fig 3. The overall agreement between the two sets of data can be considered satisfactory. However, a small discrepancy is observed in our results at high protein concentrations, where the (c_1^{II}, c_2^{II}) points appear to be located at lower c_1 and higher c_2 values than those predicted from the values of $T_{\mathrm{ph}}(c_1, c_2)$. Due to the high viscosity of the protein concentrated phase, the mechanical separation of residual protein-dilute phase from the protein-concentrate phase may be difficult to achieve. We thus expect that the protein concentrated phases may still contain small amounts (~10%) of the protein-dilute

phase. As we will discuss later, this will not be a significant source of errors in our quantitative examination of the BSA-PEG1450 interactions.

We now examine our assumption that the buffer can be regarded as one pseudo-component. This assumption is valid if the internal composition of the buffer is the same in all of our experiments. For our $T_{ph}(c_1,c_2)$ measurements, the concentration of buffer solutes inside the buffer volume is indeed kept the same. However, in the two coexisting phases (c_1^I, c_2^I) and (c_1^{II}, c_2^{II}), the thermodynamic activity of buffer solutes is the same. Thus the buffer can be approximately treated as one pseudo-component if the difference of the buffer solute concentrations in the two coexisting phases is small. This is corroborated by the observed agreement between the two sets of data in Fig 3.
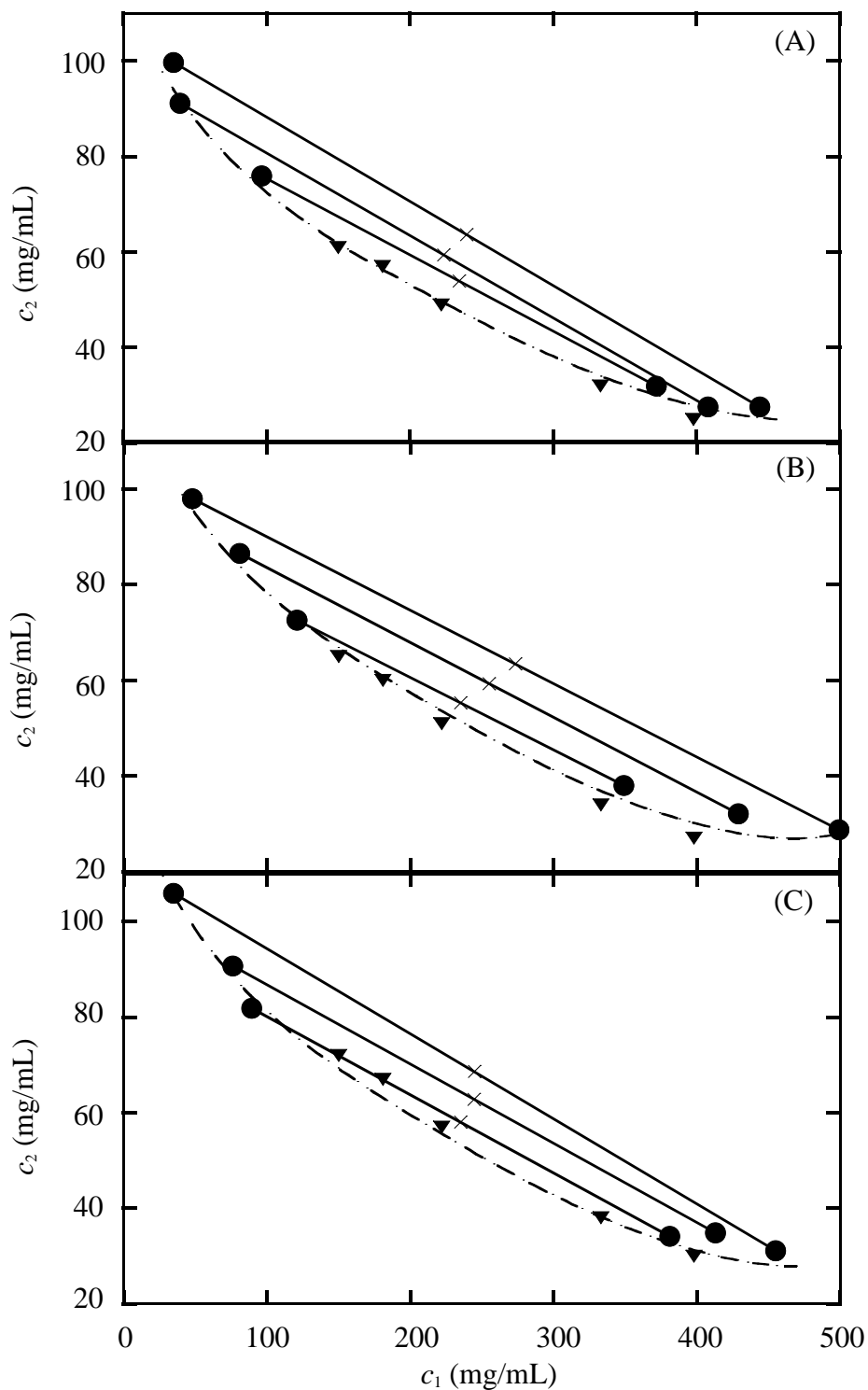
Figure 3. Measurements of BSA/PEG1450 partitioning (c_1^{I}, c_2^{I}, c_1^{II}, c_2^{II}) at three different temperatures: (A) 278, (B) 271, and (C) 268 K. The straight lines (tie lines) connect the pairs of points representing the coexisting phases (·). The crosses (×) represent the average composition of the coexisting phases. The triangles (▼) represent the values of (c_1,c_2) extracted by interpolation of our $T_{\mathrm{ph}}(c_1,c_2)$ values of Table 1. The dashed curves are guides for the eye.

			$T=268$ K		
c_1^{I} (mM)	c_1^{II} (mM)	c_2^{I} (mM)	c_2^{II} (mM)	$c'_2{}^{\mathrm{I}}$ (mM)	$c'_2{}^{\mathrm{II}}$(mM)
1.46	5.57	52.41	22.07	56.55	30.34
0.60	6.17	62.76	19.31	64.83	27.59
0.53	6.63	68.97	19.31	71.03	28.28

			$T=271$ K		
c_1^{I} (mM)	c_1^{II} (mM)	c_2^{I} (mM)	c_2^{II} (mM)	$c'_2{}^{\mathrm{I}}$ (mM)	$c'_2{}^{\mathrm{II}}$(mM)
1.81	5.27	50.34	26.21	55.17	35.17
1.20	6.48	60.00	22.07	63.45	32.41
0.72	7.53	67.59	20.00	70.34	31.72

			$T=278$ K		
c_1^{I} (mM)	c_1^{II} (mM)	c_2^{I} (mM)	c_2^{II} (mM)	$c'_2{}^{\mathrm{I}}$ (mM)	$c'_2{}^{\mathrm{II}}$(mM)
1.36	5.72	56.55	23.45	60.69	32.41
1.14	6.17	62.76	24.14	66.21	34.48
0.53	6.93	75.86	21.38	77.93	32.41

Table 1. Measurements of BSA/PEG1450 partitioning (c_1^{I}, c_2^{I}, c_1^{II}, c_2^{II}) at three different temperatures. $c'_2{}^{\mathrm{I}}$, $c'_2{}^{\mathrm{II}}$ are the effective PEG concentrations in the two coexisting phases with respect to the buffer volume, calculated by $c'_2 = c_2/(1-\phi)$.

			$T=268$ K		
c_1^{I} (mM)	c_1^{II} (mM)	c_2^{I} (mM)	c_2^{II} (mM)	$\bar{c}_1$ (mM)	$\bar{c}_2$ (mM)
1.46	5.57	52.41	22.07	3.46	37.24
0.60	6.17	62.76	19.31	3.31	40.69
0.53	6.63	68.97	19.31	3.61	44.14

			$T=271$ K		
c_1^{I} (mM)	c_1^{II} (mM)	c_2^{I} (mM)	c_2^{II} (mM)	$\bar{c}_1$ (mM)	$\bar{c}_2$ (mM)
1.81	5.27	50.34	26.21	3.46	37.93
1.20	6.48	60.00	22.07	3.31	40.69
0.72	7.53	67.59	20.00	3.61	43.45

			$T=278$ K		
c_1^{I} (mM)	c_1^{II} (mM)	c_2^{I} (mM)	c_2^{II} (mM)	$\bar{c}_1$ (mM)	$\bar{c}_2$ (mM)
1.36	5.72	56.55	23.45	3.46	40.00
1.14	6.17	62.76	24.14	3.31	43.45
0.53	6.93	75.86	21.38	3.61	47.59

Table 2. Average BSA and PEG1450 concentrations ($\bar{c}_1$, $\bar{c}_2$) at three different temperatures. The critical concentration of BSA can be estimated by the average BSA concentration, $c_1^{\mathrm{c}} \approx \bar{c}_1$.

6. Determination of Depletion Interaction from Partitioning Measurements

The thermodynamic behavior of the BSA-PEG1450-buffer system can be examined by applying the above-described thermodynamic perturbation theory to our experimental results. Our approach will be to use an expression for $\alpha(c_1)$ able to describe our two sets of data. Because both sets of data are consistent with the presence of depletion interactions, we will use an expression of $\alpha(c_1)$ that represents the free-volume fraction. The dependence of the free-volume fraction on the thickness of the depletion layer is taken into account by introducing the ratio $q = \delta/R_1$, where R_1 is the radius of the protein assumed to be a sphere and δ is the thickness of the depletion layer.[55,85,86,89] However, the parameter q is not calculated from microscopic parameters but is determined using our experimental results. For α, we will use an expression that treats proteins as hard spheres. From scaled particle theory[85,189-192] applied to hard spheres (Appendix A), the following expression has been determined:

$$\alpha = (1 - \phi)\exp(-A\eta - B\eta^2 - C\eta^3) \tag{9}$$

where $\phi = c_1 V_{\text{prot}}$, $\eta = \phi/(1 - \phi)$, $A = 3q + 3q^2 + q^3$, $B = 4.5q^2 + 3q^3$, $C = 3q^3$. Our computer simulations show that eq 9 is reasonably accurate even for very dense hard-sphere fluids (Chapter 9, Section 2). We comment that, when ϕ is small, the overlapping of the depletion layers can be neglected and eq 9 becomes $\alpha \approx 1 - (1 + q)^3 \phi$.

A set of apparent q values is determined from the partitioning measurements by applying eq 5, i.e. $c_2^I/\alpha(c_1^I, T) = c_1^{II}/\alpha(c_1^{II}, T)$, and eq 9 to the $(c_1^{\text{I}}, c_2^{\text{I}}, c_1^{\text{II}}, c_2^{\text{II}})$ values (Table 3). No dependence of q on temperature and the average PEG concentration, $\bar{c}_2$,

could be established within the experimental error. We thus report the average value: $q = 0.35 \pm 0.03$.

$\bar{c}_2$ (mg/mL)	q	q'	$\bar{c}_2$ (mg/mL)	q	q'	$\bar{c}_2$ (mg/mL)	q	q'
	$T = 268$ K			$T = 271$ K			$T = 278$ K	
54	0.39	0.34	55	0.35	0.31	58	0.37	0.32
59	0.40	0.33	59	0.33	0.29	63	0.34	0.30
64	0.37	0.32	63	0.29	0.25	69	0.34	0.30

Table 3. Normalized thickness of depletion lay, q, for the BSA-PEG1450 pair calculated from the experimental results of the isothermal partitioning measurements of BSA/PEG1450 using scaled particle theory. q''s are calculated using the approximation expression of free-volume, $\alpha \approx 1 - (1 + q)^3 \phi$.

We note that the q' values in Table 3 calculated by using $\alpha \approx 1 - (1 + q)^3 \phi$ are only 10-15% smaller than the values calculated using eq 9. We thus conclude that the overlapping of the depletion layers contributes only marginally to our q values.

As mentioned in Section 5, the protein-concentrate phases may still contain small amounts of the corresponding protein-dilute phases. However, we observe that, according to $\alpha \approx 1 - (1 + q)^3 \phi$, even a large contamination has no effect on the obtained value of q. According to the complete eq 9, we estimate that 10% of the dilute phase would produce a small increase in the q values (2-4%). We therefore conclude that contamination does not significantly change our final results.

7. Determination of Depletion Interaction from LLPS Temperature Measurements

The apparent q values are determined from the measurements of LLPS temperature by comparing our results in Fig. 2 with the LLPS phase boundary computed by applying the three equilibrium conditions of phases I and II: $\hat{\mu}_1(\text{I}) = \hat{\mu}_1(\text{II})$, $\hat{\mu}_2(\text{I}) = \hat{\mu}_2(\text{II})$ and $\widehat{\Pi}(\text{I}) = \widehat{\Pi}(\text{II})$ to eqs 2a-c:

$$\hat{\mu}_1'(c_1^{\text{I}}, T) - \frac{c_2^{\text{I}}}{\alpha^{\text{I}}}\left(\frac{\partial \alpha^{\text{I}}}{\partial c_1^{\text{I}}}\right)_T = \hat{\mu}_1'(c_1^{\text{II}}, T) - \frac{c_2^{\text{II}}}{\alpha^{\text{II}}}\left(\frac{\partial \alpha^{\text{II}}}{\partial c_1^{\text{II}}}\right)_T \tag{10a}$$

$$\ln\left(\frac{c_2^{\text{I}}}{\alpha^{\text{I}}}\right)_T = \ln\left(\frac{c_2^{\text{II}}}{\alpha^{\text{II}}}\right)_T \tag{10b}$$

$$\widehat{\Pi}'(c_1^{\text{I}}, T) + c_2^{\text{I}} - c_1^{\text{I}}\frac{c_2^{\text{I}}}{\alpha^{\text{I}}}\left(\frac{\partial \alpha^{\text{I}}}{\partial c_1^{\text{I}}}\right)_T = \widehat{\Pi}'(c_1^{\text{II}}, T) + c_2^{\text{II}} - c_1^{\text{II}}\frac{c_2^{\text{II}}}{\alpha^{\text{II}}}\left(\frac{\partial \alpha^{\text{II}}}{\partial c_1^{\text{II}}}\right)_T \tag{10c}$$

where the expressions of α and $(\partial \alpha / \partial c_1)$ are calculated by eq 9.

However this requires not only an expression for $\alpha(c_1)$ but also one for $\hat{f}'(c_1, T)$. For $\hat{f}'(c_1, T)$, we consider eq 7a applied to hard spheres interacting by square-well potentials with well magnitude ε and range of interaction λ:

$$u(r) = \begin{cases} \infty, & for\ 0 \leq r < \sigma \\ -\varepsilon, & for\ \sigma \leq r \leq \lambda\sigma \\ 0, & for\ r > \lambda\sigma \end{cases} \tag{11}$$

The expression for $\hat{f}^{(0)}(\phi)$ is derived from the Carnahan-Starling equation of state,[204] which has shown to be very accurate even for very dense hard-sphere fluids,

$$Z = \frac{1 + \phi + \phi^2 - \phi^3}{(1 - \phi)^3} \tag{12}$$

From the basic thermodynamic relations, we can derive:

$$\hat{f} = c \left(\frac{\partial \hat{f}}{\partial c} \right)_{V,T} - \widehat{\Pi}$$

$$Z = c \left(\frac{\partial (\hat{f}/c)}{\partial c} \right)_{V,T}$$

$$\hat{f}(c,T) = c \ln \left(\frac{c}{e} \right) + c \int_0^\phi \frac{Z-1}{\phi} \, d\phi \tag{13}$$

From the Carnahan-Starling equation of state, we arrived to:

$$\hat{f}^{(0)}(\phi) = c_1 \ln \left(\frac{c_1}{e} \right) + c_1 \frac{\phi(4 - 3\phi)}{(1 - \phi)^2} \tag{14}$$

The expressions of $\hat{f}^{(1)}(\phi)$ and $\hat{f}^{(2)}(\phi)$ were derived by using thermodynamic second order perturbation theory (Appendix C). The $\hat{\mu}_1'(c_1, T)$ and $\widehat{\Pi}'(c_1, T) = c_1 \hat{\mu}_1'(c_1, T) - \hat{f}'(c_1, T)$ were calculated from the expression of $\hat{f}'(c_1, T)$.

The expressions for $\hat{f}^{(1)}(\phi)/\varepsilon$ and $\hat{f}^{(2)}(\phi)/\varepsilon^2$ depend on the choice of reduced range of square-well potential, λ. According to the model of the square-well potential, ϕ_c = 0.18 ± 0.01 (see section 5) corresponds to $\lambda \approx 1.5$.[60] We thus consider the corresponding expressions at $\lambda = 1.5$.[205] We remark that, according to eq 8 and a numerical analysis of the LLPS boundary, the obtained value of q is significantly affected only by the choice of $\hat{f}^{(0)}(\phi)$ and not by that of $\hat{f}^{(1)}(\phi)/\varepsilon$ and $\hat{f}^{(2)}(\phi)/\varepsilon^2$. This quantity, contrary to $\hat{f}^{(1)}(\phi)$ and $\hat{f}^{(2)}(\phi)$, does not depend on the nature of the attractive part of the potential. We also observe that the choice of $\hat{f}^{(0)}(\phi)$ and $\alpha(\phi)$ must be consistent with respect to each other since the accuracy of their second-derivative ratio is crucial for the determination of q (see eq 8).

To numerically calculate T_{ph} as a function of c_2, we start by setting the values of $\hat{\mu}_2$ and T. Then the equilibrium conditions $\hat{\mu}_1(c_1^I, \hat{\mu}_2, T) = \hat{\mu}_1(c_1^{II}, \hat{\mu}_2, T)$ and $\hat{\Pi}(c_1^I, \hat{\mu}_2, T) = \hat{\Pi}(c_1^{II}, \hat{\mu}_2, T)$ can be rewritten as:

$$\hat{\mu}_1'(c_1^I, T) - \left(\frac{\partial \alpha^I}{\partial c_1^I}\right)_T \exp(\hat{\mu}_2) = \hat{\mu}_1'(c_1^{II}, T) - \left(\frac{\partial \alpha^{II}}{\partial c_1^{II}}\right)_T \exp(\hat{\mu}_2) \tag{15a}$$

$$\hat{\Pi}'(c_1^I, T) + \left[\alpha^I - c_1^I \frac{c_2^I}{\alpha^I}\left(\frac{\partial \alpha^I}{\partial c_1^I}\right)_T\right]\exp(\hat{\mu}_2) = \hat{\Pi}'(c_1^{II}, T) + \left[\alpha^{II} - c_1^{II}\left(\frac{\partial \alpha^{II}}{\partial c_1^{II}}\right)_T\right]\exp(\hat{\mu}_2)$$

$$\tag{15b}$$

For given values of T_{ph} and $\hat{\mu}_2$, we obtain c_1^I and c_1^{II} by numerically solving eq 15a,b (see Appendix D). The corresponding values of c_2^I and c_2^{II} are obtained by applying eq 2b. The phase boundary is then computed by repeating this procedure for several values of T_{ph} and $\hat{\mu}_2$. Finally a $T_{ph}(c_2)$ curve is generated for the chosen value of c_1 by interpolation. The comparison between the experimental data and the calculated curves (solid lines in Fig 2) has allowed us to determine q and ε for each curve.

Our results of q and ε for each experimental c_1 value are reported in Table 4. We observed that changes in the reported q values larger than 0.01 produce a noticeable discrepancy between the experimental data and the theoretical curves for all ε values. We therefore obtain the same value of q (within the experimental error) in all five cases and report $q = 0.31 \pm 0.01$. That q does not depend on c_1 is strong evidence that the chosen expressions for $\hat{f}^{(0)}(\phi)$ and $\alpha(\phi)$ are satisfactory. However, the value of ε is found to increase with c_1 at the highest protein concentrations. This observed discrepancy can be attributed to the nonsufficient accuracy in the chosen expressions for $\hat{f}^{(1)}(\phi)$ and $\hat{f}^{(2)}(\phi)$.

c_1 (mg/mL)	ε / k (K)	q
150	172	0.31 ± 0.01
180	168	0.31 ± 0.01
220	168	0.31 ± 0.01
330	181	0.31 ± 0.01
400	198	0.31 ± 0.01

Table 4. Normalized thickness of depletion lay, q, for the BSA-PEG1450 pair numerically extracted from the experimental results of the LLPS temperature measurements of BSA/PEG1450 using thermodynamic perturbation theory and scaled particle theory. k is Boltzmann constant.

We also make an attempt to estimate the value of ε from our experimental value of B. For the square-well potential, $BM_1^2/V_1 = 4 - 4[\exp(\varepsilon/RT) - 1](\lambda^3 - 1)$.[206] This equation yields $\varepsilon/k = 120$ K for $\lambda = 1.5$. Considering all approximations involved, this value is not very different from those reported in Table 4.

As already mentioned in Section 3, the thermodynamic perturbation theory described above neglects the presence of polymer-polymer interactions. These interactions should be included for a more general and accurate description of our system. We also note that this approximation may become rather inadequate if coil-coil interpenetration is significant. For polymers, the concentration at which this interpenetration starts to be important marks the passage from their dilute regime to their semidilute regime. For PEG1450, this transition occurs at $c_2 \approx 100$ mg/mL.[207] Thus our experimental PEG concentrations for $T_{ph}(c_2)$ fall inside the dilute-regime domain. We

however observe that polymer-polymer interactions in our ternary system cannot be simply approximated with those of the corresponding PEG1450-buffer system. This would ignore the effect of BSA concentration on polymer-polymer interactions, which is expected to be significant due to the high protein concentration of our mixtures. As criteria to evaluate the importance of polymer-polymer interactions, we will directly consider our ternary mixtures. We observe that our measurements on the LLPS temperature, which are performed on a *wide* range of PEG1450 concentrations, can be accurately described by the same value of $q = 0.31$. This indicates that our results do not significantly depend on PEG1450 concentration. This would not be consistent with significant PEG-PEG interactions. We thus believe that neglecting the presence of polymer-polymer interactions does not significantly affect our conclusions.

8. Intrinsic Consistency of the Scaled Particle Model and Discrepancy between Experimental q and Microscopic Estimation

Our analysis shows that the agreement on q between the two sets of data is good. This implies that the depletion-interaction model satisfactorily describes the BSA-PEG1450-buffer system. Thus eq 9 with $q = 0.31$ can be used to describe both BSA-PEG1450 interactions (from eq 3) and the effect of PEG1450 concentration on protein-protein interactions (from eq 4). We also note that this q value for the BSA/PEG1450 pair is comparable to that of $q \approx 0.3$ obtained from partitioning measurements on the γ-

crystallin/PEG1450.[4,55] The agreement between BSA and γ-crystallin suggests that depletion interactions are similar for a wide range of protein/PEG pairs.

We now make an attempt to compare the determined value of q = 0.31 with that obtained from microscopic structural parameters. The value of q is given by,

$$q \equiv \frac{\delta}{R_1} = f(R_g/R)\frac{R_g}{R}$$

where R_g is the radius of gyration of the polymer and $f(R_g/R)$ is a parameter function itself of R_g/R due to polymer deformability. We estimate R = 2.7 nm for BSA from $R = \sqrt[3]{3V_1^{sp}/4\pi N_A M_1}$, where the specific volume of BSA, $V_1 = 0.735(\text{mL/g})$, molecular weight of BSA $M_1 = 66400(\text{g/mol})$ and N_A is the Avogadro's constant. The value of R_g = 1.5 nm for PEG1450 was calculated from the known experimental dependence of R_g on PEG molecular weight, $R_g = 0.27M^{0.55}$. For our R_g/R value, computer simulations and theoretical modeling gives $f(R_g/R) = 1.010$, thereby yielding q = 0.556.

The discrepancy between calculated and experimental values of q can be related to several microscopic details such as the actual shape of the protein molecules, the conformational properties of PEG coils, and the presence of protein-PEG weak attraction. A similar discrepancy was also observed for γ-crystallin/PEG mixtures. In the next Chapter, we will examine the conformational properties of PEG coils as a possible source of the observed discrepancy.

9. LLPS of Protein-PEG-Buffer Mixtures

The experimental results reported in the previous section show that LLPS may be induced if the temperature of a protein-PEG-buffer mixture is lowered below the corresponding phase boundary. The presence of this phase transition can be predicted by a thermodynamic perturbation theory that treats the buffer as one pseudo-component and neglects polymer-polymer interactions. These two factors, which may not be ignored for high concentrations of buffer solutes or high PEG molecular weights, can bring about a more complex phase-transition behavior. As an example, we consider the mixture prepared by using BSA (200 mg/mL), PEG8000 (70 mg/mL), and sodium phosphate buffer (0.2 M, pH 7.1). This mixture undergoes LLPS not only by lowering the temperature below 273 K but also by increasing it above 300 K. For this mixture, we determine the two coexistence curves shown in Fig 4. We have also found that a small increase in phosphate concentration significantly reduces the temperature gap between the two boundaries.

The coexistence curve with the upper critical point can be qualitatively described by invoking the same factors used to explain the BSA-PEG1450-acetate buffer system. However, the coexistence curve with a lower critical point can be explained by considering the corresponding PEG8000-phosphate buffer mixture. For this system, LLPS with a lower critical point is also observed when the temperature is increased above 363 K, consistent with previous results. Thus the effect of BSA is to move this phase boundary toward lower temperatures. Since this phase transition is driven by the

PEG-buffer system, a wide range of protein systems can undergo LLPS when the temperature is increased.

We remark that two LLPS boundaries may be a valuable tool for protein crystallization and biomaterials science in general. For instance, since it is believed that critical fluctuations enhance the nucleation of protein crystals,[88] the presence of two critical points can extend the composition domain where critical fluctuations would occur. This could be used for optimizing protein crystallization.
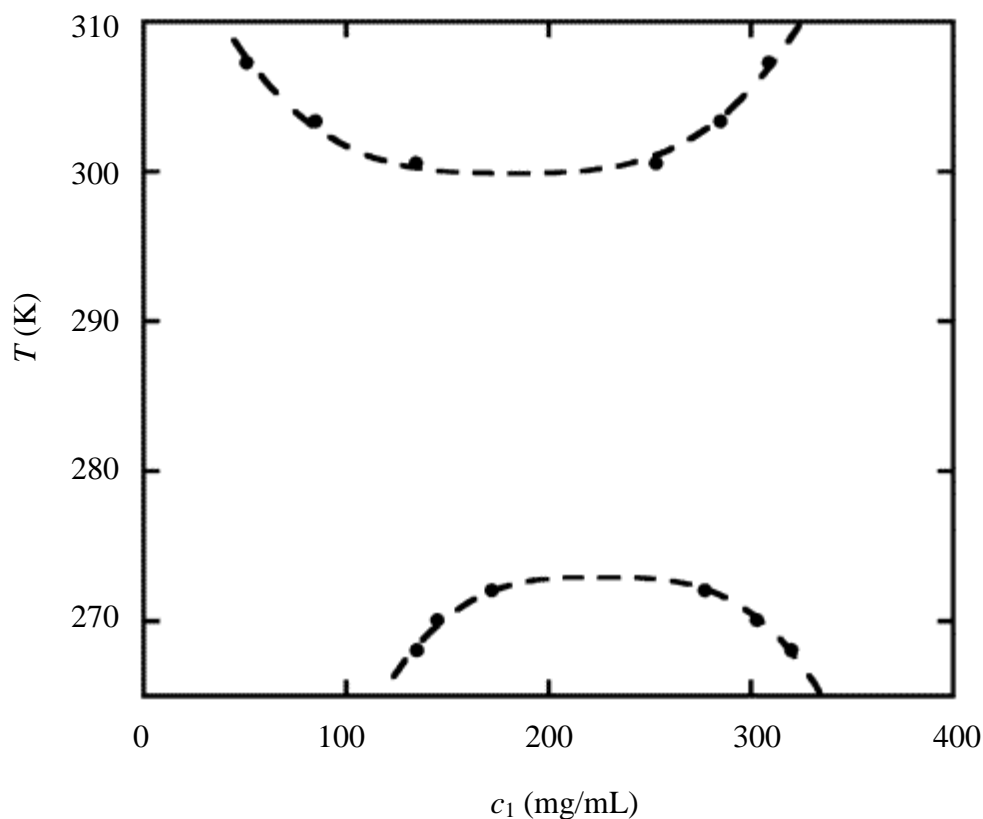
Figure 4. Two coexistence curves of the ternary mixture with the composition: BSA, 200 mg/mL; PEG8000, 70 mg/mL; sodium phosphate buffer, pH 7.1, 0.2 M. The points were determined by measuring the protein concentration in the two coexisting phases at several temperatures. The solid curves are guides for the eye.

Chapter 9

Effect of Polymer Conformation on

Depletion Interactions by

Monte Carlo Simulations

1. Introduction

In the last Chapter, we have shown that the insertion of non-absorbing polymer effectively induces protein condensation. This phenomenon is explained using a model describing depletion interactions. This model is based on an expression of free-volume fraction obtained from scaled particle theory. Free energy is expressed in terms of protein volume fraction and reduced thickness of depletion layer. The validity of this expression for the BSA-PEG-buffer system has been examined through checking experimentally the consistency of two thermodynamically independent quantities, which are both related to the free-volume fraction. However, the q value for the BSA/PEG1450 pair calculated from the protein radius and polymer gyration radius was only about 50% of the experimentally determined value. This discrepancy was also observed for other protein-PEG systems.

In order to explain this discrepancy and make better estimation of the q value, we need to consider other non-ideal effects of the protein-polymer systems, such as 1) specific structural/conformational properties of PEG coils; 2) the role of temperature on free volume due to protein-protein attractive interactions; 3) protein-PEG specific net interactions; 4) deviation of protein shape from that of a sphere. In this dissertation, we will examine the first two effects. Specifically, we have used Monte Carlo simulations to investigate the effect of PEG conformational degrees of freedom on the q value in this chapter. We have also examined the role of temperature on the free volume fraction of fluid of hard spheres with attractive interaction by treating the polymer as a hard sphere in the following chapter.

2. AO Sphere

We have first tested the accuracy of our algorithm. We have performed simulations on the free volume fraction, α, by inserting an AO-sphere (see Chapter 1, section 6) into a hard-sphere suspension. The ratio of the AO-sphere radius to the hard-sphere radius is the reduced thickness of depletion layer, q. Simulations were conducted at several values of q and hard-sphere volume fraction, ϕ. The number of hard spheres, N, is fixed. The value of ϕ is calculated from the radius of sphere, R, by $\phi = 4\pi R^3 N/3$. The results are reported in Appendix I table 1. For this hard-sphere system, four theoretical expressions of α are currently available:[196] 1) the scaled particle theory (SPT) expression,[191] 2) the Carnahan Starling (CS) expression,[198] 3) the Mansoori Carnahan Starling Leland (MCSL) expression,[208] and 4) the Andrews Ellerby (AE) expression.[209,210] We define $\eta \equiv \phi/(1-\phi)$, and the free volume fraction expressions are:

$$\alpha_{\text{SPT}} = (1-\phi)\exp\left(-(3q+3q^2+q^3)\eta - \left(\frac{9}{2}q^2+3q^3\right)\eta^2 - 3q^3\eta^3\right)$$

$$\alpha_{\text{CS}} = (1-\phi)\exp\left(-(3q+3q^2+q^3)\eta - \left(\frac{9}{2}q^2+3q^3\right)\eta^2 - (3q^3)\eta^3 + q^3\phi\eta^3\right)$$

$$\alpha_{\text{MCSL}} = (1-\phi)\exp\left((3q^2-2q^3)\ln(1+\eta) - (3q+6q^2-q^3)\eta - (3q^2+4q^3)\eta^2 - 2q^3\eta^3\right)$$

$$\alpha_{\text{AE}} = (1-\phi)\exp\left(-\frac{(3q+3q^2+q^3)(\eta+\eta^2)}{\left(1+\left(1-\frac{15}{14}q\right)\eta + \left(\frac{87}{7\sqrt{2}\pi}-\frac{18}{\pi^2}-\frac{15}{14}\right)q\eta^2\right)}\right) \tag{1}$$

The comparison between our simulations and the theoretical curves is shown in Fig 1. We report $\alpha - (1+q)^3\phi$ as a function of ϕ. The quantity $\alpha - (1+q)^3\phi$ is the volume

fraction associated with the overlap of the depletion layers (see Chapter 8, section 6), which is directly related to hard-sphere distribution within the suspension.

We observe that our simulation results are in very good agreement with all theoretical curves with the exception of that associated with AE theory.
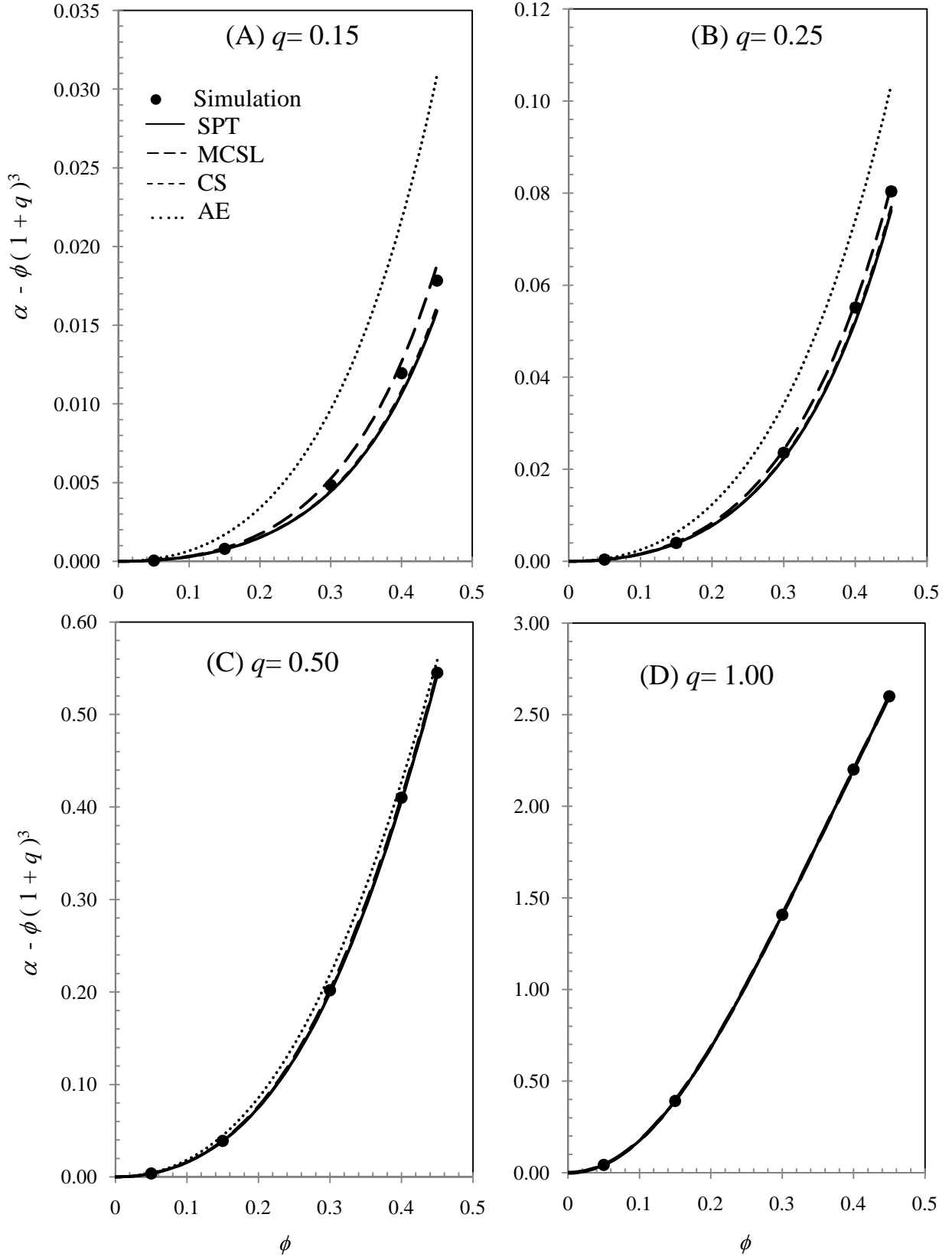
Figure 1. Our simulation and theoretical curves for $\alpha - (1 + q)^3 \phi$ as a function of ϕ at several q.

3. Deformability and Penetrability of Polymer Chain

On a molecular level, the reduced thickness of the depletion layer for a polymer coil can be calculated using, $q = \delta/R = fR_g/R$,[95, 96] where f is a function of polymer property. Since the depletion layer is defined as the region around the protein molecule which is not accessible to the center of polymer coil, it is clear that its reduced thickness, q, is proportional to the gyration radius normalized by the effective radius of protein molecule. If the polymer is a sphere then $R_g = \sqrt{3/5}R_{HS}$ and $q = R_{HS}/R$. This implies that $f = \sqrt{5/3}$. However, for a polymer coil, the correction factor f will be smaller than $\sqrt{5/3}$ because a coil is deformable and penetrable in the presence of globular compact particles (e.g. protein).[95, 122] Moreover, $f = f(R_g/R)$ itself will be a function of the ratio between polymer gyration radius to the protein radius.[95, 96, 98] If the protein molecule is relatively small compared to the polymer coil, it can easily penetrate the polymer coil, i.e. it can get closer to the center of polymer, and the correction factor would be significantly smaller than one. In contrast, when the protein molecule is relatively large with respect to the polymer coil, it will be hard for it to deform/penetrate the polymer coil and thus the correction factor will be large. We further note that penetrability also depends on the conformation of the polymer.

In the case of the BSA/PEG1450 pair, where $R_g/R = 1.5/2.7 = 0.56$, the correction factor for the random-coil model, $f(R_g/R) = 1.01$, was used to estimate q (Chapter 8).

We note that the ideal-coil model is extensively used to describe the depletion interactions for coilloid/polymer systems.[93-96] However, the ideal-coil model ignores

structural details of polymers such as the chemical bond lengths and bond angles and conformational details related to steric effects between the monomeric units and gauche-trans properties. In the case of colloidal particles (with radius of ≈ 100 nm), neglecting structural properties such as bond length and angles may be acceptable because the bond length is very small compared to the radius of the colloidal particle.[93-96] However, many proteins, including those investigated here, have a radius of only a few nanometers. Thus, this approximation may not be suitable for small proteins. We therefore use our simulations to examine the effect of structural parameters on the depletion interaction.

The first structural effect we introduce with respect to the random coil is the existence of backbone bonds. The polymer was modeled by a linear chain consisting of M segments (representing the bonds) with given length, l. These attached segments can freely rotate in any direction. The protein molecules were modeled as hard spheres in the absence of attractive interaction. We performed simulations to determine the free-volume fraction for the polymer with M segments as a function of the volume fractions, ϕ, of hard-spheres. Simulations were performed on a cubic box with unit volume. The hard-sphere radius, R, is calculated from the specified value of ϕ. The value of α was determined as a function of ϕ and the results were fitted to the free-volume expression of scaled particle theory, eq 1.

Using eq 1, we determined the function $q(\phi)$. A representative plot of q as function of ϕ for $l/R = 0.067704$ and $M = 101$ is shown in Fig 2. The value of $M = 101$ corresponds to the number of backbone bonds in PEG1450. We can observe that q decreases as ϕ increases.
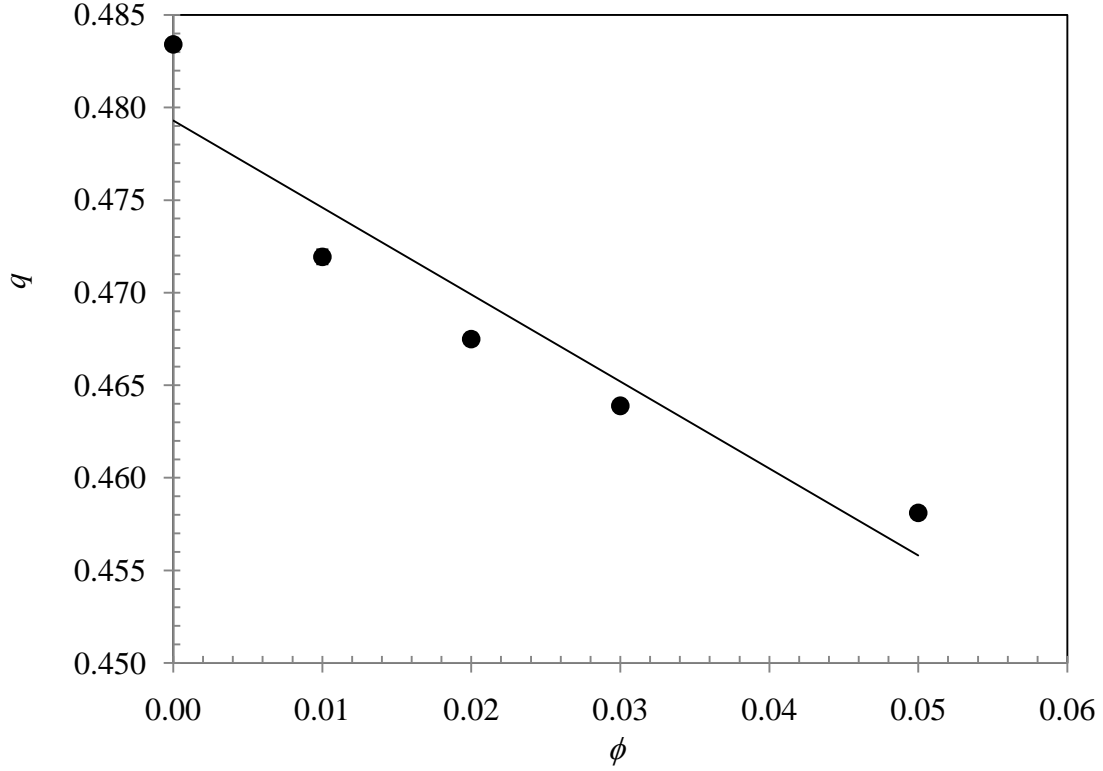
Figure 2. Simulation results for q as a function of ϕ for a coil with M=101 and l/R= 0.067704. Our results do not fit well a linear behavior.

However the dependence of q on ϕ is not linear. Interestingly, as we can see in Fig 3, a linear dependence on $\sqrt{\phi}$ well describes the behavior of q. Hence, in order to determine the value of q in the limit of $\phi = 0$ (i.e. in the absence hard-sphere interactions), we have used a weighted linear regression to fit q as a function of $\sqrt{\phi}$ with a linear function. For the case shown in Fig 3, the value of $q = 0.4832\pm0.0007$ is obtained by extrapolation. In Fig 3, we also show our simulations results for an equivalent AO sphere with R_{AO}/R=0.4832 as a control. The corresponding $q(\phi)$ (shown in Fig 3) was obtained using eq 2. We obtain $q(\phi) = R_{HS}/R = 0.4832\pm0.0008$ is independent of ϕ within the simulation error as expected. The observed decrease of q as ϕ increases implies that the polymer coil is penetrable and/or deformable as the hard-sphere concentration increases.
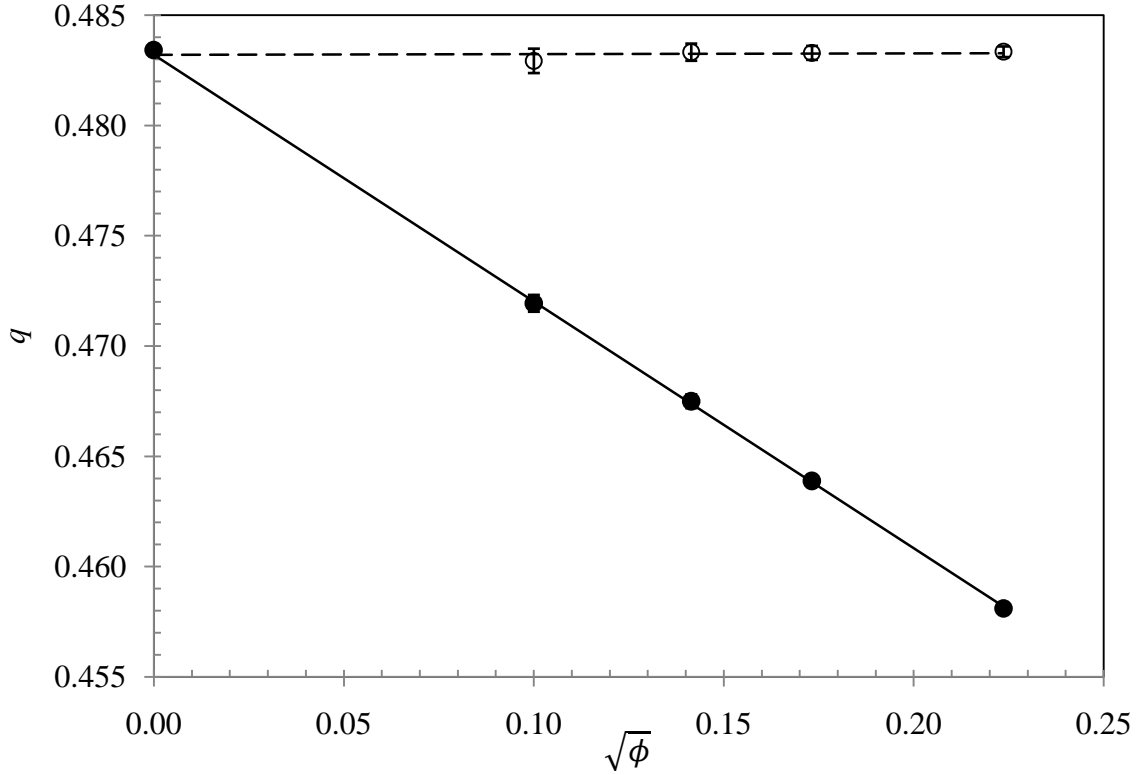
Figure 3. Depletion-layer thickness, q as a function of $\sqrt{\phi}$ for a coil with M=101 and l/R= 0.067704. (solid circles) Data are fitted to a linear function (solid line). The value of q is obtained is 0.4832±0.0007. The slope is -0.113±0.004. Data for an equivalent AO sphere with R_{AO}/R= 0.4832 (open circles) are also fitted to a linear function (dash line). The value of q obtained is 0.4832±0.0008. The slope is 0.0003±0.004.

For a given polymer coil with M segments of length l, we have also determined its gyration radius, R_g, as a multiple of R as (Appendix G). Note that R_g/R for the segment-chain depends on M and l/R. We have therefore characterized the dependence of R_g/R as a function of ϕ. A representative result of R_g/R is shown in Fig 4. We observed that the R_g/R decreases as ϕ increases. This is consistent with coil deformability. However, that the obtained change of R_g/R in the investigated range of ϕ is within 0.6%. This change is much lower than the corresponding change of 5.3% observed for q. That q decreases

significantly more than R_g/R indicates that the effect of penetration is more important than that of deformation. Similar results for $q(\phi)$ and $R_g/R(\phi)$ have been also obtained for other values of l/R (appendix I table 2, 3).
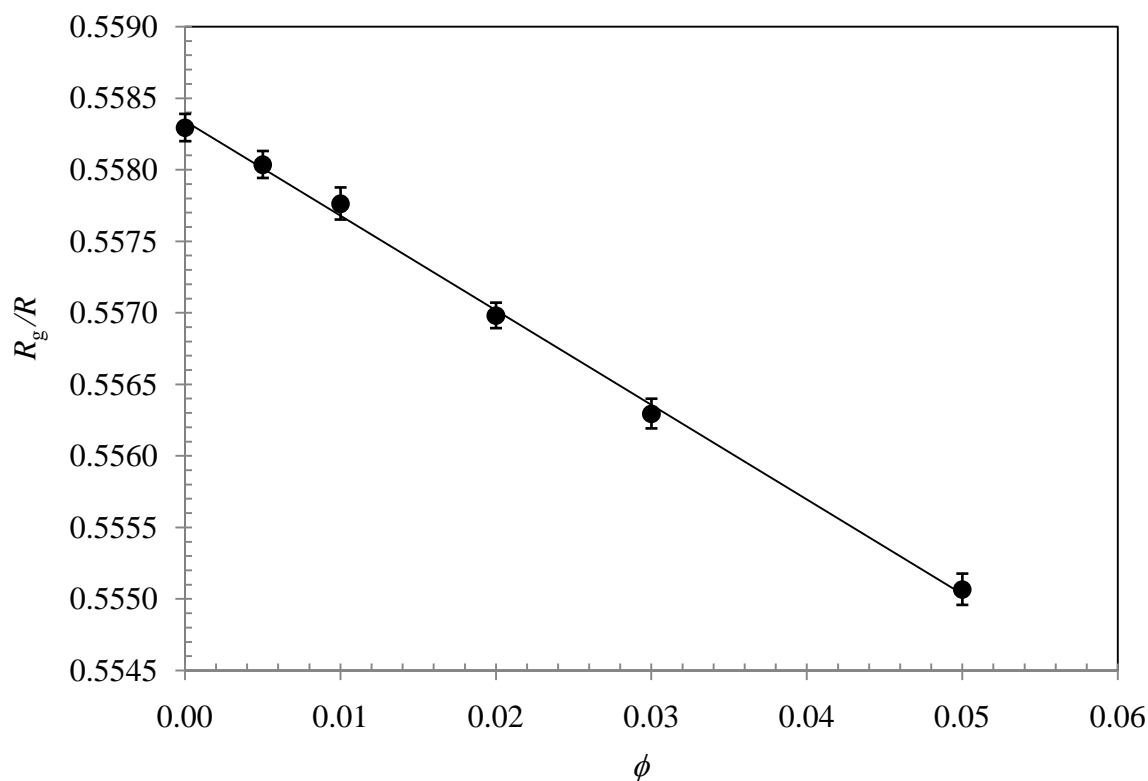
Figure 4. R_g dependence on ϕ for a segment chain with M=101 and l/R= 0.067704.

Using the values of q and R_g in the limit of ϕ= 0, we have determined the correction factor for the segment chain with M=101 by using $f = q\,R/R_g$. For M=101, R_g/R was changed by choosing different ratios of l/R. The corresponding correction factors f were calculated and reported in Fig 5 at different R_g/R. In Fig 5, the theoretical values of correction factors for ideal chain are also shown. They are calculated using the following equation based on Eisenriegler theory (see appendix B),

$$f = \frac{R}{R_g}\left[\left(1 + \frac{6}{\sqrt{\pi}}\left(\frac{R_g}{R}\right) + 3\left(\frac{R_g}{R}\right)^2\right)^{\frac{1}{3}} - 1\right] \qquad (2)$$

$$f = \frac{2}{\sqrt{\pi}} \qquad \text{when} \quad \frac{R_g}{R} \to 0 \qquad (3)$$

The vertical dash line in Fig 5 indicates the value of correction factor at $R_g/R = 0.556$. This value corresponds to the BSA-PEG1450 pair. This Fig shows that the segment-chain is more penetrable than the ideal chain in the presence of spherical particles. Thus the reduced thickness of the depletion layer obtained for the segment chain, $q = 0.48$, is lower than that for the ideal chain, $q = 0.57$ based on Eisenriegler theory. The q value for the segment chain is closer to the experimental value of $q = 0.31$.
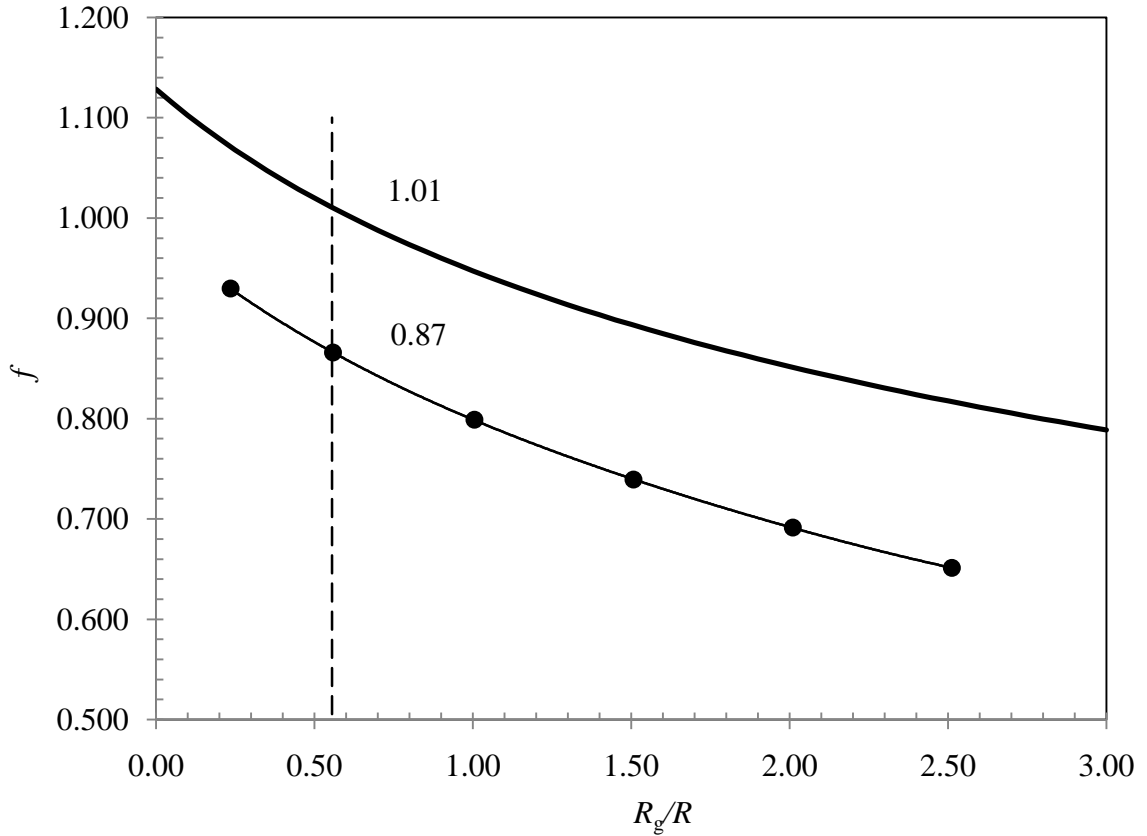
Figure 5. Correction factor as a function of the ratio R_g / R. The solid curve is f for ideal chain. The solid circles are the values for the segment chain with M =101. The vertical dash line indicates the value at $R_g/R = 0.556$ for the BSA-PEG1450 pair.

We attribute the increase in penetrability observed for the segment chain compared to ideal chain to the discrete nature of the changes of polymer conformation occurring in the presence of spheres. We verify this conclusion by determining f as a function of M at constant R_g/R. In Fig 6, we plot f as a function of $M^{-0.5}$ to demonstrate that, as the number of segments increases (and discreteness of chain decreases) at constant R_g/R, f increases. In the limit of $M \to \infty$ (i.e. infinitesimally short segment length), the segment chain coincides with the ideal chain. The value of f extrapolated in this limit is 0.99 for R_g/R=0.556. This value is 2% lower than the theoretical value of 1.01 obtained using eq 2. A similar discrepancy from theory was obtained from Monte Carlo simulations reported by Tuinier and Lekkerkerker on the ideal chain.[95]
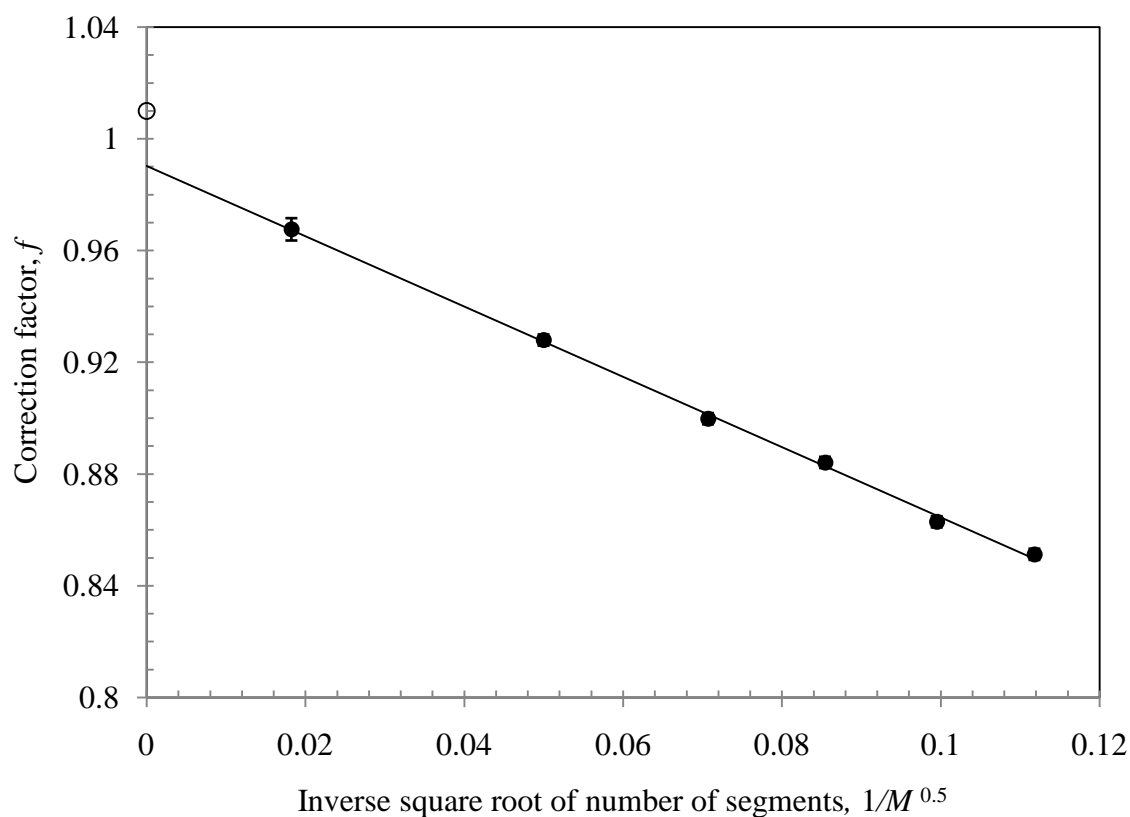
Figure 6. Correction factor of coils increase as a function of $1/M^{0.5}$ (solid circles). The straight line is a linear fit through the data. The open circle represents the value obtained from eq 2 using R_g/R=0.556.

To further investigate the role of polymer structure on $f\left(R_g/R\right)$, we examined the effect of a fixed angle between neighboring segments. In other word, we have examined a fixed-angle segment chain.
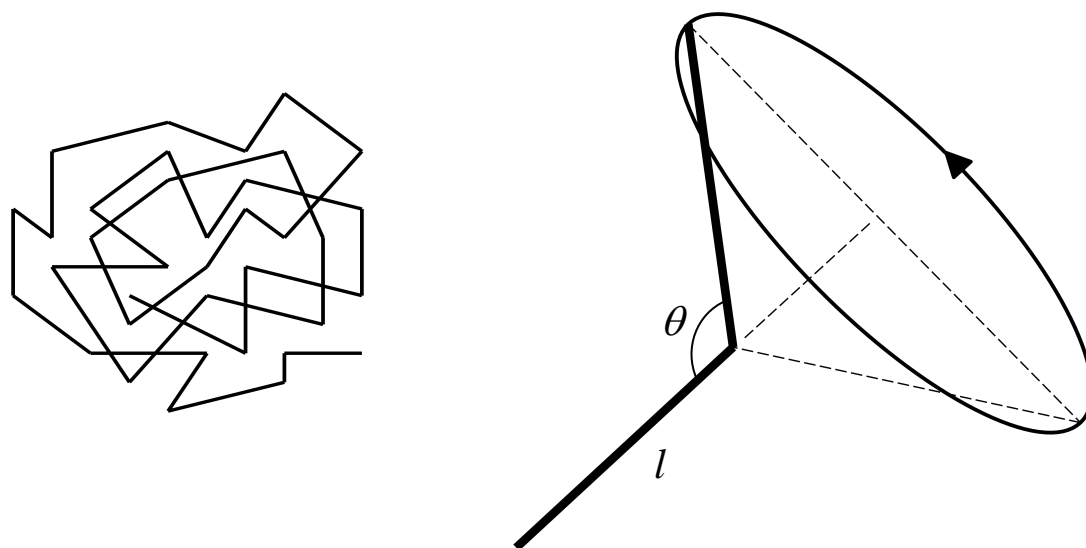
Figure 7. Fixed-angle segment chain. The fixed angle is θ.

Here, a segment can only rotate with respect to its neighboring segments and the angle, θ, between segments is fixed as shown in Fig 7. Simulations have been performed with the chain with $M=101$ and at several fixed angles: $\theta = 30°$, $60°$, $90°$, $109.5°$, $120°$ and $150°$. The corresponding correction factors f are plotted in Fig 8 as a function of R_g/R. We observe that the presence of a fixed angle generally reduces penetrability to the segment chain. Moreover, penetrability for a given R_g/R and M increases as θ increases. We attribute this behavior to an increase of compactness as θ decrease. Interestingly, f for the fixed-angle segment chain with $\theta=120°$ nearly coincides with f for the segment chain. In the case of PEG, we have $\theta=109.5°$. For this angle, f is only about 2% higher than f

for the segment-chain. For the BSA-PEG1450 pair ($R_g/R = 0.556$), we obtain $q = 0.49$ at $\theta = 109.5°$.
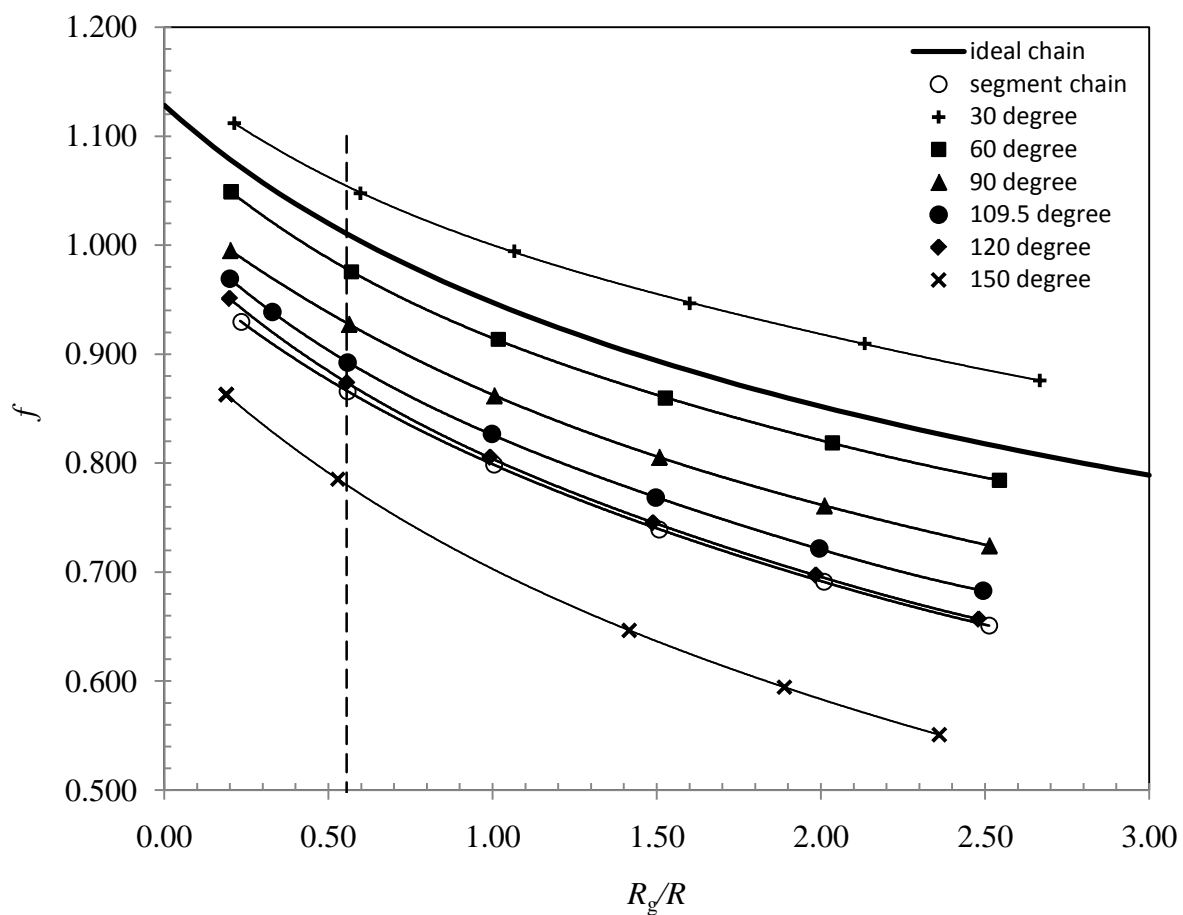
Figure 8. Correction factors as a function of R_g/R and θ for $M=101$.

4. Inconsistency of the Gyration Radius and Bond Length of Polymer

In Section 3, we have demonstrated that a segment chain is more penetrable than an ideal chain with the same R_g/R. Then, by introducing a fixed bond angle, we can mimick PEG1450 structure by setting $M = 101$ and $\theta = 109.5°$. The obtained value of q of 0.49 is closer than that of the ideal chain to the experimental value. Our simulations could be used not only to determine R_g/R but also R_g for the fixed-angle segment chain as a function of M and θ. In order to determine absolute values of R_g for PEG, we need to know the value of the segment length l. PEG backbone consists of 1/3 C-C bonds and 2/3 C-O bonds. We have used the values of 1.54Å and 1.43Å for the C-C and C-O bonds in ethylene glycol respectively.[211] We have determined the value of $l = 1.47$ Å as a weighed average between these two bond-length values. If we set the length of the segment to $l = 1.47$ Å, we can determine R_g as function of ϕ for a given R. We show our simulation results in Fig 9 for the segment chain and the fixed-angle segment chain with $\theta = 109.5°$. We further observe that R_g can be also calculated at $\phi = 0$ using well-established theoretical equations:

$$R_g = \sqrt{\frac{Nl^2}{6}} \qquad \text{for segment chain} \tag{4}$$

$$R_g = \sqrt{\frac{Nl^2 \left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\cos^k(\pi - \theta)\right]}{6}} \qquad \text{for fixed} - \text{angle segment chain}$$

In Fig 9, we can see that R_g does not significantly depend on ϕ in both cases. We can also see that the theoretical values obtained using eq 4 agrees quite well with our simulations.

The obtained value of R_g for $M = 101$ and $\theta = 109.5°$ can be compared with the experimental value of $R_g = 15$ Å for PEG 1450.[89] This experimental value is 80% larger than the theoretical one for both chain models. Clearly, this large discrepancy leads to an inconsistency between the R_g value calculated using structural parameters of PEG1450 (i.e., $M = 101$ and $l = 1.47$ Å) and the experimentally observed R_g. To remove this inconsistency, we have introduced a novel two-parameter model so that the theoretical value of R_g obtained for the chosen values of M and l reproduces the experimental value of R_g.
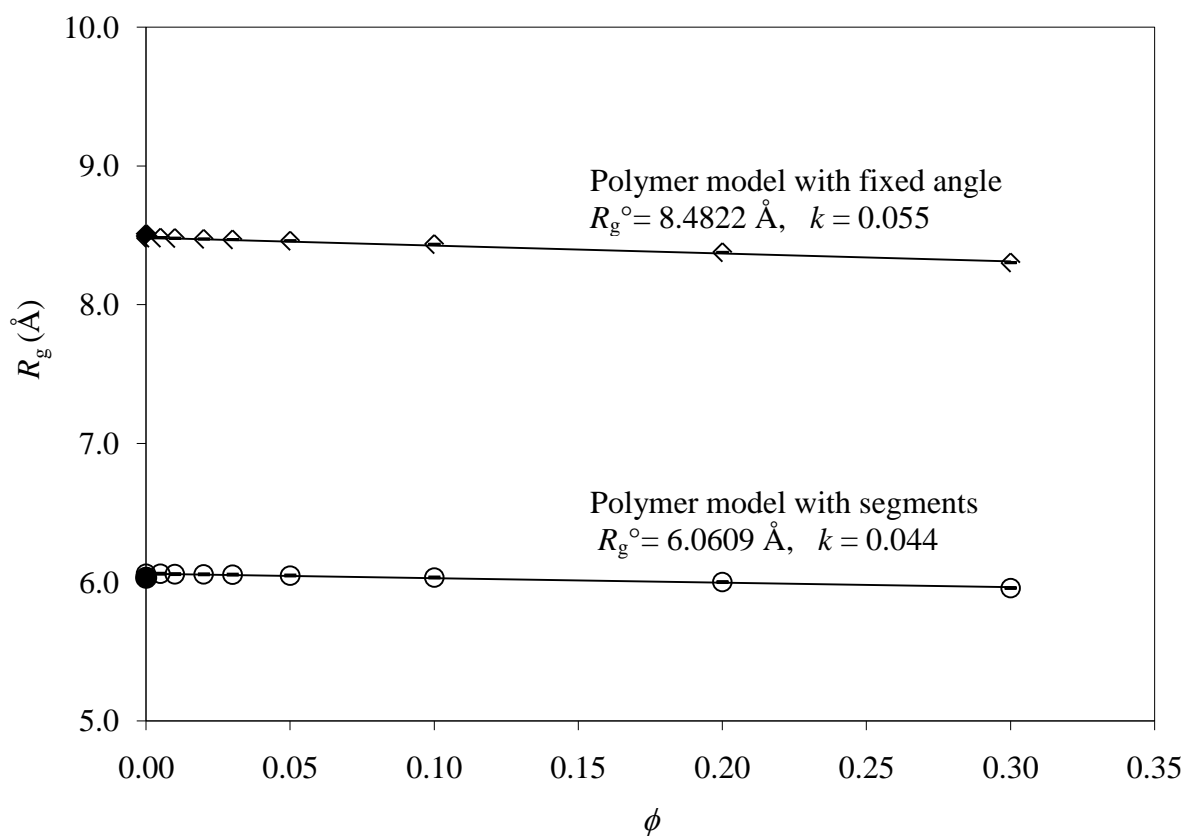
Figure 9. Gyration radius, R_g, for the segment chain with $M = 101$ and $l = 1.47$ Å (open circle) and the corresponding fixed-angle polymer chain with $\theta = 109°$ (open diamond). The data are fitted with linear function: $R_g = R_g° (1 - k\phi)$. The solid circle and diamond represent the theoretical values of gyration radius.

5. Two-Parameter Self-Avoiding Chain Model

The experimental gyration radius of PEG scales with molecular weight according to

$$R_g = b(M_w/M_w^\circ)^a \tag{5}$$

where $a = 0.55$ and $b = 0.27$ Å in water at 293.15K.[89] In eq 5, M_w is the molecular weight of PEG and $M_w^\circ = 1$g/mol. For an ideal chain, which has no rotational restriction, the scaling exponent is 0.50 independent of M. However, a real polymer chain is also affected by steric exclusion effects between segments. This results in an expansion of the coil and a scaling exponent larger than 0.5. [212] Moreover, the solvent also contributes to the value of a. If the polymer-solvent interaction is energetically favorable (good solvent), the polymer chain will expand and a increases.[212] In contrast, the polymer chain will contract and a decreases when the segment-segment interaction is favorable (poor solvent). [212] In our experimental conditions of $T = 298$K and lower temperatures, water is a good solvent for PEG. We now observe that the steric exclusion effect can be modeled by assuming that each segment end point is a small hard sphere with radius r. Values of a larger than 0.5 can be successfully obtained in this way. However, to reproduce the experimental value of b in eq 5, we need to introduce another parameter in the coil model. As a second parameter, we introduce the energy difference between *trans* and *gauche* conformations in the polymer chain. In summary, the experimental gyration radius of PEG, and its dependence on molecular weight can be obtained by introducing a self-avoiding chain model that includes two adjustable parameters: an effective hard-sphere radius for the segment end points and a *trans-gauche* transition energy.

In Fig 10, we illustrate the two important features of this model. In Fig 10a, we can see that the solid black circles with radius r characterize the steric effects associated with segment end points. The angle Φ has been introduces to characterize conformational states. In Fig 10b, the conformational potential energy $V(\Phi)$ used in our model is shown. Specifically, we set $V(\Phi) = 0$ for $0 \leq \Phi < \pi/2$ (*trans* state) and $V(\Phi) = \epsilon$ for $\pi/2 \; \Phi < \pi$ (*gauche* state).
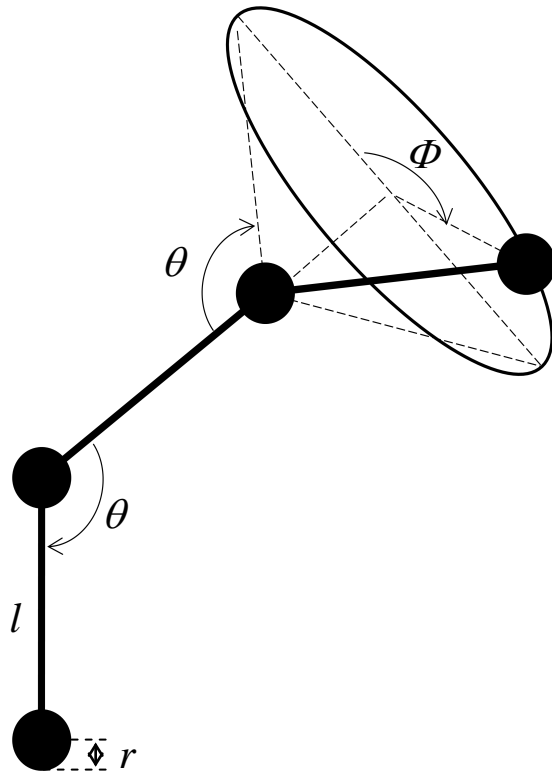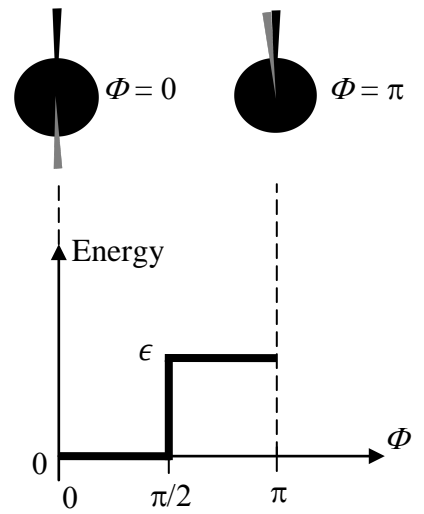
(A)   (B)

Figure 10. Self-avoiding chain model with segment length, l, bond angle, θ, and two adjustable parameters, the effective atomic radius $\hat{r}$ and the trans-gauche transition energy, $\hat{\epsilon}$.

Using this two-parameter chain model, we have determined R_g at $\phi=0$ as a function of M and the reduced parameters $\hat{r} \equiv r/l$ and $\hat{\epsilon} \equiv \epsilon/l$. Specifically, we performed simulations with self-avoiding chains with three values of M (101, 137 and 206) and calculated the corresponding scaling factor a using eq 5. Values for R_g/l for $M=101$ and θ $=109.5°$ and the corresponding values of a for different sets of $\hat{r}$ and $\hat{\epsilon}$ are reported in Table 4 of Appendix I. Our results are also shown in Fig 11.
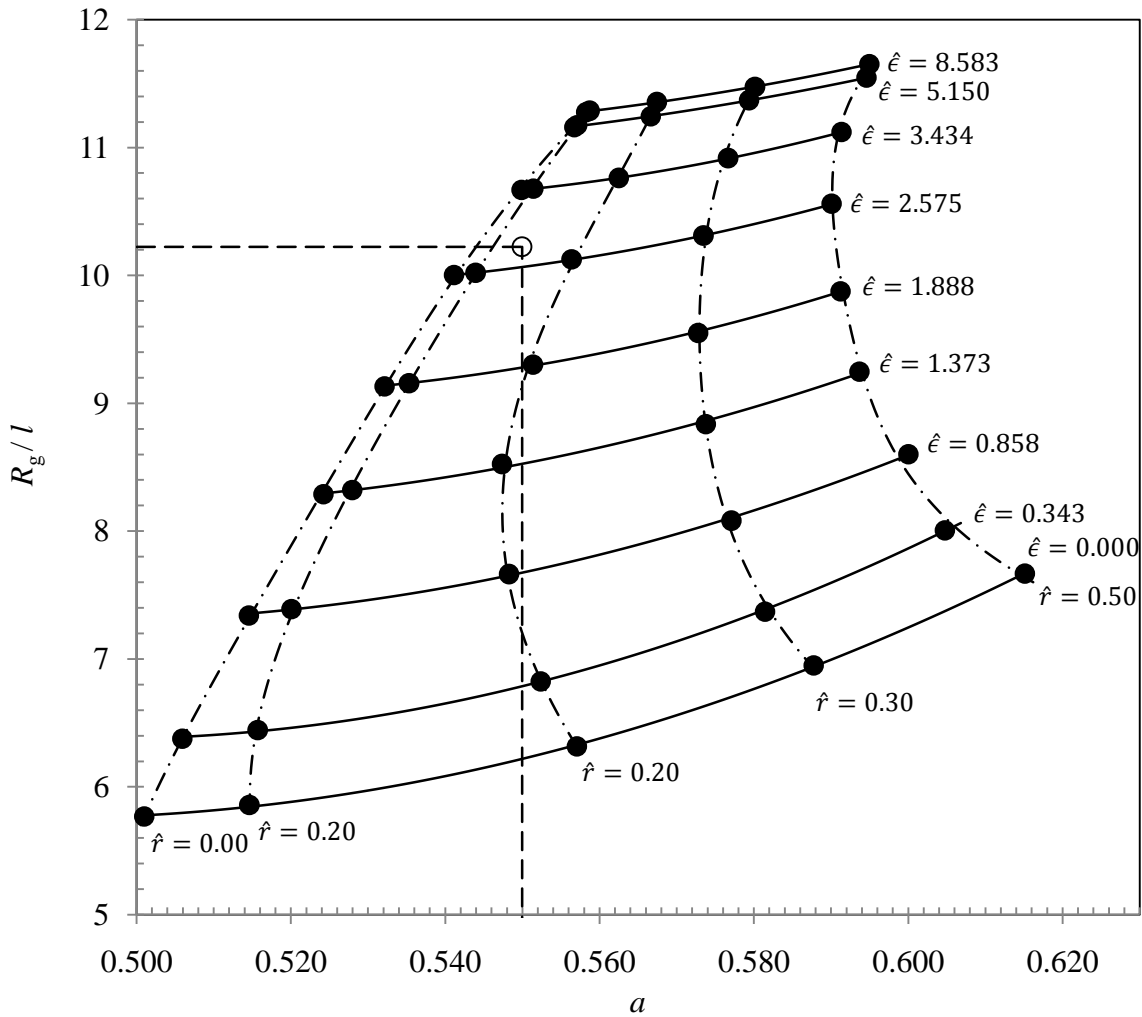
Figure 11. Values of R_g/l and a calculated for several sets of the two parameters $(\hat{r}, \hat{\epsilon})$. The coordinates of $(a, R_g/l)$ are denoted by solid circles. The solid curves indicate the points with same $\hat{\epsilon}$, and the dashed dotted curves indicate the points with same $\hat{r}$. The chosen values of $\hat{r} =0.15$ and $\hat{\epsilon} =2.747$ (open circle) best fit the experimental values of $a=0.55$, $b=0.27$.

We now examine the behavior of R_g in Fig 11. The solid curves connect points with same $\hat{r}$ and dashed curves connect points with the same $\hat{\epsilon}$. In Fig 11, when $\hat{r} = 0$ and $\hat{\epsilon} = 0$, there is no rotational restriction on the azimuth angle Φ and the scaling exponent determined by simulation is equal to 0.50 within the simulation error ± 0.001. This value corresponds to the value of ideal chains. We can also see that R_g and a both increase with $\hat{r}$, at constant $\hat{\epsilon}$. This is consistent with steric effect favoring the extended conformation of polymer chain and a corresponding stronger dependence of R_g on M.

We now examine the behavior of R_g and a as a function of $\hat{\epsilon}$ at constant $\hat{r}$. As $\hat{\epsilon}$ increases, the more extended *trans* states become increasingly more favored compared to the *gauche* ones. Thus, R_g increases with $\hat{\epsilon}$ as shown in Fig 11. However, the dependence of scaling exponent, a, on $\hat{\epsilon}$ is not always monotonic. To describe the observed behavior of a, we consider the following expression for a obtained from differentiation of eq 5 when M is large:

$$a = M\big[\ln(R_g(M+1)) - \ln(R_g(M))\big] \tag{6}$$

The dependence of a on $\hat{\epsilon}$ is then given by:

$$\frac{\partial a}{\partial \hat{\epsilon}} = M\left[\frac{\partial \ln(R_g(M+1))}{\partial \hat{\epsilon}} - \frac{\partial \ln(R_g(M))}{\partial \hat{\epsilon}}\right] \tag{7}$$

We now consider the case of $\hat{r} = 0$. At $\hat{\epsilon} = 0$ both *trans* and *gauche* states have equal probability. However, as $\hat{\epsilon}$ increases, *trans* states become more probable than *gauche*

states. Since R_g increases with the frequency of *trans* states, $\partial \ln(R_g)/\partial \hat{\epsilon}$ is positive. We now consider the effect that M has on $\partial \ln(R_g)/\partial \hat{\epsilon}$. In Fig 12, we plot $\ln R_g$ as a function of $\hat{\epsilon}$ for two different values of M (see solid curves). From the initial slopes, we can see that $\partial \ln(R_g)/\partial \hat{\epsilon}$ increases with M. This can be understood by observing that the higher M, the higher the number of gauche to *trans* changes occur upon increasing $\hat{\epsilon}$. This implies that the difference on the right side of eq 7 is positive. Thus $\partial a/\partial \hat{\epsilon} > 0$ consistent with the behavior obtained for $\hat{r} = 0$ and $\hat{r} = 0.1$ in Fig 12. We now introduce the effect of $\hat{r}$. Since $\hat{r} > 0$, the probability of *trans* states is higher than that of *gauche* states even at $\hat{\epsilon} = 0$. As $\hat{r}$ increases at $\hat{\epsilon} = 0$, the probability of gauche states reduces due to the corresponding increase of steric effects. We therefore obtain that $\partial \ln(R_g)/\partial \hat{\epsilon}$ decreases as $\hat{r}$ increases (see the initial slope for solid and dashed curves in Fig 12). As M increases at constant $\hat{r}$ and $\hat{\epsilon} = 0$, the number of *gauche* states may even decrease due to a corresponding increase of steric effects. This implies that the corresponding number of *gauche* to *trans* changes upon increasing $\hat{\epsilon}$ may also decrease. In this case, the difference on the right side of eq 7 is negative and $\partial a/\partial \hat{\epsilon} < 0$. This is consistent with the observed behavior at low $\hat{\epsilon}$ and large $\hat{r}$. As $\hat{\epsilon}$ increases, the role of steric exclusion becomes relatively weaker and $\partial a/\partial \hat{\epsilon} > 0$ is recovered.
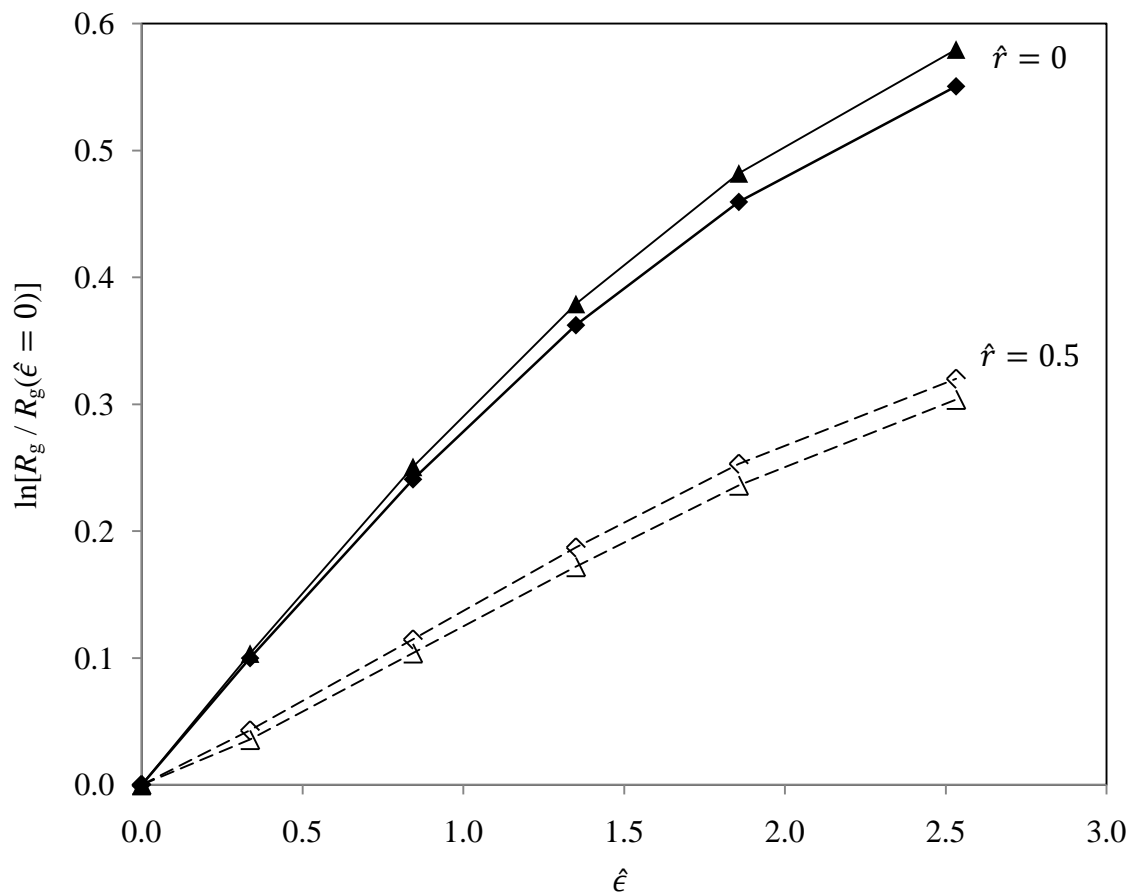
Figure 12. Logarithm of the ratio R_g/R_g^o as a function of the reduced *trans-gauche* conformational transition energy, where R_g^o is R_g at $\hat{\epsilon} = 0$. The diamond and the triangle marks denote the polymer chain with 101 and 206 segments respectively. The solid lines with solid marks represent the polymer chain with $\hat{r} = 0$, while the dash lines with open marks represent the chain with $\hat{r} = 0.5$.

6. Determination of q using the Self-Avoiding Chain Model

In the previous section, we have introduced a two parameter self-avoiding chain model. Using the obtained map of a-R_g/l data (see Fig 11), we have then identified the values of the two parameters ($\hat{r} = 0.15, \hat{\epsilon} = 2.701$) that best fitting the experimental scaling properties of PEG. The value of $\hat{\epsilon} = 2.701$ corresponds to a molar trans-gauche transition energy of $\epsilon = 1600$ cal/mol at 298.15°C. Interestingly, this value is very close to the experimental value of $\epsilon = 1400 \pm 150$ cal/mol for pure diethyl ether in the liquid state at the same temperature.[213]

For this specific case, we have performed simulations on the free-volume as a function of ϕ to determine q in the limit of $\phi = 0$. The correction factors f as a function of R_g/R have been plotted in Fig 13.
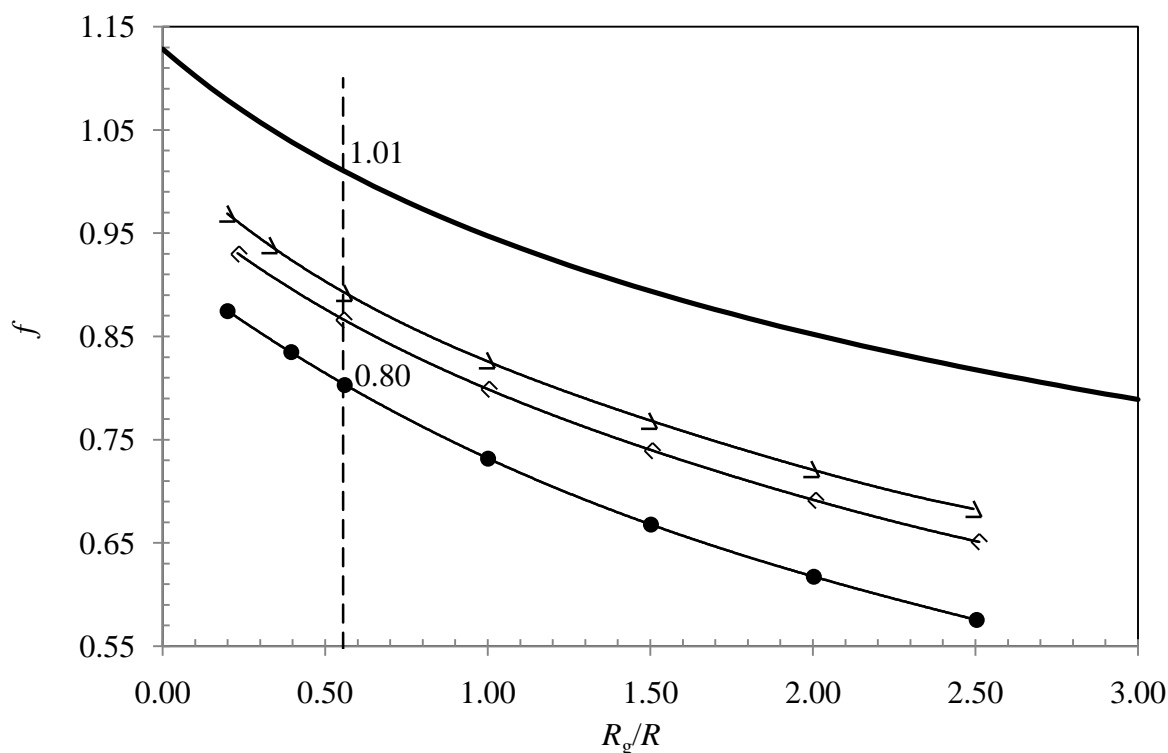
Figure 13. Correction factor f as a function of R_g/R: ideal chain (solid curve), segment-chain (solid diamonds), fixed-angle segment chain (open circles), self-avoiding chain (closed circles). The vertical dashed line corresponds to the R_g/R value for PEG1450-BSA pair.

In Fig 13, the correction factor for the BSA-PEG1450 pair (R_g/R=0.556), is equal to 0.80, and the corresponding q value is 0.44. This value is in better agreement with the experimental one than those previously obtained for the segment chain and the fixed-angle segment chain. This implies that the two-parameter self-avoiding chain is more penetrable. This conclusion can be related to a corresponding larger chain expansion. We summarize the q values obtained for the BSA-PEG1450 pair for each polymer model in Table 1. Therefore, we conclude that the conformation of polymer chain in solution contributes significantly to the depletion interaction between protein molecules caused by non-absorbing polymer.

Ideal Chain	Segment Chain	Fixed-angle	Self-avoiding	Experimental
$q = 0.56$	$q = 0.48$	$q = 0.49$	$q = 0.44$	$q = 0.31$

Table 1. The experimental, theoretical and simulation values of reduced depletion layer thickness, q, of BSA/PEG1450 pair.

7. Effect of ϕ on q

The q values obtained from simulations were all extracted in the limit of infinite dilution, i.e. $\phi = 0$. Because the polymer coils can be penetrated by the spherical particles, the value of q will depend on ϕ as already shown for the segment chain (see Fig 3). We have used our simulation results, $\alpha(\phi)$, to determine $q(\phi)$ using eq 1. Our results are shown in Fig 14. The q value decreases as the volume fraction increases. This is

consistent with spheres being able to penetrate the polymer coil and, in turns, reduce the average thickness of depletion layers around them. In Fig 14, we have indicated the experimental range of protein volume fractions given in Chapter 8 by the two dashed vertical lines. Within this range, the change of q values is around 10% (Fig 14). Therefore, the q value for the BSA-PEG1450 pair determined by simulation in the experimental range of protein volume fraction is 0.39 ± 0.02. This value of q is even closer to the experimental value 0.31 ± 0.01 extracted from $T_{\text{ph}}(c_2)$ measurements.
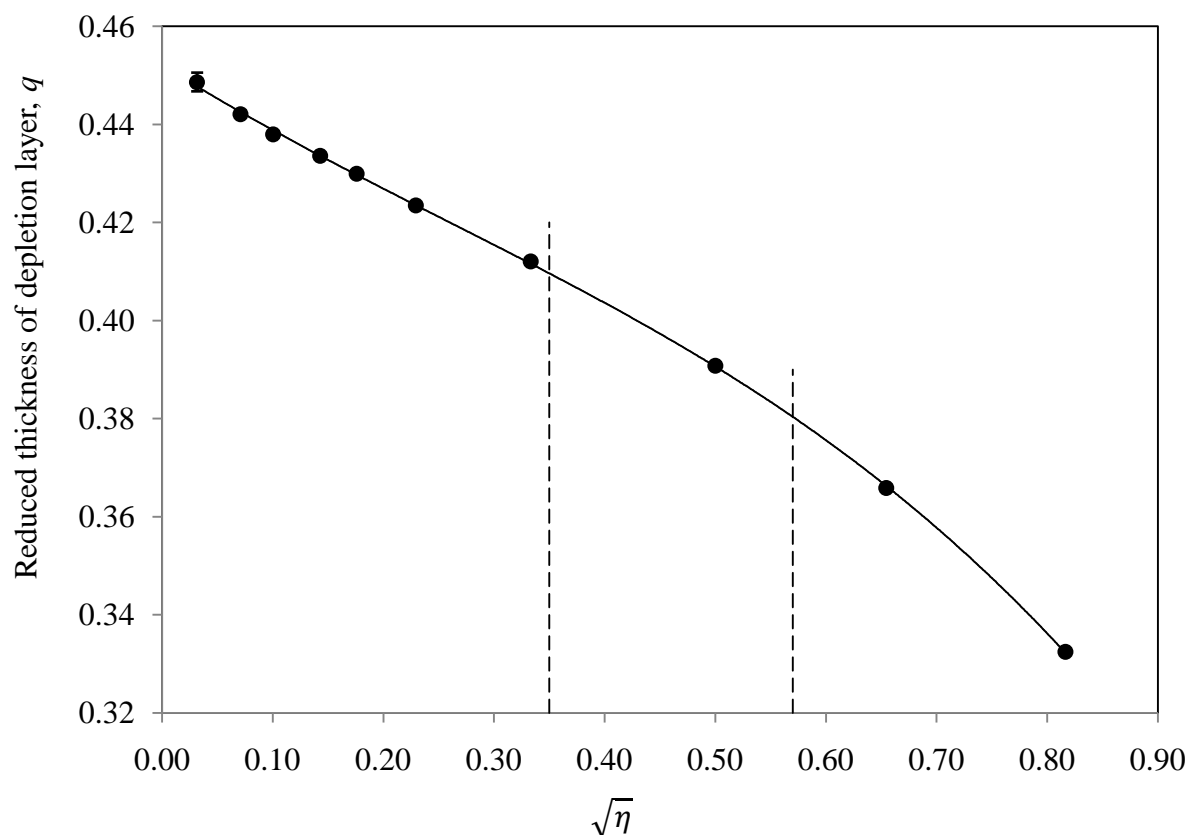
Figure 14. The q values for the BSA-PEG1450 determined by Monte Carlo simulations for the self-avoiding chain at volume fractions. Note that $\eta \equiv \phi/(1 - \phi)$. The experimental range of protein volume fractions described in Chapter 8 is indicated by the two dashed vertical lines. For comparison, we also include our simulation results for the corresponding segment chain.

In summary, we have constructed a polymer model that allows us to determine more accurate estimates of q from the microscopic parameters such as R_g and R. This yields values of q closer to the experimental one for the BSA-PEG1450 pair. It is important to observe that estimation of q can be further improved by considering other factors such as the deviation from the spherical shape and attractive interaction between the polymer coil and the protein. These factors should be assessed in future work.

Chapter 10

Computer Simulation Study on the Effect

of Temperature on Free-Volume Fraction

1. Introduction

In the case of a hard-sphere suspension, the free-volume fraction, α (ϕ), is independent of temperature. However, if attractive interactions are added to the hard-core repulsion of hard-spheres, α becomes a function of T. This is directly relevant to protein solutions, for which protein-protein attractive interactions are often encountered. Attractive interactions promote particle clustering. We therefore expect that α is larger than that predicted for pure hard-spheres. The effect of temperature on α can be described by introducing the following high-temperature power-series expansion:

$$\ln\alpha = \ln\alpha^{(0)} + \zeta_1 \frac{1}{RT} + \zeta_2 \left(\frac{1}{RT}\right)^2 + \cdots \tag{1}$$

For simplicity, we can describe attractive interactions using the square-well potential (see Section 2 of Chapter 8). We can rewrite the previous expression as:

$$\ln\alpha = \ln\alpha^{(0)} + \hat{\zeta}_1 \frac{\varepsilon}{RT} + \hat{\zeta}_2 \left(\frac{\varepsilon}{RT}\right)^2 + \cdots \tag{2}$$

where ε is the reduced depth of potential well, and the coefficients $\hat{\zeta}_i = \zeta_i/\varepsilon^i$ is a function of the reduced range of square-well potential, λ, and the volume fraction, ϕ. It is important to note that $\alpha^{(0)}$ is the free-volume fraction at $T\rightarrow\infty$. This coincides with the free-volume fraction for pure hard spheres. We have performed particle insertion into a hard-sphere suspension with attractive interactions. For simplicity, we have considered the case in which the inserted particle is a AO sphere (Chapter 1) with radius qR. We first use our simulations to examine the effect of temperature on free-volume fraction that is directly relevant to our experimentally investigated BSA-PEG1450-buffer system. We will then describe the effect of temperature as a function of q and λ more generally.

2.   Effect of Attraction on α.

Implications for the BSA-PEG1450-buffer system

Our measurements of $T_{ph}(c_2)$ and second virial coefficient described in Chapter 8 indicate that protein-protein interactions are attractive for the BSA-buffer system at pH 5.2. Modeling these interactions using a square-well potential has allowed us to obtain ε = 1.5± 0.2 J/mol and λ =1.5. The value of q that best describes our results is $q = 0.31$, which is significantly smaller than the value of $q = 0.56$ obtained from R_g/R and $f(R_g/R)$, where $f(R_g/R)$ is the correction factor for an ideal chain. In the previous chapter, our simulations together with a more accurate polymer model for PEG1450 have shown that $q=0.39$ is a better microscopic estimation of the reduced thickness of depletion layer.

To examine the role of temperature on α, we have performed simulations to determine α as a function of ε at $\lambda=1.5$ and at several values of ϕ (0.1, 0.2 and 0.3). We have specifically considered the case of $q = 0.25$ and $q = 0.50$ in order to cover both the experimentally-determined value of q and corresponding microscopic estimations. Our results are shown in Fig 1a,b, where we plot $\ln(\alpha/\alpha^{(0)})$ as a function of $\hat{\varepsilon} \equiv \varepsilon/RT$. We note that $\ln(\alpha/\alpha^{(0)})$ describes the contribution of the attractive interaction to the free volume change associated with inserting the AO sphere (see eqs 1,2). It is important to remark that the upper limit of $\hat{\varepsilon}$ was chosen to be low enough that phase transitions were not observed during our simulation. Fig. 1 shows that $\ln(\alpha/\alpha^{(0)})$ increases with $\hat{\varepsilon}$. The effect is larger at high volume fractions of spheres. This behavior can be related to particle clustering due to attractive interactions. The dashed vertical line indicates the value of $\hat{\varepsilon}$ associate with the BSA-buffer system at the lowest experimental temperature

(-15°C). At this $\hat{\varepsilon}$ value and $q = 0.25$, we find that α is 0.1% higher than $\alpha^{(0)}$ at $\phi = 0.1$, and becomes 0.6% higher than $\alpha^{(0)}$ at $\phi = 0.3$. At $q = 0.50$, α is 1.1% higher than $\alpha^{(0)}$ at $\phi = 0.1$, and becomes 10% higher than $\alpha^{(0)}$ at $\phi = 0.3$. We have then used the scaled particle theory (eq 1 in Chapter 9) to obtain the corresponding values of q.   At $q = 0.25$, the obtained q is 1.0% smaller at $\phi = 0.1$, and 1.5% smaller at $\phi = 0.3$. At $q = 0.50$, the obtained q is 3.1% smaller at $\phi = 0.1$, and 5.7% smaller at $\phi = 0.3$. Thus, we remark that the deviations of α and q due to the presence of attractive interaction are smaller or comparable to the experimental error. Thus, neglecting the effect of temperature on α is a valid approximation within our experimental domain.
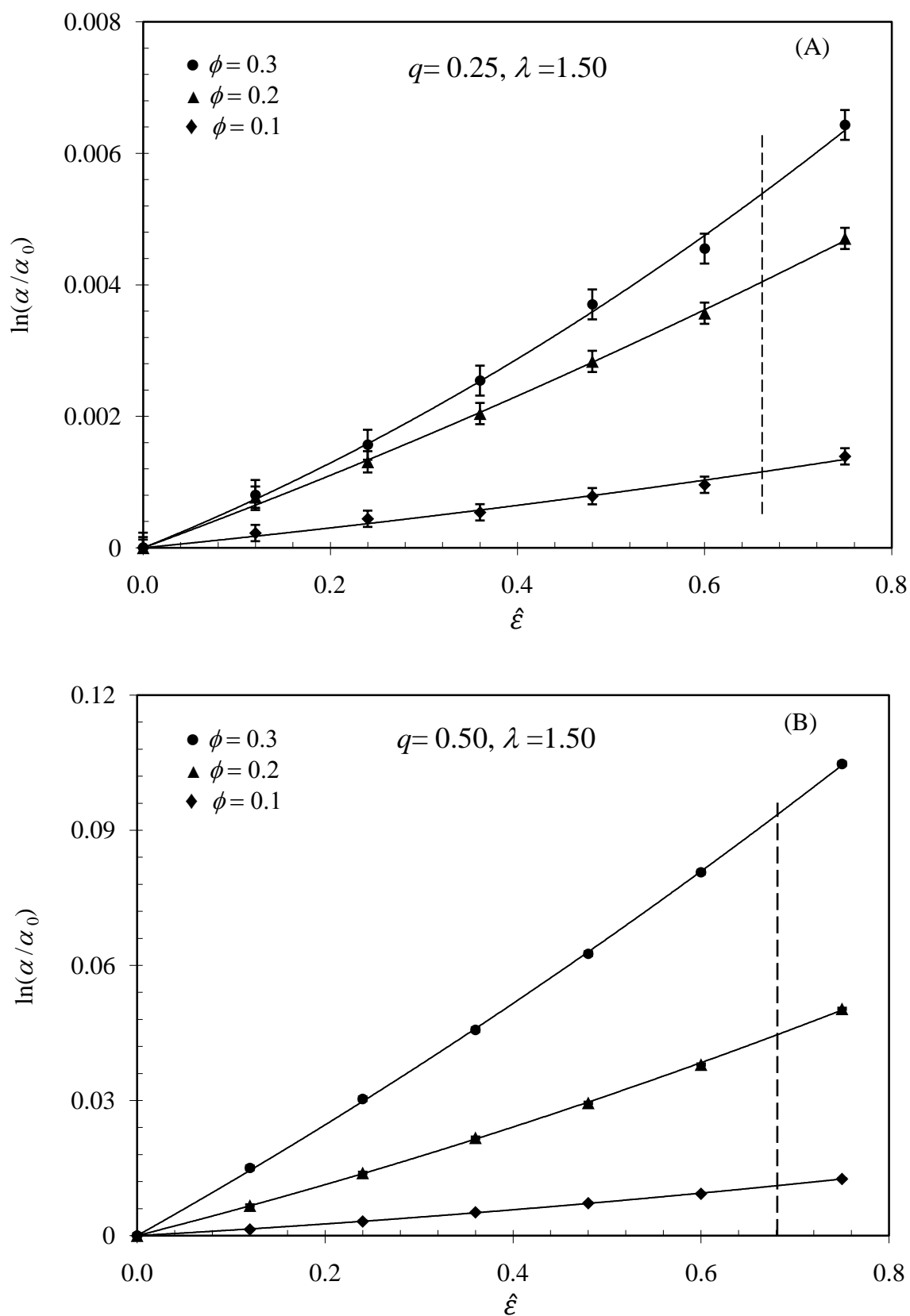
Figure 1. Effect of $\hat{\varepsilon}$ on the free volume fraction α. (A) q=0.25 and λ=1.5; (B) q=0.50 and λ=1.5. Vertical dashed lines indicates ε=1.5kJ/mol and T=-15°C.

3. Dependence of $\hat{\zeta}_1$ on λ, q and ϕ

We now examine the behavior of the first order coefficient, $\hat{\zeta}_1(\lambda, \phi, q)$, in eq. 2. This parameter, when positive, describes the increase in α with respect to $\alpha^{(0)}$. The value of $\hat{\zeta}_1(\lambda, \phi, q)$ was determined by linear extrapolation of $\ln(\alpha/\alpha^{(0)})/\hat{\varepsilon}$ at $\hat{\varepsilon} = 0$ (high-temperature limit). Our simulation results are shown in Fig 2(a,b,c). In this Fig, we plot $\hat{\zeta}_1$ as a function of λ (with $1 < \lambda \le 2$) at several values of ϕ (0.1, 0.2, 0.3 and 0.4) and q (0.25, 0.5 and 1.0). For λ values close to $\lambda = 1$, $\hat{\zeta}_1$ increases with λ. This can be understood by considering the fact that increasing the range of attractive interactions enhances particle clustering. This effect is larger at higher values of ϕ (see Fig 2). We can also see that the higher q the higher $\hat{\zeta}_1$ for given values of ϕ and λ. It is however important to observe that the behavior of $\hat{\zeta}_1$ at relatively larger values of λ is quite complex. We can notice that $\hat{\zeta}_1$ displays a maximum when plotted as a function of λ. Furthermore, $\hat{\zeta}_1$ becomes even *negative* when $\lambda - 1$ is appreciably larger than q.

To physically understand this behavior, we should compare representative particle configurations in the case of small and large λ values. This is illustrated in Fig 3. In Fig 3A, we consider the case of λ small. Here, we can appreciate that the number of particle contacts is larger when clustering occurs. Thus, the right configuration is favored with respect to the left configuration. Correspondingly, for a given q value, the free volume associated with the right configuration is larger than that associated with the left configuration. Since an increase in λ (for λ small) increases clustering, $\hat{\zeta}_1$ is expected to increase with λ. We now consider the case in which λ is large. This is shown in Fig 3B.
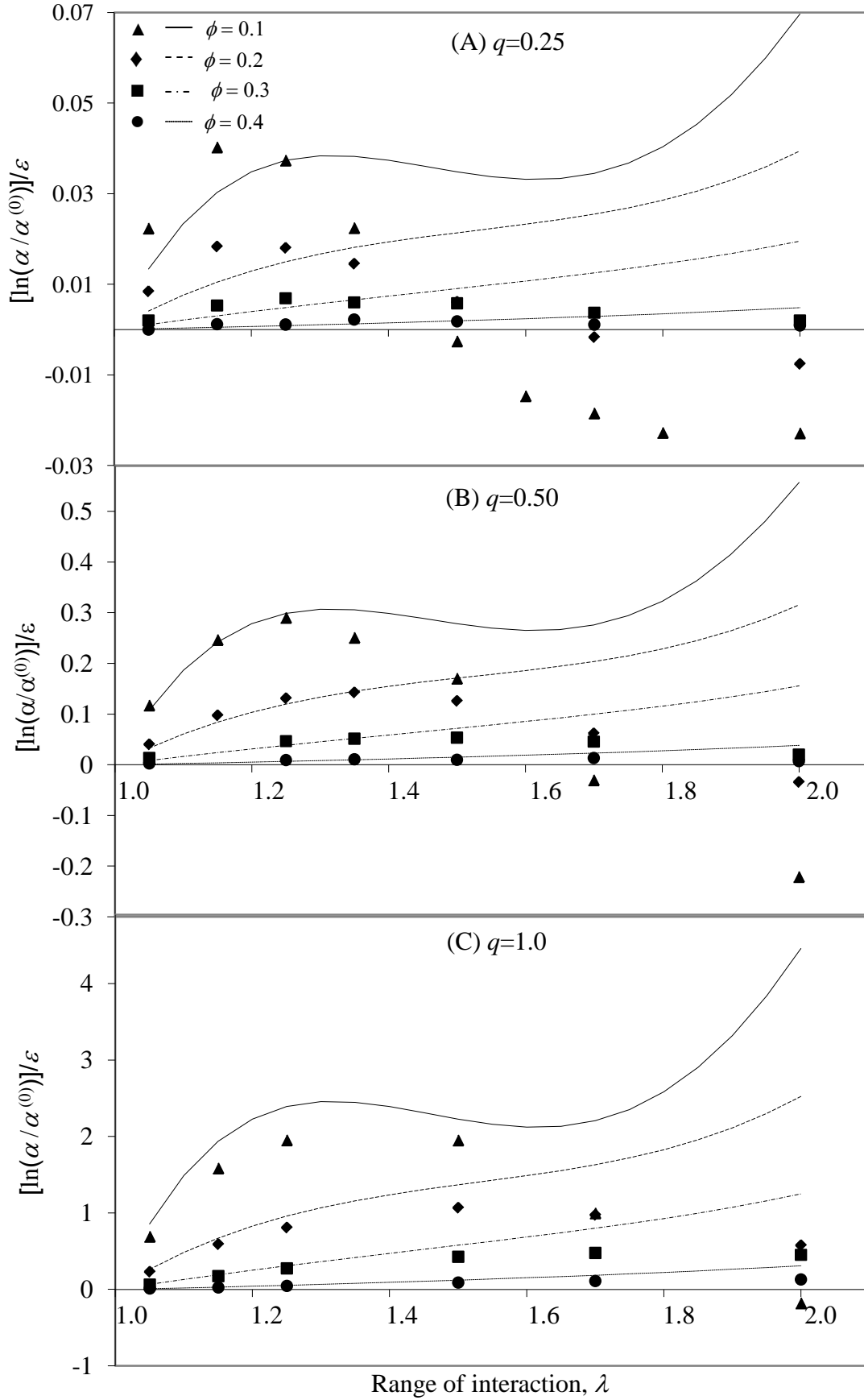
Figure 2. $[\ln(\alpha/\alpha^{(0)})]/\varepsilon$ as a function of λ at several values of ϕ and q.

UNFAVORED

FAVORED

(A) Short λ

FAVORED

UNFAVORED

(B) Long λ
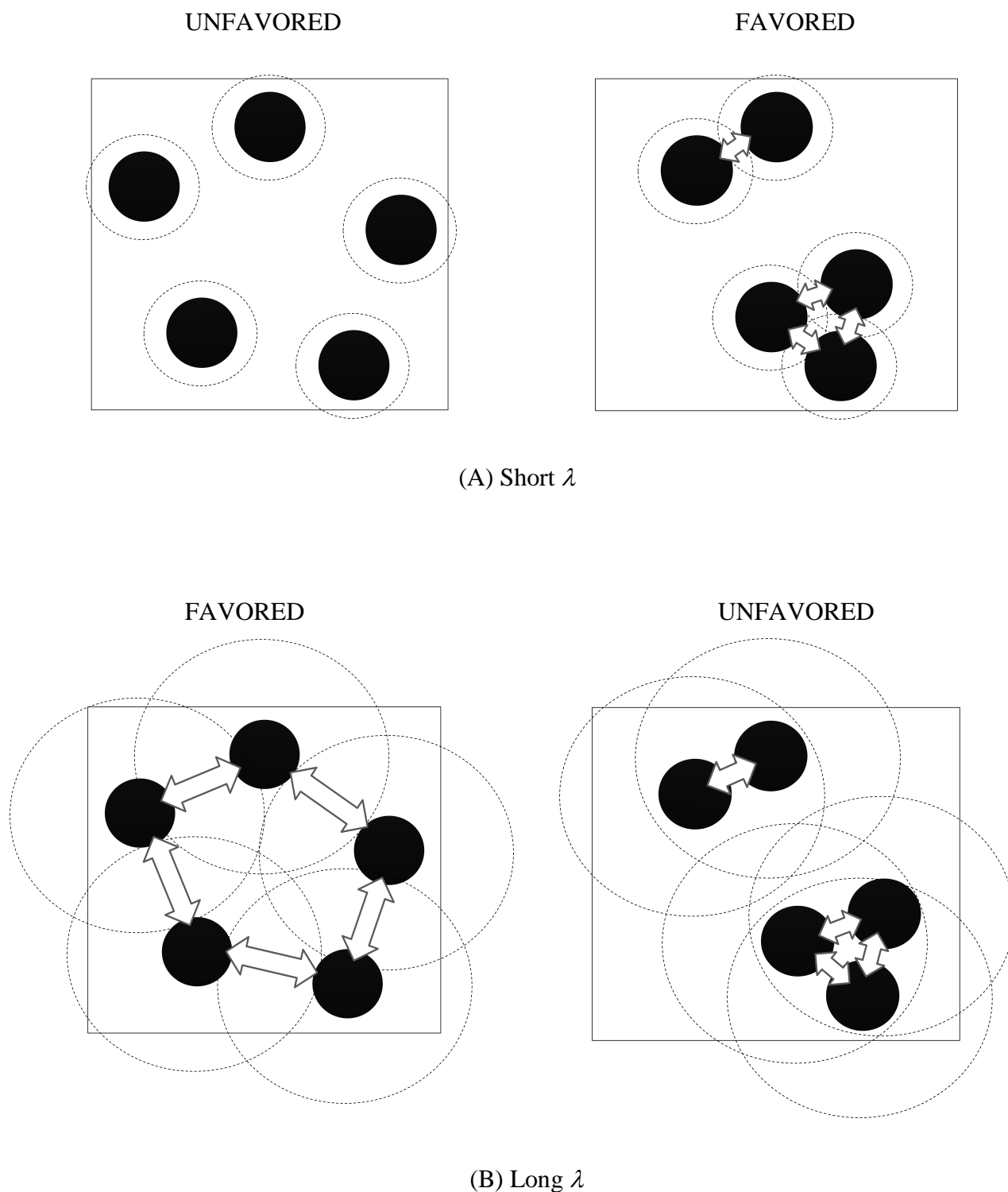
Figure 3. (A) Attraction favors clustering configurations of particles when λ is small; (B) Attraction favors dispersing configurations of particles when λ is large. The solid circles denote hard spheres, the dashed circles denote the range of attractive interaction, and the arrows indicate particle-particle attractive contacts responsible for lowering the system internal energy.

Since λ is large, particle-particle attraction occurs even if particles are not close to each other. In other words, clustering becomes no longer necessary in order to lower the internal energy of the system. On the other hand, particle clustering may even reduce the number of contacts if λ is large as shown in Fig 3B. Thus, the left configuration becomes favored with respect to the right configuration. In these conditions, particle-particle attraction has the effect of *lowering* the free volume. The specific value of λ at which the effect of particle-particle attraction on free volume switches behavior is connected to the thickness of the depletion layer, q, because α depends on q. Fig. 4 shows the effect of range of interaction on free volume.

We can also use scaled-particle theory to determine the expression for $\hat{\zeta}_1(\lambda, \phi, q)$. According to scaled-particle theory,[191,192] the work, W, to create a cavity, which is directly related to free volume, depends on the pressure of the particle suspension, Π:

$$\beta W = -\ln(1 - \phi) + 3(q + q^2)\eta + \frac{9}{2}q\eta^2 + \beta V q^3 \Pi \tag{3}$$

where $\eta \equiv \phi/(1 - \phi)$, $\beta \equiv 1/kT$, and k is the Boltzmann constant, and V is the volume of one hard sphere given by $V \equiv 4\pi R^3/3$.

Using the square-well potential, we can apply first-order perturbation theory with respect to the hard-sphere pressure, Π_0, and obtain the following expression for Π:

$$\Pi = \Pi_0 - 12\varepsilon\rho\phi \frac{\partial \left(\phi \int_1^\lambda g_0(r)r^2 dr\right)}{\partial \phi} \tag{4}$$

where the ρ is the number density of the particles.

If we set $W = \ln\alpha$ as in the case of a hard-sphere suspension, we obtain from eq 2:

$$\frac{\ln(\alpha/\alpha_0)}{\hat{\varepsilon}} = 12\phi^2 q^3 \frac{\partial \left(\phi \int_1^\lambda g_0(r)r^2 dr\right)}{\partial \phi} \tag{5}$$

Details on the derivation of eq 5 are given in Appendix E. Theoretical curves obtained using eq 5 are also shown in Fig 2. We observe that the theoretical values display a trend similar to our simulation results. However discrepancy between the theoretical curves and the simulation results increases with λ. A careful examination of scaled particle theory leads to the conclusion that this discrepancy arises from the basic assumption: $W = \ln\alpha$. When $\varepsilon = 0$, i.e. the hard-sphere case, the free energy change of inserting the AO sphere is equal to the work for creating a cavity. However, when $\varepsilon \neq 0$, these two quantities are not necessary equal. According to scaled particle theory, both hard-sphere interactions and attractive interactions contribute to the surface tension of the cavity. In other words, the cavity (i.e., the AO-sphere) screens the attractive interactions between the particles. On the other hand, in our simulation, the insertion of the AO-sphere does not screen the attractive interactions between the particles. Therefore, when λ is longer than the diameter of cavity, the free energy change of inserting second solute is smaller than the work of creating the cavity (Fig 5).
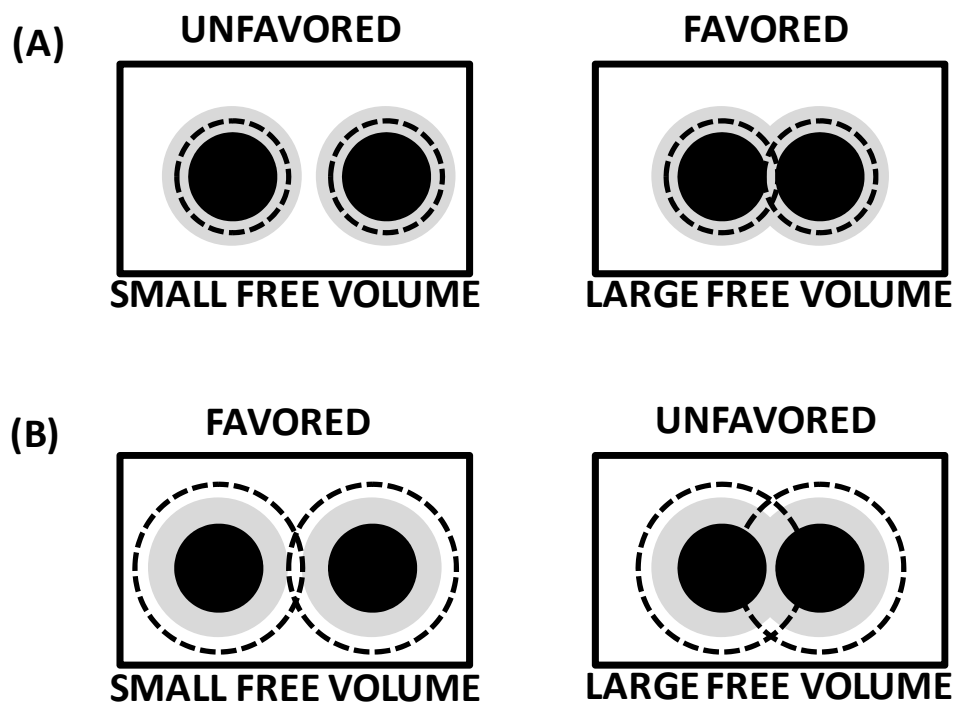
Figure 4. Free volume in the case of short (A) and long (B) range of interactions. The dashed circles denote the range of interaction. The gray layers denote the depletion layers with thickness proportional to q. The white area inside the box represents the free volume.
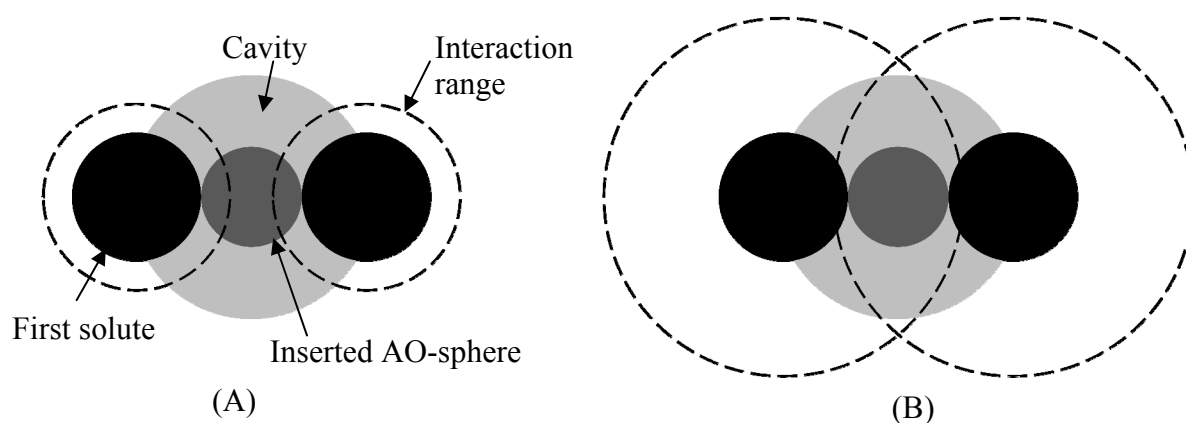
Figure 5. Insertion of the AO sphere. (A)When the range of interaction is short, the surface tension of the cavity is affected by both hard sphere interaction and the attractive interaction. (B) When the range of interaction is long, the attractive interaction does not contribute to the surface tension of the cavity.

Chapter 11

LLPS Isothermally Induced by

Protein Cross-linking

1. Introduction

In chapter 8, we have demonstrated that LLPS of protein solution can be induced by adding PEG, leading to the formation of a protein-rich phase. We now remark that protein condensation can be also significantly enhanced in the presence of protein oligomerization. Protein oligomerization can naturally occur or can be induced using chemical cross-linking. In this chapter, we show that condensation of BSA or lysozyme can be isothermally induced by the addition of a chemical cross-linker, glutaraldehyde. The morphology of protein condensed phases can be controlled by rationally choosing the experimental conditions. We will see that the coupling between LLPS and oligomerization plays a crucial role in the observed morphology. This study is relevant to both protein condensation diseases and the preparation of protein materials, because crowding, oligomerization and cross-linking are important aspects of these areas of research.

2. Isothermal Induced LLPS of BSA

We performed protein cross-linking reactions in aqueous solutions of BSA using glutaraldehyde as the cross-linking agent. Although the mechanism of protein-glutaraldehyde binding is rather complex due to the participation of glutaraldehyde oligomers, it is well-established that this bifunctional cross-linker mainly reacts with the ε-amino group of the lysines located on the protein surface (chapter 1, section 8). Cross-

linking experiments were performed as a function of glutaraldehyde concentration at a protein concentration of 10 mg/mL and at pH 5.2, close to the isoelectric point.

At this pH, LLPS of BSA-buffer solutions does not occur for protein concentrations as high as 400 mg/mL and temperatures as low as -15 °C. However, if PEG is added to the albumin-buffer solutions, LLPS can be readily induced by lowering the sample temperature as shown in chapter 8.

We performed BSA cross-linking in the presence and absence of PEG at 25 °C. The results are summarized in Fig 1a-c for the cross-linking experiments in the presence of PEG (PEG8000 6.0% (w/w)) and in Fig 1d for those in the absence of PEG. Fig 1a-c illustrates the presence of protein-rich droplets typical of LLPS at three cross-linker concentrations. The LLPS boundary of the corresponding cross-linker-free system is located at T_{ph}^{0}= -12 °C. As we will discuss later in more details, protein oligomerization is responsible for increasing the LLPS temperature, thereby bringing the system into a nonequilibrium state at the reaction temperature (25 °C). Contrary to ordinary LLPS, this process is irreversible due to chemical cross-linking. Hence, the final product is cross-linked protein microspheres. For the experiments in the absence of PEG, protein aggregates with no defined morphology were observed by both light microscopy and scanning electron microscopy. These amorphous aggregates are shown in Fig 1d and Fig 2. Thus, our results demonstrate that the presence of the LLPS boundary in the phase diagram of protein solutions qualitatively affects the morphology of the protein condensed phase generated from cross-linking.
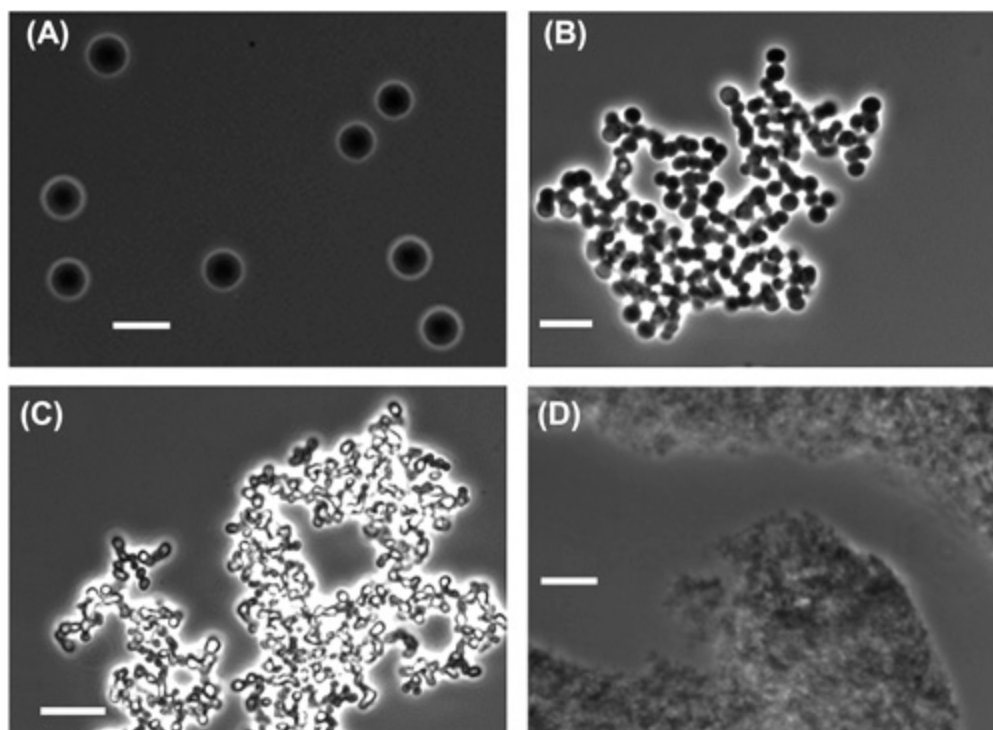
Figure 1. Images taken with a light microscope using phase contrast. LLPS induced by BSA cross-linking at 25 °C and at three representative glutaraldehyde concentrations: (A) c_{CL} = 0.075%; (B) c_{CL} = 0.15%; and (C) c_{CL} = 0.30%. The cross-linked microspheres were obtained from 10 mg/mL BSA in aqueous sodium acetate buffer, 0.1 M, pH 5.2, PEG8000 6.0% (w/w) at 25 °C. The LLPS temperature of the cross-linker-free system is T_{ph}^{0} = -12 °C. (D) Cross-linked amorphous aggregates obtained from 10 mg/mL albumin in aqueous sodium acetate buffer, 0.1 M, pH 5.2, c_{CL} = 0.1% at 25 °C. The length of the horizontal bars is 10 μm.
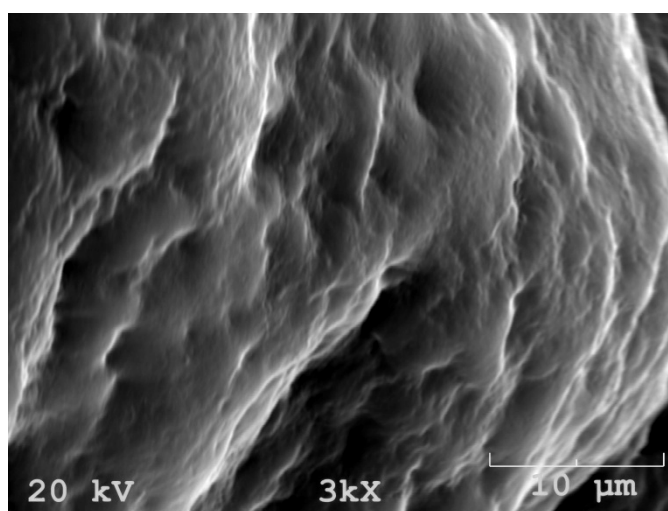
Figure 2. Images taken with SEM. Cross-linked amorphous aggregates obtained from 10 mg/mL albumin in aqueous sodium acetate buffer, 0.1 M, pH 5.2, c_{CL} = 0.1% at 25 °C. The length of the horizontal bars is 10 μm.

From Fig 1a-c, we can also observe that the average radius of the final cross-linked microspheres decreases as the glutaraldehyde concentration increases. To explicitly show this behavior, we report the average radius, R, of the microspheres as a function of cross-linker concentration in Fig 3a. Correspondingly, in Fig 3b, we show that R^{-1} linearly decreases with c_{CL}, approaching zero as the cross-linker concentration vanishes. The observed behavior suggests that the final size of the microspheres is governed by the nucleation of the protein-rich droplets from the metastable solution. According to nucleation theory,[214] the radius of the critical nucleus decreases as the supersaturation increases. Thus, if a higher cross-linker concentration brings the system into a higher final supersaturation with respect to LLPS, the average radius of the nucleating particles decreases as the cross-linker concentration increases.
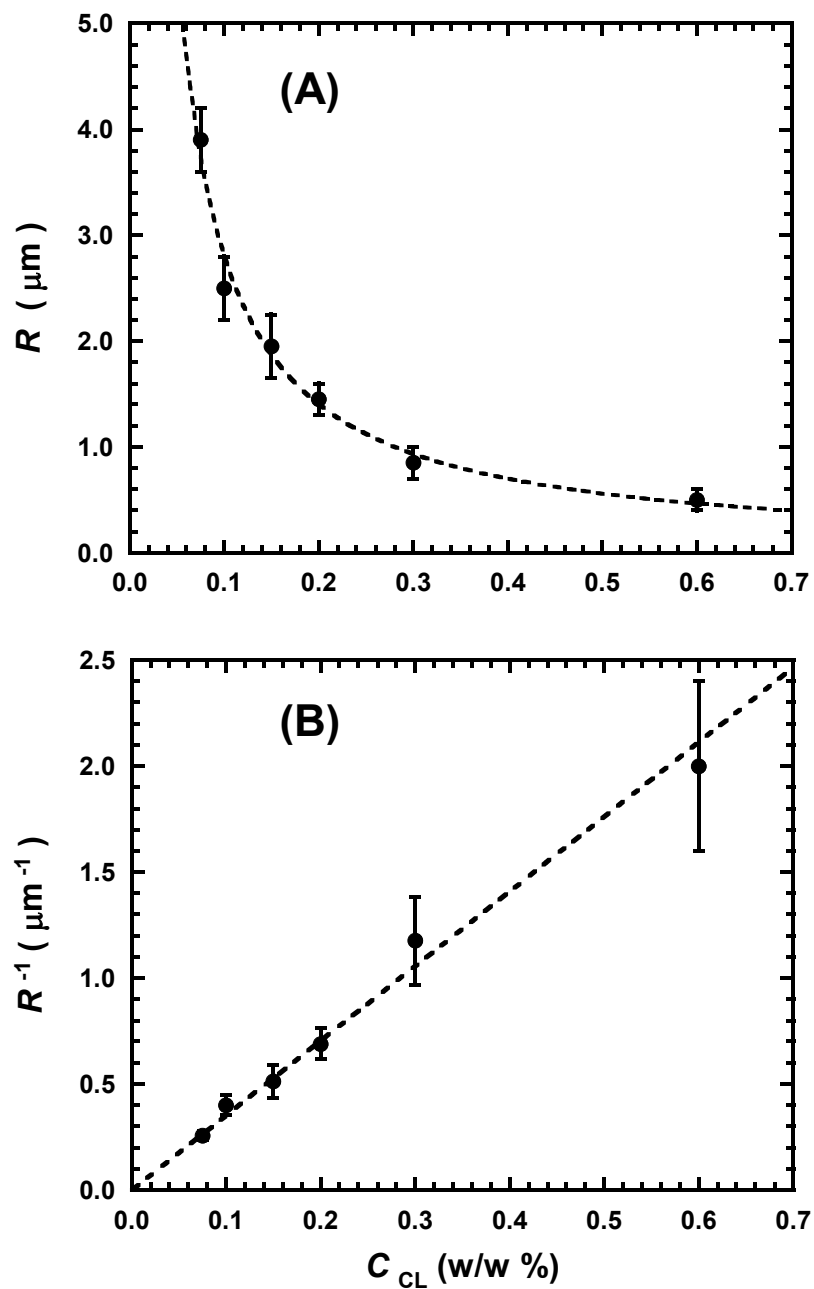
Figure 3. (A) Average radius of BSA microspheres, R, as a function of glutaraldehyde concentration, c_{CL}. The experimental conditions are those described for Figure 1. (B) Inverse of R as a function of c_{CL}. The R^{-1} data were fitted to straight lines.

3.   Isothermal Induced LLPS of lysozyme

We also performed protein cross-linking reactions in aqueous solutions of chicken egg-white lysozyme at 25 °C using glutaraldehyde as the cross-linking agent. Cross-linking experiments were performed as a function of glutaraldehyde concentration at a protein concentration of 10 mg/mL.

Lysozyme is known to undergo LLPS without the assistance of PEG. LLPS in lysozyme aqueous solutions has been previously reported at pH 4.5 and 7.1. Consequently, we performed our cross-linking experiments at these experimental conditions. We observed that lysozyme has a net positive charge at these two pH values (pI ≈11).[215] This implies that salt must be added to the protein solutions to screen electrostatic repulsive interactions between the protein molecules. For all our lysozyme experiments, we experimentally measured the LLPS temperature, T_{ph}^{0}, of the corresponding cross-linker-free systems.

Our results are summarized in Fig 4a,b. Fig 4 illustrates the presence of protein-rich droplets for lysozyme at pH 4.5 and 7.1. We found that the phase-separation process of albumin solutions at pH 5.2 is significantly faster than that of lysozyme solutions at pH 4.5-7.1. This can be related to the significantly lower number of lysines of lysozyme (six) as compared to BSA (sixty). Moreover, contrary to the BSA case, the size of the cross-linked protein microspheres was found to be independent of the cross-linker concentration. This suggests that the final size of the microspheres does not decrease as the rate of nucleation increases. In this case, the growth rate of the droplets may also play a significant role.
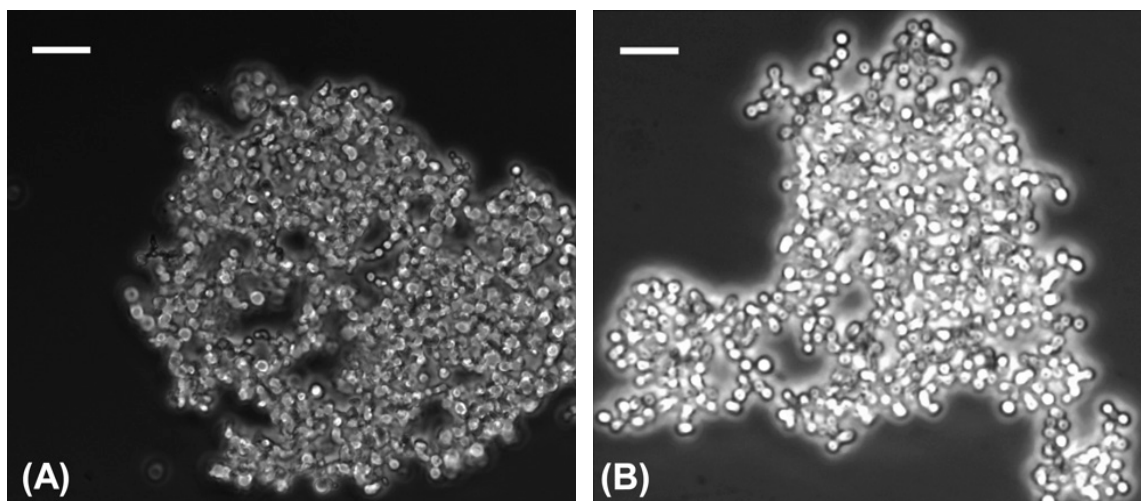
Figure 4. Images taken with a light microscope using phase contrast. LLPS induced by lysozyme cross-linking at 25 °C. Cross-linked lysozyme microspheres are reported for two representative cases: (A) 10 mg/mL lysozyme in aqueous sodium acetate buffer, 0.1 M, pH 4.5, NaCl 0.5 M, $c_{CL} = 0.1\%$, $T_{ph}^{0} = -12$ °C and (B) 10 mg/mL lysozyme in sodium phosphate buffer, 0.1 M, pH 7.1, NaCl 1.0 M, $c_{CL} = 0.1\%$, $T_{ph}^{0} = -1.7$ °C. The length of the horizontal bars is 10 μm.

Because the phase-separation process of the lysozyme solutions is relatively slow, we accurately examined the effect of pH and NaCl concentration on the kinetics of LLPS separation by determining the turbidity of the protein samples as a function of time from the beginning of the reaction. As shown in Fig 5, the turbidity is observed to sharply increase after a well-defined and reproducible induction time, t_{ind}, which can be used to characterize the rate of phase separation. In Fig 6, we show that t_{ind}^{-1} increases with the cross-linker concentration, c_{CL}, as expected. At a given pH and cross-linker concentration, t_{ind}^{-1} increases with salt concentration. This effect, which is larger at pH 4.5 than at pH 7.1, can be related to the corresponding effect of the salt concentration on T_{ph}^{0} because a higher value of T_{ph}^{0} implies a smaller difference between the initial LLPS temperature

and the cross-linking temperature. Moreover, an increase in T_{ph}^{0} is also related to a corresponding increase in the magnitude of protein-protein attractive interactions. This favors cross-linking by promoting more contacts between the protein molecules. Finally, Fig 6 also shows that t_{ind}^{-1} significantly increases with the pH at a given cross-linker concentration. This effect is mainly related to the corresponding increase in chemical reactivity of the lysine ε-amino groups of the protein. Indeed, the t_{ind}^{-1} values at pH 7.1 are significantly higher than those at pH 4.5 even when T_{ph}^{0} at pH 4.5 is higher than at pH 7.1.
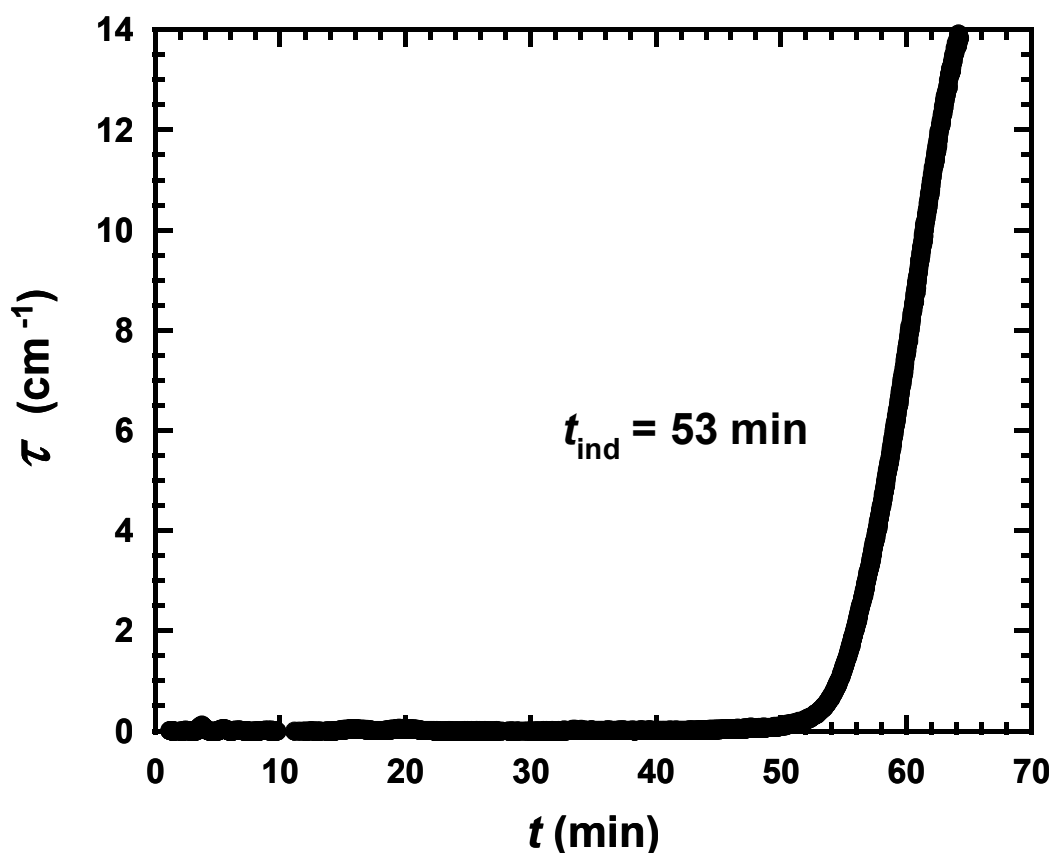
Figure 5. Turbidity, τ, as a function of time, t, after cross-linking has started. This turbidity profile was obtained at 25 °C for the representative case: 10 mg/mL lysozyme in aqueous sodium acetate buffer, 0.1 M, pH 4.5, NaCl 0.5 M, $c_{CL} = 0.5\%$.
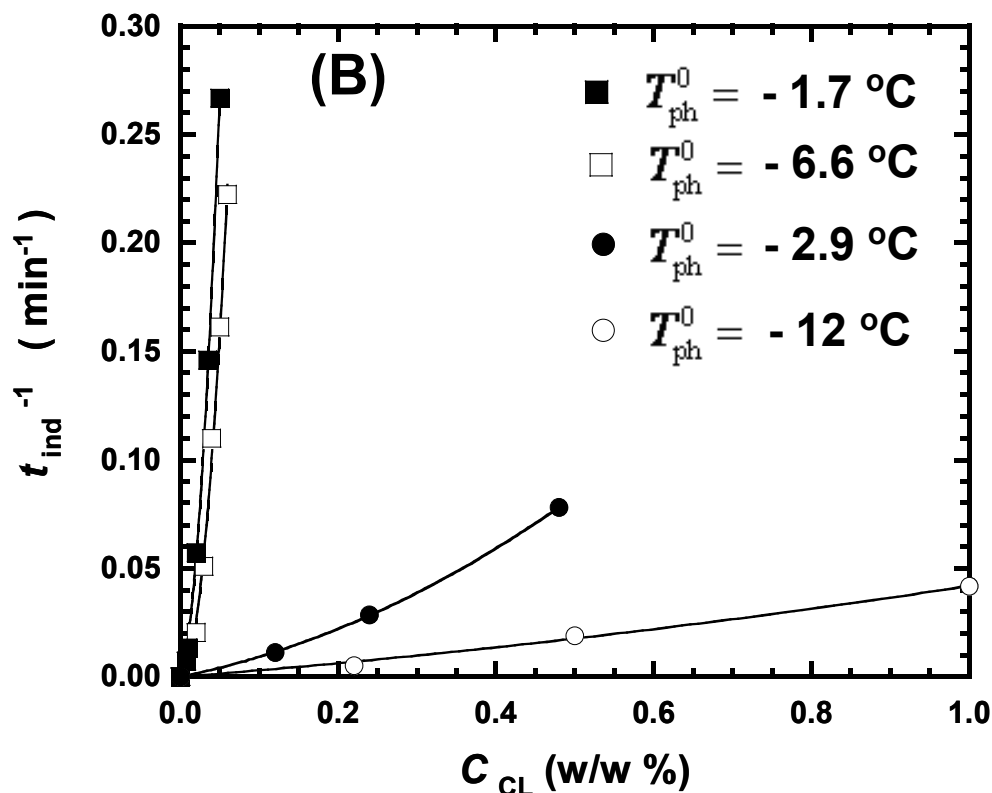
Figure 6. Inverse of the induction time, t_{ind}, as a function of c_{CL} for four representative cases at 25 °C: 10 mg/mL lysozyme in aqueous sodium acetate buffer, 0.1 M, pH 4.5, NaCl 0.5 M, T_{ph}^0= -12 °C (open circles); 10 mg/mL lysozyme in aqueous sodium acetate buffer, 0.1 M, pH 4.5, NaCl 1.0 M, T_{ph}^0= -2.9 °C (closed circles); 10 mg/mL lysozyme in aqueous sodium phosphate buffer, 0.1 M, pH 7.1, NaCl 0.5 M, T_{ph}^0= -6.6 °C (open squares); and 10 mg/mL lysozyme aqueous sodium phosphate buffer, 0.1 M, pH 7.1, NaCl 1.0 M, T_{ph}^0= -1.7 °C (closed squares). The solid curves are guides for the eye.

To further characterize the effect of pH on phase separation, we also performed cross-linking of the lysozyme at pH 9.1 closer to the isoelectric point. At this pH, measurements of static and dynamic light scattering show that the lysozyme undergoes aggregation even in the absence of cross-linkers. Interestingly, the aggregation rate decreases as the NaCl concentration increases. This behavior can be related to the weakening effect of salts on electrostatic attractive interactions between the oppositely

charged surface groups of the protein molecules. However, we remark that the LLPS temperature increases with the NaCl concentration. This implies that salt favors LLPS with respect to aggregation. We also investigated protein aggregation in the presence of PEG. We found that both aggregation rate and LLPS temperature increase with the PEG concentration. This behavior is consistent with the presence of depletion forces, which enhance protein-protein attraction, thereby favoring both aggregation and LLPS.

The effect of cross-linking on the phase separation of lysozyme solutions is consistent with the observed behavior. In the presence of NaCl at 0.5 M and higher concentrations, we observed the formation of droplets (Fig 7a). On the other hand, at low salt concentrations or in the presence of PEG, we observed the formation of amorphous macroscopic aggregation (Fig 7b). For the latter case, the corresponding sample turbidity steadily increases with time, contrary to the typical turbidity profile associated with oligomerization-induced LLPS. Hence, the presence of the LLPS boundary in the protein-solution phase diagram may not necessarily lead to the formation of protein-rich droplets if protein aggregation is relatively fast. From these observations, we deduced that specific conditions of pH, PEG, and salt concentrations can be chosen so that LLPS is favored with respect to aggregation.

In the following two sections, we will examine two important aspects of oligomerization-induced LLPS in more detail: (1) the effect of oligomerization on the LLPS temperature and (2) the kinetic evolution of oligomerization-induced LLPS. We will focus on the following two systems: albumin-PEG8000-buffer at pH 5.2 and lysozyme-NaCl-buffer at pH 4.5. For the lysozyme system, we successfully used DLS to

monitor the onset of phase separation since the corresponding cross-linking rate was sufficiently slow.
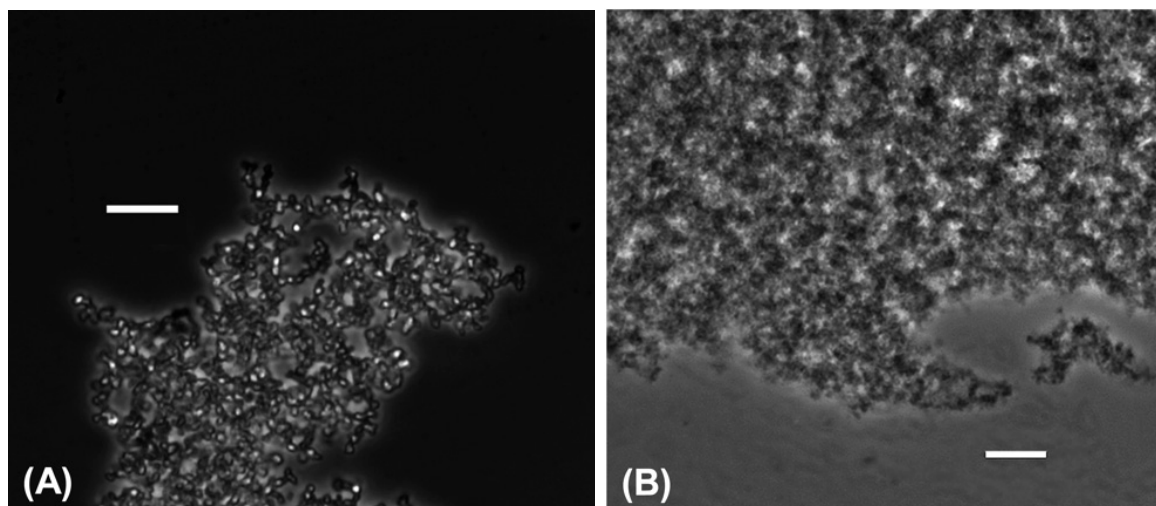
Figure 7. Images taken with a light microscope using phase contrast. (A) LLPS induced by lysozyme cross-linking at 25 °C. The experimental conditions are 10 mg/mL lysozyme in sodium borate buffer, 0.2 M, pH 9.0, NaCl 0.5 M, $c_{CL} = 0.1\%$, $T_{ph}^0 = -4.3$ °C. (B) Amorphous aggregation induced by lysozyme cross-linking at 25 °C. The experimental conditions are 10 mg/mL lysozyme in sodium borate buffer, 0.2 M, pH 9.0, PEG8000 2.5%, $c_{CL} = 0.1\%$, $T_{ph}^0 = -4.3$ °C. The length of the horizontal bars is 10 μm.

4. Effect of Oligomerization on LLPS Temperature

As the degree of oligomerization increases, the LLPS boundary moves toward higher temperatures. To experimentally demonstrate this effect, we performed measurements of the LLPS temperature, T_{ph}, as a function of time, t, starting from the initial LLPS temperature, T_{ph}^{0}, of the protein monomer. In Fig 8a, we experimentally demonstrate the increase of T_{ph} with time for both protein cases. As expected, the rate of increase of the LLPS temperature found for albumin was significantly larger than that found for lysozyme.

To further examine the effect of oligomerization on the LLPS temperature, we measured the protein contribution to light-scattering intensity, i_s, as a function of time starting from the initial intensity, i_s^{0}. In Fig 8b, we report the absolute temperature ratio: T_{ph}/ T_{ph}^{0} as a function of the corresponding intensity ratio i_s/i_s^{0}. Since the light-scattering intensity is directly proportional to the mass-average molecular weight of the protein oligomers, the ratio i_s/i_s^{0} is equal to the mass-average degree of oligomerization. We can see that the T_{ph}/ T_{ph}^{0} curves reported as a function of i_s/is^{0} in Fig 8b are close to each other, even though the T_{ph} curves in Fig 8a are significantly separated. Thus, the effect of oligomerization on the LLPS temperature is similar for both proteins.
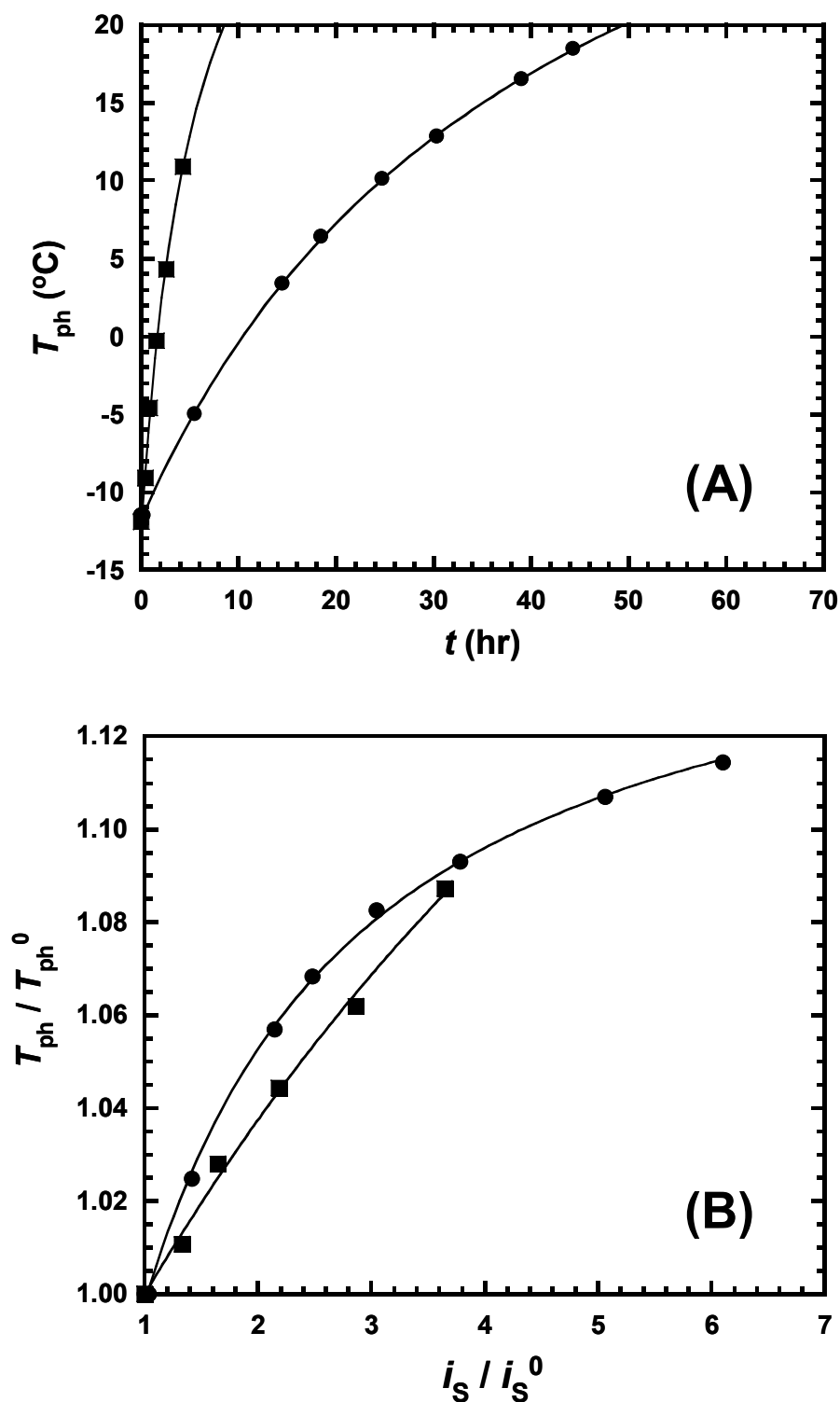
Figure 8. (A) LLPS temperature, T_{ph}, as a function of time, t, during cross-linking at 25 °C. These data were taken for 10 mg/mL albumin in sodium acetate buffer, 0.1 M, pH 5.2, PEG8000 6.0%, $c_{CL} = 0.015\%$ (squares), and 10 mg/mL lysozyme in sodium acetate buffer, 0.1 M, pH 4.5, NaCl 0.5 M, $c_{CL} = 0.050\%$ (circles). (B) Corresponding ratio of absolute temperatures, T_{ph}/T_{ph}^{0}, as a function the ratio of light-scattering intensities: i_s/i_s^{0}. The solid curves are guides for the eye.

The increase of T_{ph} with the degree of polymerization has been theoretically investigated using a perturbation method developed by Banaszak (Appendix F).[132] Since models based on hard spheres have been used to describe the phase behavior of protein solutions,[60] those based on hard-sphere chains[126-132] are used here to examine the effect of protein oligomerization. For monodisperse hard-sphere chains, Banaszak et al. have reported the increase of critical temperature as a function of chain length.[132] We used their model to compute T_{ph}/ T_{ph}^0 as a function of protein volume fraction, ϕ, and chain length. Here, we will focus on the monodisperse dimer. This species represents the smallest hard-sphere chain.

For the dimer, we compute T_{ph}/ T_{ph}^0 of 1.4 at our experimental protein concentration ($\phi = 0.007$) and T_{ph}/ T_{ph}^0 of 1.3 around the critical point. We can therefore conclude that all our experimental values of T_{ph}/ T_{ph}^0, reported in Fig 8b are significantly lower than these theoretical predictions. The same conclusion is obtained if the Baxter potential (sticky spheres)[216] is employed instead of the square-well potential. A more direct, yet approximate, comparison between our experimental results and dimer prediction can be made if we consider the experimental T_{ph}/ T_{ph}^0 values corresponding to a number-average degree of oligomerization equal to 2. If we assume the most probable distribution,[212] we must consider the values of T_{ph}/ T_{ph}^0 at i_s/ i_s^0 of 4 for the comparison. We obtain T_{ph}/ T_{ph}^0 of 1.1 for both protein cases. This value corresponds to an increase in temperature that is only about 25% of that predicted by the model. It is however important to bear in mind that the accuracy of this comparison is affected by the polydisperse nature of the cross-linked protein oligomers. Fortunately, an experimental comparison of LLPS phase boundaries between monodisperse dimer and monomer has been reported using γ D-

crystallin as a model protein.[134] As in our case, their experimental T_{ph}/T_{ph}^0 values are also significantly lower than the theoretical prediction: T_{ph}/T_{ph}^0 is 1.05 at ϕ of 0.007 and 1.1 at ϕ of 0.2, close to the critical point. Thus, although the model used previously qualitatively describes the increase of LLPS temperature with the degree of oligomerization, it significantly overestimates the experimental values of T_{ph}/T_{ph}^0.

5. Kinetic Evolution of Oligomerization-Induced LLPS

For both lysozyme and BSA solutions, we used dynamic light scattering (DLS) to examine their macromolecular size distribution as a function of time after the addition of glutaraldehyde. The representative distribution histogram is shown in Fig 9. Fig 10 summarizes our DLS results on lysozyme at three representative glutaraldehyde concentrations. During the initial stage of cross-linking, the macromolecular distribution is monomodal. The corresponding average hydrodynamic radius, R_h, was calculated and reported as a function of time, t, starting from the initial hydrodynamic radius, R_h^0, of the protein monomer (circles in Fig 10). The value of R_h increases with t due to protein oligomerization. After a given induction time, there was a sharp change in total scattered intensity, and the distribution became bimodal. The second peak corresponds to protein clusters with an average hydrodynamic radius, $R_h^{(c)}$, larger than 100 nm (squares in Fig 10). The size of these mesoscopic scattering elements increases with time. From Fig 10, we can see that the protein clusters are detected at approximately the same oligomer average radius and, consequently, at the same degree of oligomerization, independent of the cross-linker concentration.
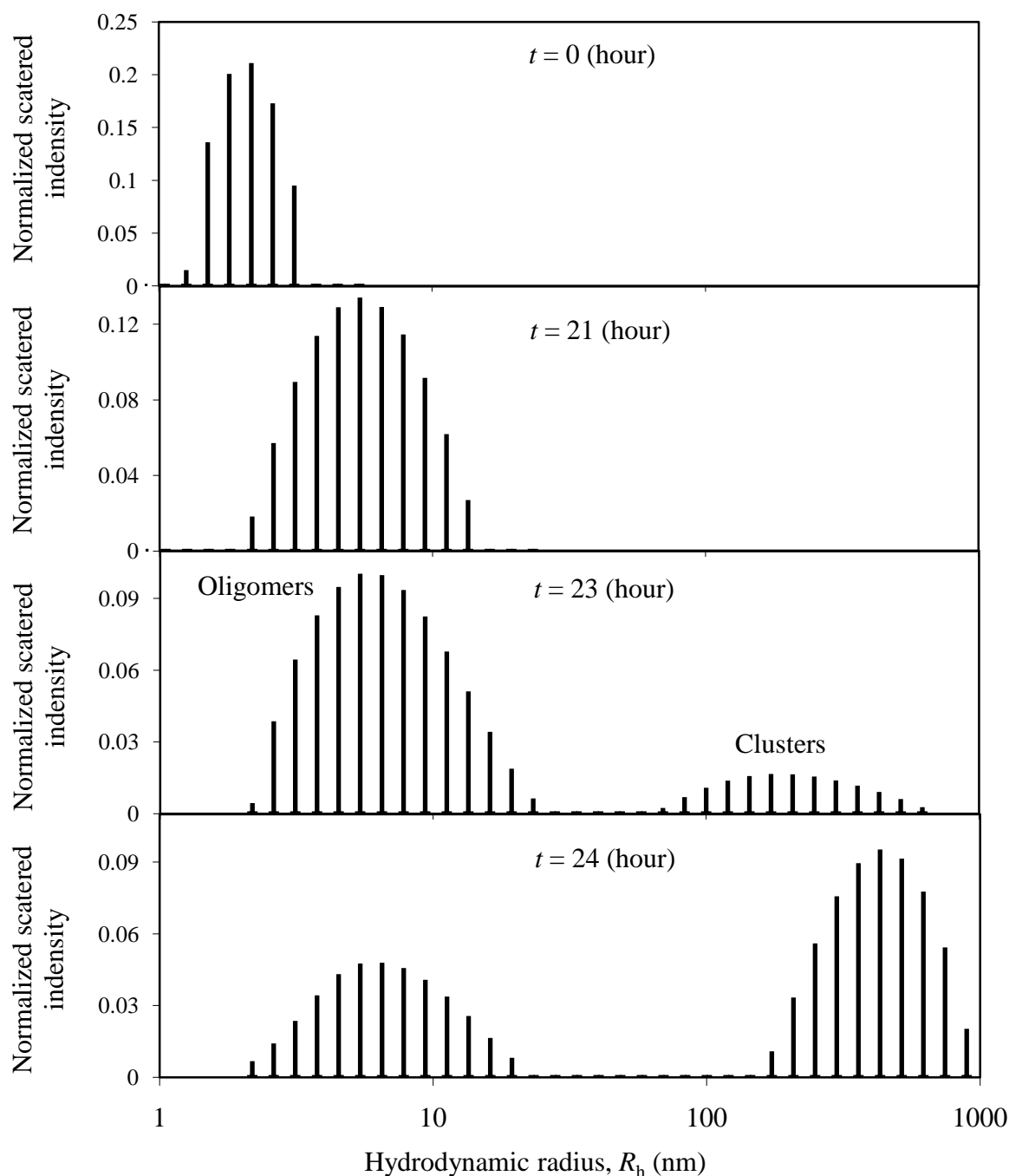
Figure 9. Histogram of size distribution of lysozyme during cross-linking at 25 °C and at four representative time. These data were taken for 10 mg/mL lysozyme in sodium acetate buffer, 0.1 M, pH 4.5, 0.5 M NaCl with 0.07% glutaraldehyde. The phase separation temperature of monomer obtained under this condition by turbidity measurement is -9.8 °C.
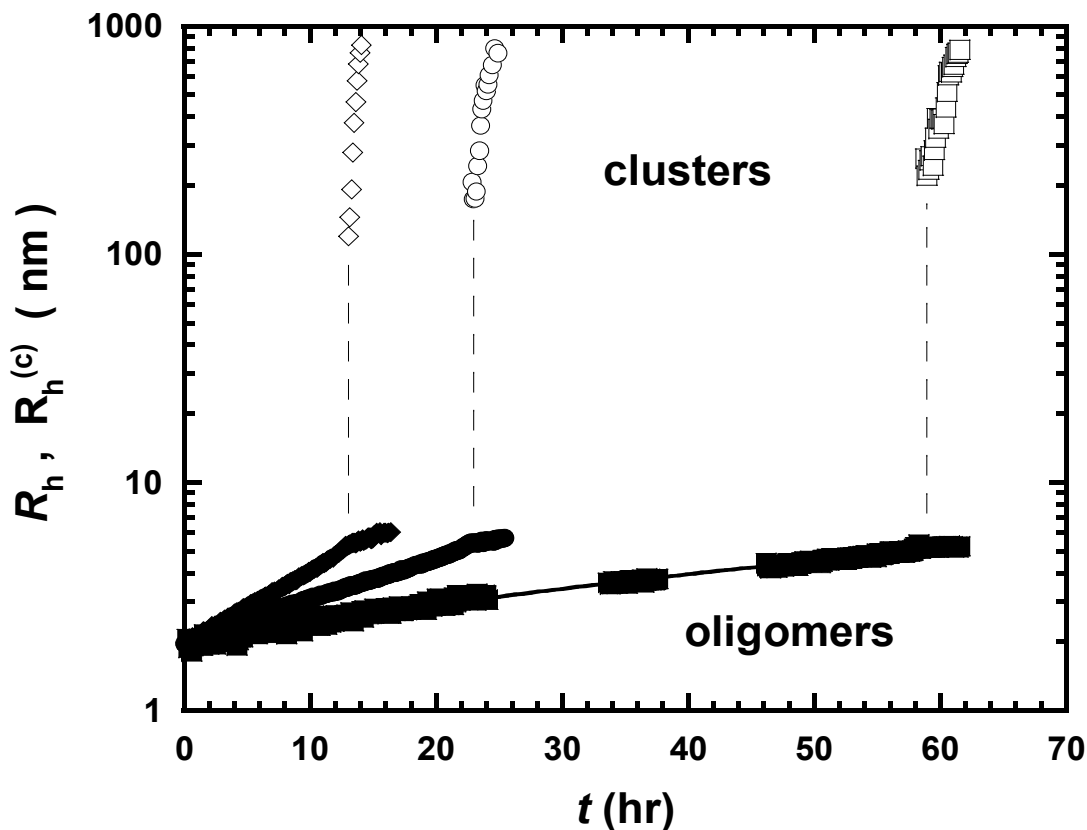
Figure 10. Average hydrodynamic radius of oligomers (R_h, solid data points) and mesoscopic clusters ($R_h^{(c)}$, open data points) as a function of time, t, during cross-linking at 25 °C and at three representative glutaraldehyde concentrations: 0.045% (squares), 0.070% (circles), and 0.090% (diamonds). The dashed vertical lines identify the light-scattering induction time for the formation of protein clusters. These data were taken for 10 mg/mL lysozyme in sodium acetate buffer, 0.1 M, pH 4.5, 0.5 M NaCl. The solid curves are guides for the eye.

In the case of BSA, we obtained similar results. However, the formation of protein clusters could not be quantitatively monitored by DLS due to the generally fast kinetics of phase separation. Moreover, for those cases in which the rate of phase separation was sufficiently reduced by using very low glutaraldehyde concentrations, the hydrodynamic radius of the incipient protein clusters was found to be already significantly larger than 1 μm, which is outside the DLS domain.

Using our DLS results, we examined the effect of cross-linker concentration on both the rate of protein oligomerization and the size of the protein clusters. To examine the protein-oligomerization rate, we determined the initial slope: $(dR_h/dt)_{t=0}/R_h^0$ as a function of glutaraldehyde concentration for both protein cases. This quantity is related to the initial oligomerization rate $(-dc_1/dt)_{t=0}$, where c_1 is the mass concentration of the monomer. Since R_h is inversely proportional to the z-average diffusion coefficient:

$$\frac{R_h}{R_h^0} = \frac{\sum_{i=1}^{\infty} i c_i}{\sum_{i=1}^{\infty} i \alpha_i c_i} \tag{1}$$

where c_i is the mass concentration of protein species i, and α_i is the ratio of the initial hydrodynamic radius to the hydrodynamic radius of species i. We have α_1 of 1 for the monomer by definition and $\alpha_i < 1$ for the protein oligomers (with $i \neq 1$). By differentiating equ 1 with respect to time and taking the limit of $t \to 0$, we obtain:

$$\frac{1}{R_h^0}\left(\frac{dR_h}{dt}\right)_{t=0} = 2(\alpha_2 - 1)\frac{1}{c_1^0}\left(\frac{dc_1}{dt}\right)_{t=0} \tag{2}$$

where $\alpha_2 < 1$ is the ratio of R_h^0 to the hydrodynamic radius of the dimer, and c_1^0 is the monomer mass concentration at $t = 0$ (i.e., the total mass concentration). We therefore conclude that $(dR_h/dt)_{t=0}/R_h^0$ is directly proportional to $(-dc_1/dt)_{t=0}$.

In Fig 11a,b, we report $(dR_h/dt)_{t=0}/R_h^0$ as a function of the glutaraldehyde concentration, c_{CL}, for both BSA (Fig 11a) and lysozyme (Fig 11b). As expected, we observed that the oligomerization rate for BSA is significantly higher than that for lysozyme at all experimental glutaraldehyde concentrations. The increase in reaction rate with glutaraldehyde concentration shows a significant deviation from linearity in both cases, and the reaction order with respect to glutaraldehyde was found to be 1.90 ± 0.05 for albumin and 1.6 ± 0.1 for lysozyme. We further observe that, for glutaraldehyde

concentrations higher than $c_{CL}{}^0 \approx 0.025\%$ (w/w), the oligomerization rate can be regarded as directly proportional to $(c_{CL}-c_{CL}{}^0)$ for both protein cases (Fig 11a,b). Since the concentration of glutaraldehyde oligomers increases with c_{CL}, our results are consistent with a reaction mechanism in which the presence of glutaraldehyde oligomers is necessary for protein cross-linking.

We now examine the dependence of the hydrodynamic radius, $R_h{}^{(c)}$, of the incipient protein clusters on cross-linker concentration. Since the values of $R_h{}^{(c)}$ are closely related to the corresponding critical radii for the nucleation of the protein-rich droplets, they can be used to examine the effect of cross-linker concentration on nucleation. In Fig 11a, we report $R_h{}^{(c)}$ at the light-scattering induction time (vertical dashed lines in Fig 10) as a function of glutaraldehyde concentration, c_{CL}, for lysozyme. We can see that $R_h{}^{(c)}$ decreases as c_{CL} increases. Correspondingly, $(R_h{}^{(c)})^{-1}$ linearly increases with c_{CL} in Fig 12b. Our results imply that the final supersaturation with respect to nucleation increases with cross-linker concentration. This behavior, which is analogous to that shown for the radius of BSA microspheres in Fig 3a,b, corroborates the hypothesis that the final size of albumin microspheres is controlled by nucleation, whereas that of lysozyme microspheres is significantly affected by the droplet growth rate.
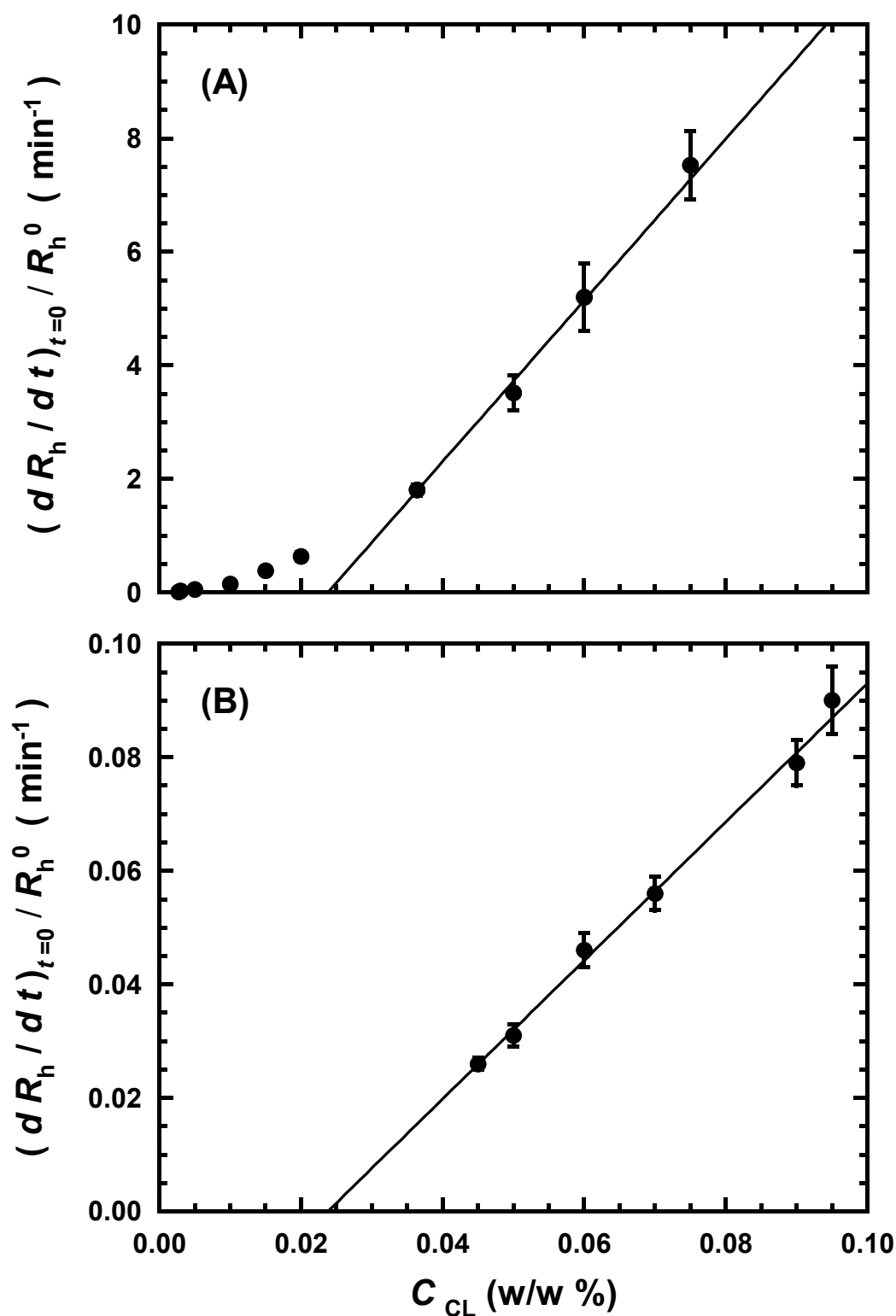
Figure 11. Derivative $(dR_h/dt)_{t=0}/R_h^0$ as a function of glutaraldehyde concentration, c_{CL} at 25 °C. (A) 10 mg/mL BSA in sodium acetate buffer, 0.1 M, pH 5.2, PEG8000 6.0% and (B) 10 mg/mL lysozyme in sodium acetate buffer, 0.1 M, pH 4.5, 0.5 M NaCl. The data with $c_{CL} > 0.025\%$ were fit to straight lines.
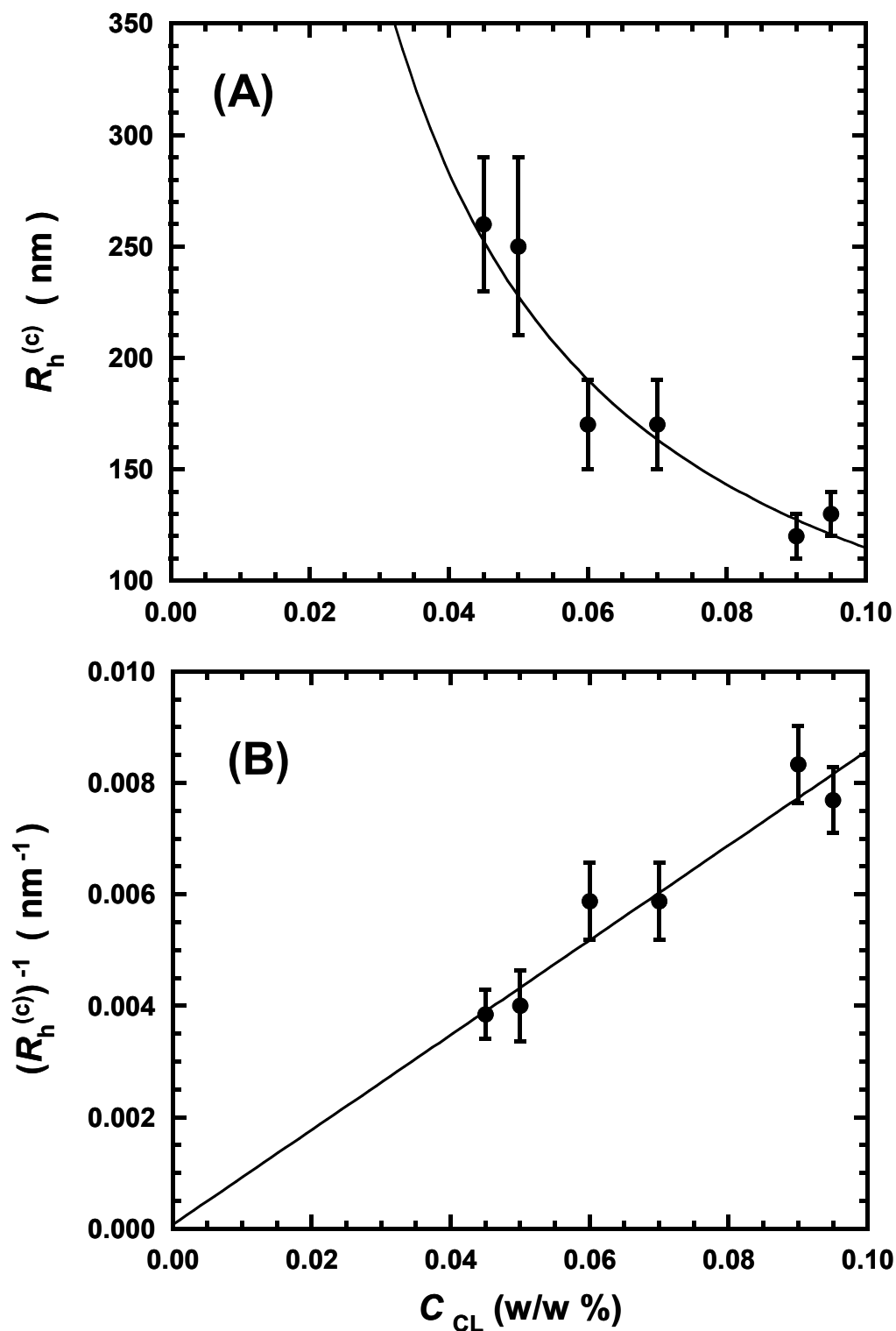
Figure 12. (A) Average hydrodynamic radius, $R_h^{(c)}$, as a function of glutaraldehyde concentration, c_{CL}. The solid line is an inversely proportional fit. The experimental conditions are those described in Figure 7. (B) Inverse of $R_h^{(c)}$ as a function of c_{CL}. The $(R_h^{(c)})^{-1}$ data were fitted to straight lines.

Chapter 12

Laccase Condensations

1. Introduction

Laccase is a family of copper-containing oxidases which catalyze oxidation of phenol and aniline derivatives.[166-168] It is one of a few enzymes that are practically used in industry. Laccase is widely used in the petroleum, paper, and textile industries.[23, 167-170] However, during the industrial processing,[108] free laccase is usually vulnerable to thermal and chemical denaturation in water and in organic solvents. Recent studies have shown that protein immobilization can increase the stability of enzyme. A widely used approach for protein immobilization is to cross-link proteins in its condensed phase (such as crystal and aggregates) by adding chemical cross-linkers.[15,22-28] Roy et al.[217] have recently prepared cross-linked crystals of laccase using ammonium sulfate as precipitant and glutaraldehyde as cross-linker. Cabana et al.[218] have prepared cross-linked laccase aggregates using respectively PEG and ammonium sulfate as precipitants and glutaraldehyde as cross-linker. In both cases, increases of stability at elevated temperature and in organic solvent were observed for the cross-linked crystals and aggregates of laccase. Moreover, BSA has been used as protein feeder to prepare cross-linked enzyme coaggregates.[219] A further increase of enzyme stability has been observed in the BSA-laccase coaggregates. However, to the best of our knowledge, the phase behavior of laccase-precipitant-buffer system and the effect of types of precipitant on the morphology of the produced laccase aggregates is not understood. In this chapter, we provide a preliminary investigation on laccase condensation in the presence of PEG and $(NH_4)_2SO_4$. We will see how the nature of precipitant has a strong effect on the type of laccase condensation. We will also investigate BSA-laccase co-aggregation in the

presence of a crosslinker. Three new important aspects are introduced compared to previous investigations on enzyme systems: 1) thermal induction of enzyme reversible condensation; 2) enzyme condensation at the liquid-liquid interface; 3) the novel preparation of enzymatically-active cross-linked BSA-laccase coaggregates. The results reported in this chapter will provide the basis for further systematic studies on enzyme condensation processes and the formation of cross-linked enzyme aggregates.

2.   Protein-Protein Interactions in Laccase Aqueous Solutions

Because protein condensation is driven by the net attractive interaction between proteins, we first investigate the protein-protein interaction in laccase aqueous solution. One important factor affecting protein-protein interactions is the electrostatic interaction between proteins. Hence, protein-protein interactions for laccase are expected to strongly depend on pH and ionic strength. One way to probe protein-protein interactions is to characterize the dependence of protein diffusion coefficient, D, as a function of protein concentration, C. Within the low protein concentration range, the behavior of $D(C)$ is linear. The lower the slope, the stronger the attraction between protein molecules is. Moreover, that D increases with C represents a sufficient condition for the presence of repulsive protein-protein interactions. In these conditions, protein condensation is not expected to occur. We have investigated the effect of pH and ionic strength on $D(C)$ for laccase using dynamic light scattering (DLS).

In Fig 1, we show $D(C)$ data obtained at several pH values and the same, low ionic strength of 0.035 M. The samples at pH 3.8, 4.5 and 5.0 were prepared with sodium acetate buffer, while those at pH 5.9 and 7.1 were in sodium phosphate buffer. In this figure, we can see that D increases with C in all cases. However, the increase in D observed in phosphate buffer is significantly larger than that observed in acetate buffer. Thus, repulsive protein-protein interactions weaken as pH is lowered. The observed effect of pH on $D(C)$ is consistent with the fact that the isoelectric point of laccase is pI=3.5±0.5. Because the net charge on laccase is small close to the isoelectric point, the contribution of repulsive electrostatic interactions is small. As pH increases, laccase molecules are negatively charged and the repulsive electrostatic interaction becomes significant. In Fig 2, we show the effect of ionic strength on $D(C)$ by adding NaCl to the laccase samples. We investigate the effect of ionic strength at pH 7.1 (phosphate buffer) and pH 4.5 (acetate buffer). At pH 7.1, the slope of $D(C)$ decreases as the ionic strength increases to 1.0 M. A further increase of the ionic strength to 2.0 M does not appreciably change the slope of $D(C)$. This behavior is expected. Indeed, electrostatic repulsive interactions between charged macromolecules are screened in the presence of salt ions, the effect being larger at lower ionic strengths. We also note that the intercept of $D(C)$ decreases as salt concentration increases. This is related to a corresponding increase in solution viscosity due to NaCl. However, our results at pH 4.5 surprisingly reveal the opposite behavior: the slope of $D(C)$ at the ionic strength of 1.0 M was found to be higher than that observed in the absence of NaCl. We explain this unusual behavior by noting that the laccase net charge in these conditions is small. Thus, the effect of salt on electrostatic screening of the net charge is small. In these conditions, attractive

electrostatic interactions between oppositely charged amino acids on the surface of laccase proteins may be significant. Here, an increase of ionic strength may weaken this electrostatic attraction leading to an increase of the $D(C)$ slope. These results at pH 4.5 suggest that strong anisotropic attractive interactions between laccase molecules occur. These interactions can favor protein aggregation compared to other forms of protein condensation such as LLPS.
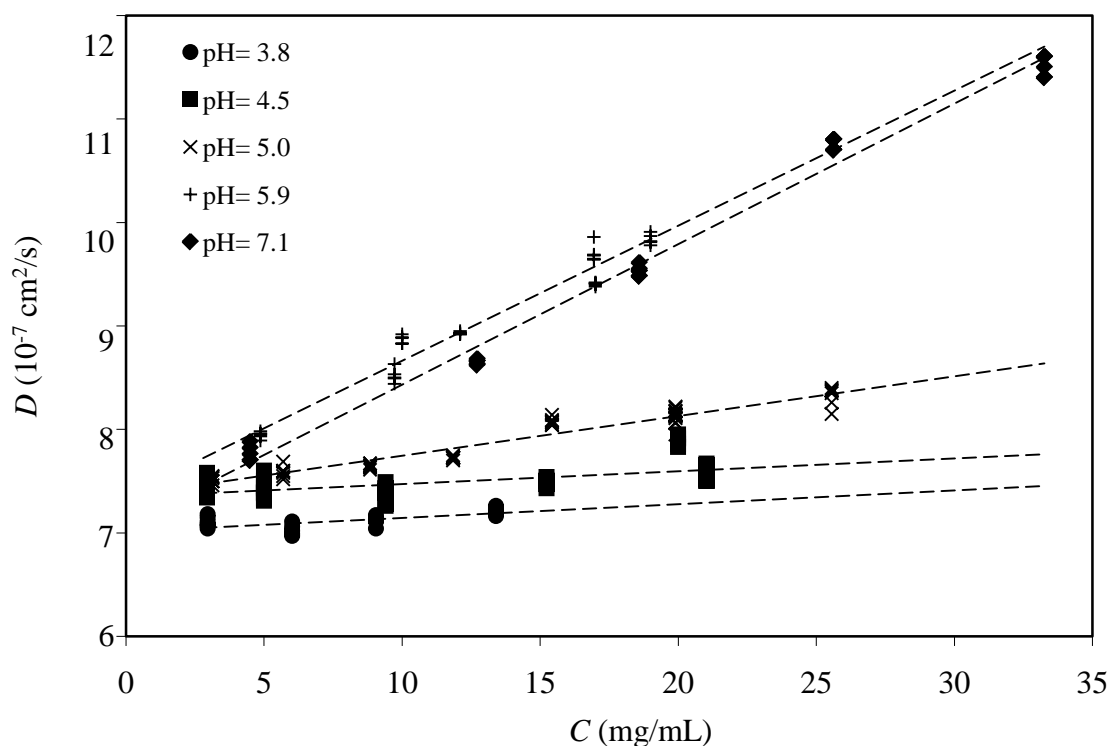
Figure 1. Diffusion coefficient, D, of laccase as a function of laccase concentration, C, at several pH values. The dashed lines are linear fits to the data.
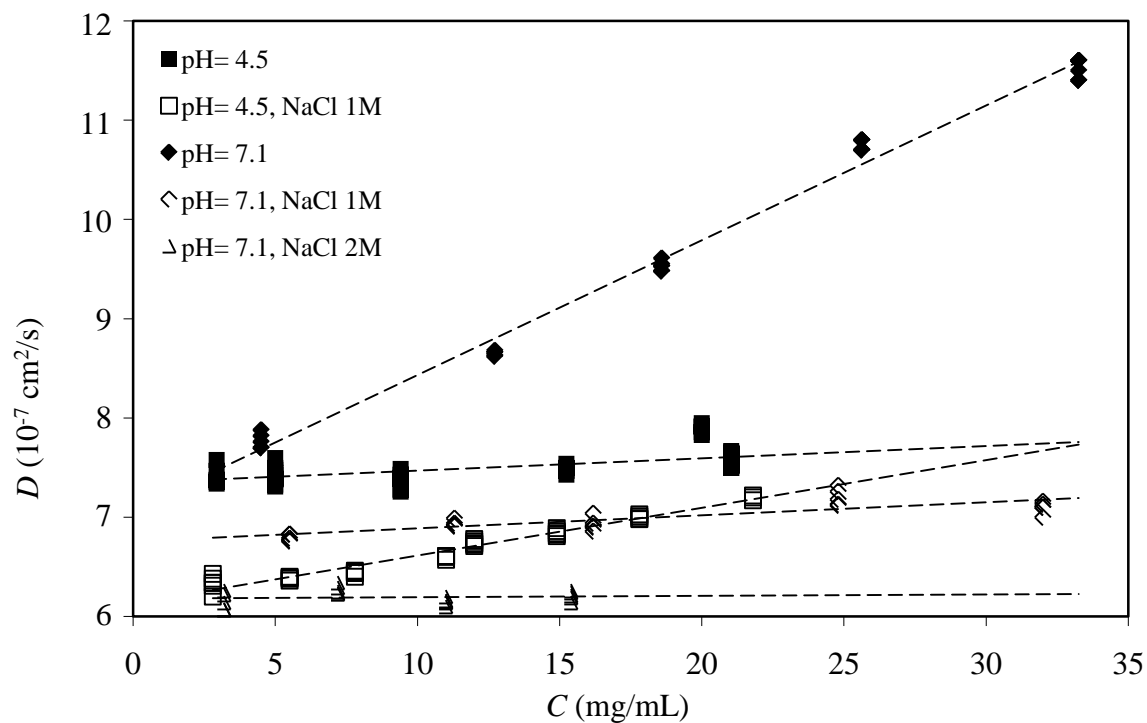
Figure 2. Diffusion coefficient, D, of laccase as a function of laccase concentration, C, at several ionic strengths at pH 4.5 and pH 7.1. The dashed lines are linear fits to the data.

3. Observation of Laccase Condensations

We have investigated the laccase condensations at pH 4.0. As shown in the previous section, laccase interactions are not significantly repulsive and precipitant agents such as PEG and salts can be used to bring about protein condensation. We further observe that the enzymatic activity of laccase at pH 4.0 is significantly higher than that observed at higher pH values.

We first investigate the effect of salt on laccase condensation. A stock solution of purified laccase in 0.1 M acetate buffer at pH 4.0 was mixed with salt stock solutions at 4°C. In order to assess the effect of the nature of the salt, we have used four different salts: $(NH_4)_2SO_4$, Na_2SO_4, NaCl, and $NaClO_4$. In all cases, the final laccase concentration was 20mg/mL and the salt ionic strength was 2.0 M.

The solutions were immediately stirred and incubated at 4 °C for 12 hours. In the case of $NaClO_4$, the laccase sample became opaque one minute after mixing, while samples remained transparent for the other salt cases even after 12 hours. In Fig 3(a, b), we show phase-contrast light-microscopy images for laccase aggregation induced by sodium perchlorate. The aggregates appear to be amorphous loose aggregates with a poor contrast with respect to the background. This protein-aggregation process was found to be irreversible. Indeed laccase aggregates could not be dissolved by both temperature increase and extensive dilution. The suspension had a laccase activity lower than 10% of the salt-free laccase sample. A similar irreversible $NaClO_4$–induced aggregation process was observed for BSA. Because $NaClO_4$ is an oxidizing agent, this irreversible aggregation process can be attributed to oxidation of protein molecules. Since the

NaClO$_4$–induced irreversible aggregation of laccase leads to a significant loss of enzymatic activity, we did not investigate this salt case further. However, we remark that concentration gradients of NaClO$_4$ are often used to purify proteins such as laccase by ionic exchange chromatography. Thus, our results raise some concerns on the use of NaClO$_4$ in protein-purification protocols.

No aggregation was observed for laccase in the presence of the other three salts at the same ionic strength of 2.0 M at 4°C and 22°C (room temperature). Since (NH$_4$)$_2$SO$_4$ has a high solubility in water, we were able to significantly increase the ionic strength of the protein samples by preparing laccase (20 mg/mL) solutions with 2.0 M (NH$_4$)$_2$SO$_4$. Interestingly, while the final solutions remained transparent at 4°C, sample opacification was observed by increasing sample temperature to 22°C. Sample opacification was found to be reversible; i.e. the samples returned to complete transparency by lowering the temperature or by sample dilution. Light microscopy showed that the precipitates were compact aggregates with a high contrast with respect to the background (Fig 3c,d). These observed laccase aggregates may be described as clusters of small protein-rich droplets with a diameter smaller than 1μm (Fig 3d). The suspension yielded the same laccase activity as that of salt-free laccase solution. Finally, we have checked the effect of PEG on laccase condensation. In this investigation, we have use PEG with two different molecular weights: 1450 g mol^{-1} (PEG1450) and 8000 g mol^{-1} (PEG8000). A 36% (w/w) PEG stock solution was added to a laccase solution in 0.1 M acetate buffer at pH 4.0 and 22 °C, and the final solution was 20 mg/mL in laccase and 12%(w/w) in PEG. The laccase samples appeared slightly opaque immediately after PEG addition. As shown in Fig 3(e, f) by light microscopy, the observed weak sample opacification is related to

dispersed aggregates. These aggregates have a globular compact shape typical of LLPS. The amount of laccase aggregates in the presence PEG8000 appeared to be significantly lower than that obtained in the case of $(NH_4)_2SO_4$. Sample opacification in the presence of PEG was found to be reversible; i.e. the samples returned completely transparent by sample dilution. However, temperature variations did not appreciably change sample opacification by visual inspection. The sample suspension yielded the same laccase activity as that of salt-free laccase solution.

We have also examined condensation of laccase at pH 5.9 in 0.015 M phosphate buffer. As shown by our DLS measurements, protein-protein interactions in the laccase-buffer system are significantly more repulsive. In the case of $(NH_4)_2SO_4$, laccase condensation was observed as at pH 4.0 (Fig 4a, b). However, a higher salt concentration (2.4 M instead of 2.0 M) was required to obtain the laccase aggregates.
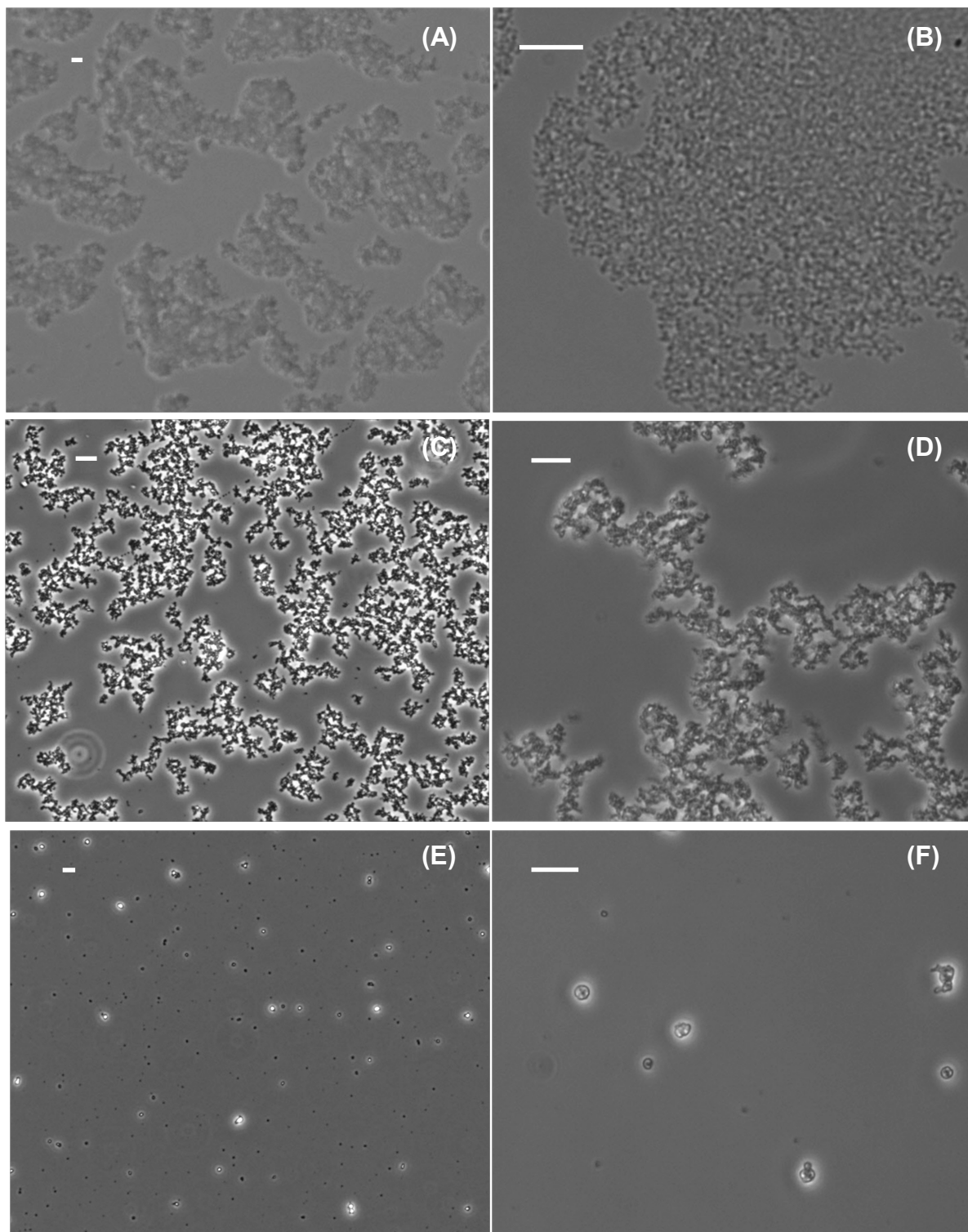
Figure 3. Light microscope images of laccase aggregates precipitated by NaClO$_4$ (A and B), (NH$_4$)$_2$SO$_4$ (C and D), and PEG8000 (E and F) at pH 4.0. The length of the white horizontal bar is 10μm.

In the case of PEG, no aggregation of laccase was observed. According to DLS experiments, 1.0 M NaCl can be used to screen electrostatic interactions. We have therefore prepared a laccase solution with 1.0 M NaCl and 6% (w/w) PEG. In this condition, laccase aggregation was found to be similar to that observed at pH 4.0 in the presence of PEG8000 (see Fig 4c, d).
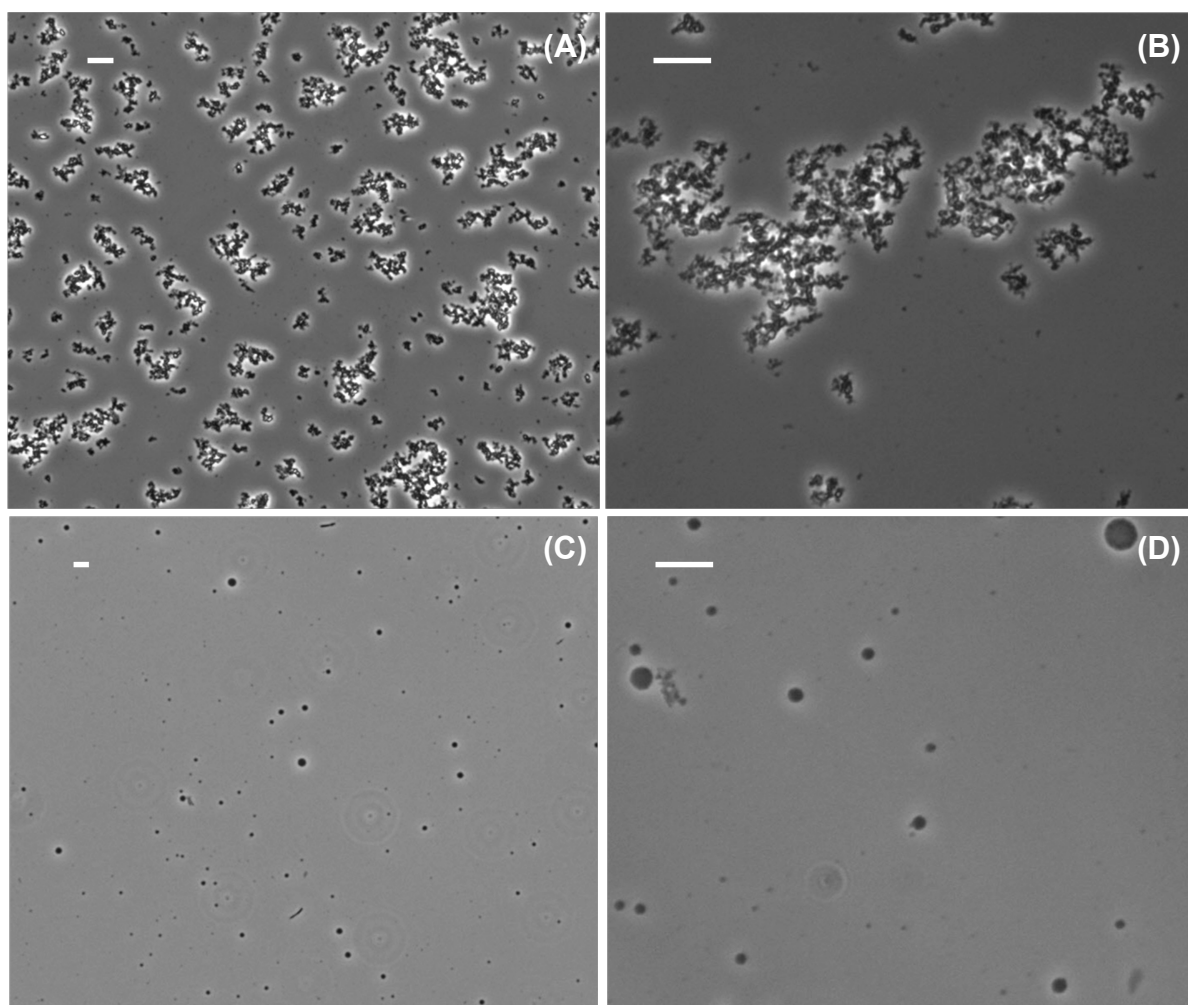
Figure 4. Light microscope images of laccase aggregates precipitated by $(NH_4)_2SO_4$ (A and B), and PEG 8k (C and D) at pH 5.9. The length of the horizontal white bar is 10μm.

4.   Phase Boundary for the Laccase-$(NH_4)_2SO_4$-Buffer System

In the previous section, we have observed that laccase condensation can be induced in the presence of $(NH_4)_2SO_4$.   This process is reversible and can be induced by increasing sample temperature. In order to quantitatively characterize the effect of temperature on laccase condensation, we have characterized the phase-separation boundary for the laccase-$(NH_4)_2SO_4$-buffer ternary system at pH 4.0 by turbidity measurements.

Laccase samples initially transparent at low temperature were prepared. Their turbidities were then measured as temperature was increased. A representative turbidity profile is shown in Fig 5.
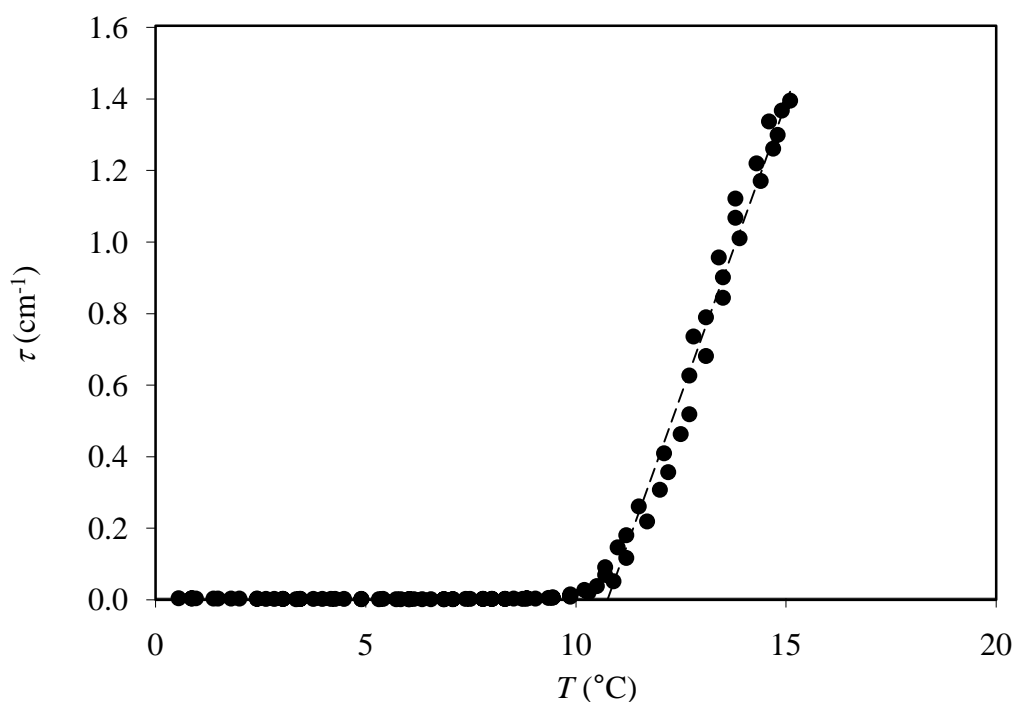
Figure 5. Turbidity of a representative laccase sample (20 mg/mL laccase, 2.0 M $(NH_4)_2SO_4$, 1.0 M NaCl, 0.1 M acetate buffer, pH 4.0) as a function of temperature during a heating process.

As we can see in the figure, an abrupt increase of turbidity during the heating process occurs at a well defined temperature, T_{ph}, which characterizes the onset of laccase condensation. We have measured T_{ph} as a function of $(NH_4)_2SO_4$ concentration at two laccase concentrations, 10 and 20 mg/mL, and at both pH 4.0 and 5.9 (see Fig 6).
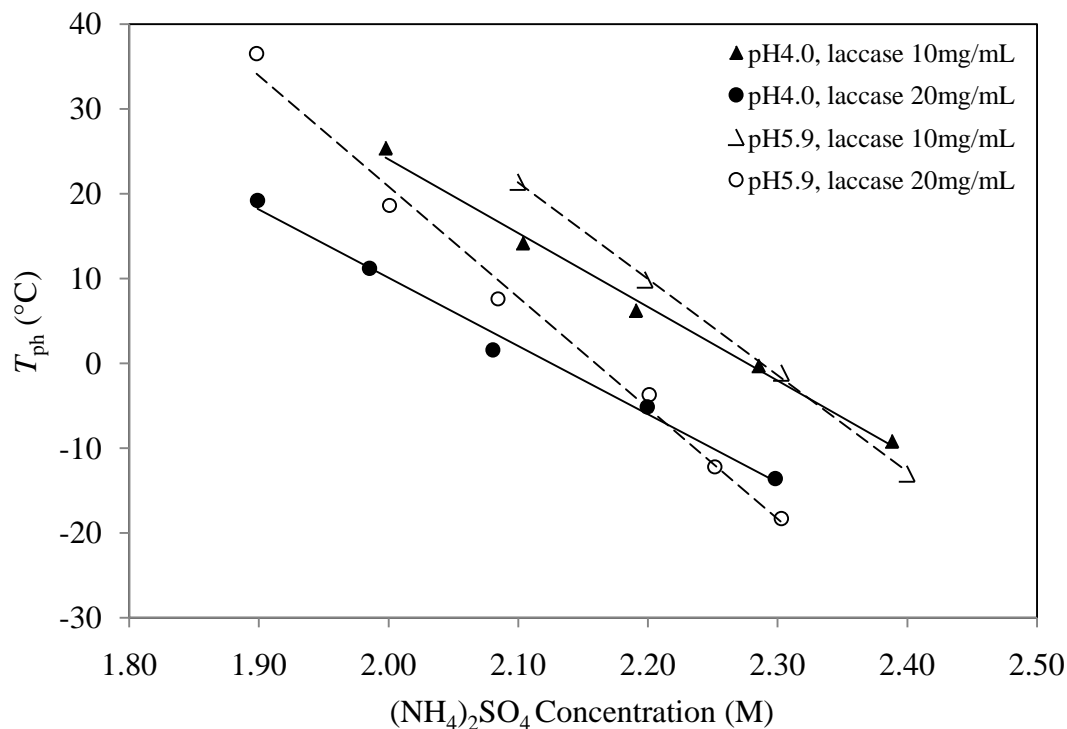
Figure 6. Phase transition temperatures of laccase solutions as a function of $(NH_4)_2SO_4$ concentration. The solid and dashed lines are linear fits to the data.

In this figure, we can see that T_{ph} decreases as $(NH_4)_2SO_4$ or laccase concentration increases. At a given pH, the effect of $(NH_4)_2SO_4$ concentration on T_{ph} is independent of laccase concentration. On the other hand, the effect of $(NH_4)_2SO_4$ concentration on T_{ph} was found to be higher at pH 5.9 than at pH 4.0. Specifically, the magnitudes of the slopes of the T_{ph} lines at pH 5.9 in Fig 6 were found to be 60% larger than those at pH 4.0.

The slopes are significantly different, and the T_{ph} lines at different pH intersect with each other. In other words, while an increase in pH produces an increase in T_{ph} at low temperature at relatively low $(NH_4)_2SO_4$ concentration, the opposite occurs at high $(NH_4)_2SO_4$ concentration. We can explain this behavior by observing that salts such as $(NH_4)_2SO_4$ act on the thermodynamics of protein solutions in two ways: 1) they introduce common-ion effects in the case of charged proteins and screen electrostatic interactions; 2) they are excluded from the surface of protein molecules due to protein preferential solvation. Both actions contribute to an increase in protein-protein attraction. At pH 4.0, the laccase net charge is small. Thus, the common-ion effect of $(NH_4)_2SO_4$ is small at this pH. On the other hand both effects of $(NH_4)_2SO_4$ contribute at pH 5.9, where the laccase net charge is relatively large.

In order to confirm that the observed aggregates are a laccase condensed phase, we have determined the activity of laccase in the supernatant solution and compared it to that of a corresponding homogeneous laccase solution. Laccase solutions (20 mg/mL in 0.1-M acetate buffer, pH 4.0) with four different concentrations of $(NH_4)_2SO_4$ (1.9, 2.0, 2.1 and 2.2 M) were investigated. These solutions were incubated at 22°C for 24 hours. In the case of the lowest $(NH_4)_2SO_4$ concentration, no phase separation occurs according to our phase-boundary measurements (Fig 6). As a reference case, we consider laccase activity in a $(NH_4)_2SO_4$-free solution. We find that the laccase activity for this case is 6.6 U/mg (1U=1μmol/min is the unit of substrate consuming rate) before and after the incubation time. All four incubated solution/suspensions had the same laccase activities (6.6 U/mg). The solution/suspensions were then centrifuged at 22°C and the laccase activity of the supernatant was measured as a function of $(NH_4)_2SO_4$ concentration. As shown in Fig 7,

the activity fraction retained in the supernatant was found to decrease as the $(NH_4)_2SO_4$ concentration increases. This result confirms the presence of laccase aggregates.
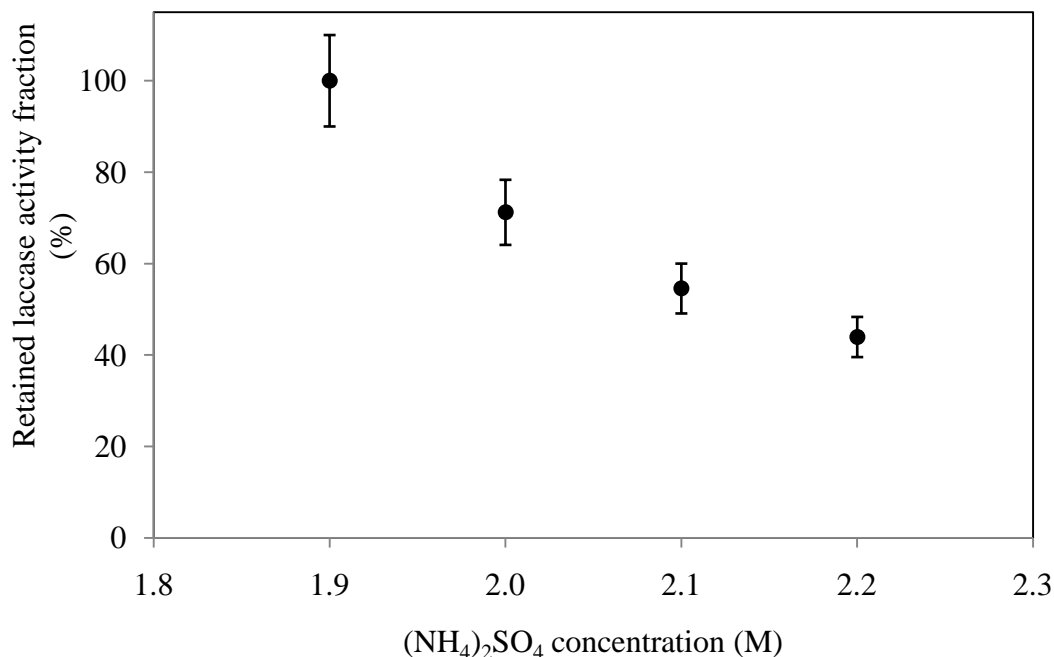
Figure 7. laccase activity in the supernatant as a function of $(NH_4)_2SO_4$ concentration.

For comparison, we have also investigated the effect of $(NH_4)_2SO_4$ on BSA and lysozyme condensation at pH 4.0. In both cases, condensation was observed in the presence of $(NH_4)_2SO_4$ in similar experimental conditions. Representative light-microscopy images for both BSA and lysozyme aggregates are shown in Fig 8. The morphology of these aggregates is similar to those observed in the case of laccase. In the case of BSA, we also observed a rapid increase of turbidity as temperature increases at a well-defined value of T_{ph}. The dependence of T_{ph} on $(NH_4)_2SO_4$ concentration, which is shown in Fig 9, is similar to that observed in the case of laccase. BSA aggregates were

also found to be reversible, although the temperature and dilution driven dissolution processes were found to be somewhat slower.
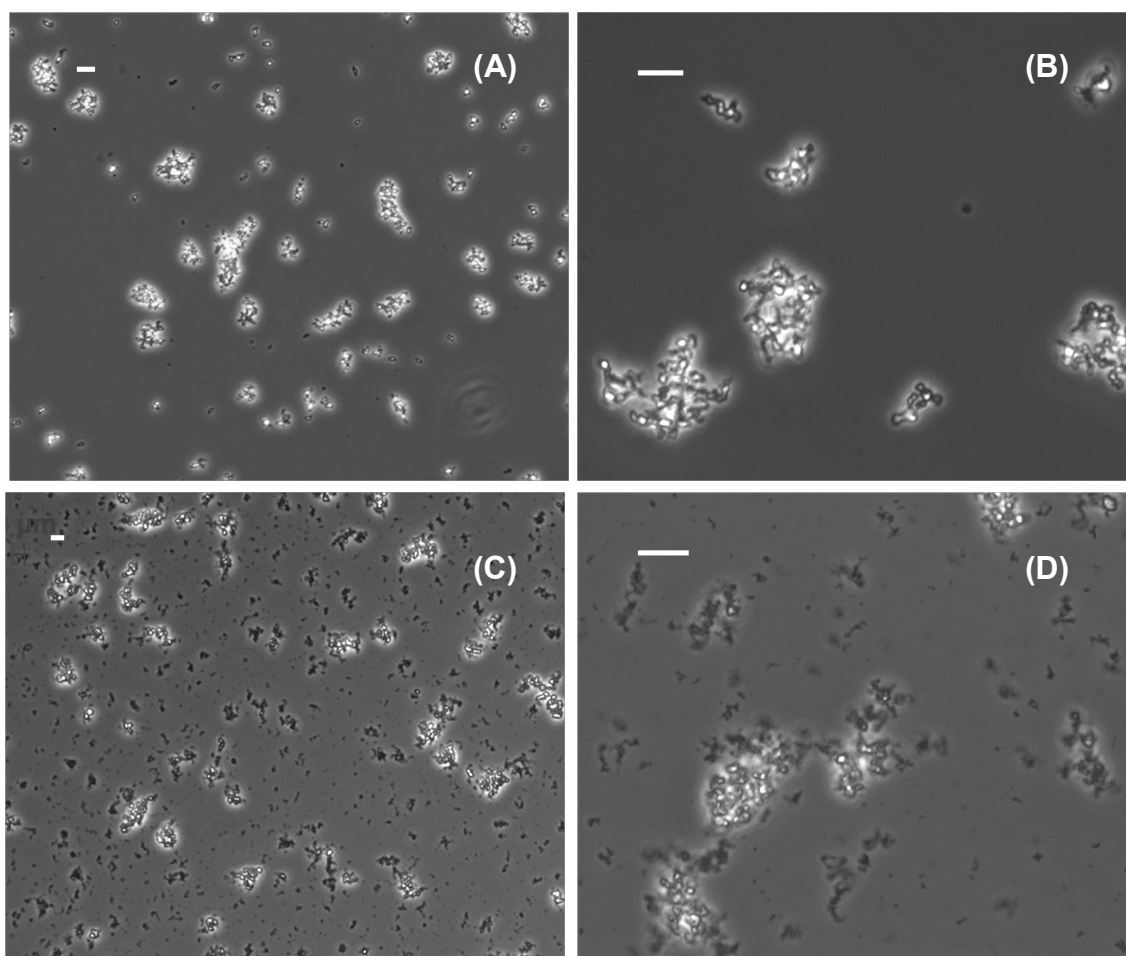
Figure 8. Light microscope images of BSA (A and B) and lysozyme (C and D) aggregates precipitated by $(NH_4)_2SO_4$ at pH 4.0. The white bar indicates 10μm.
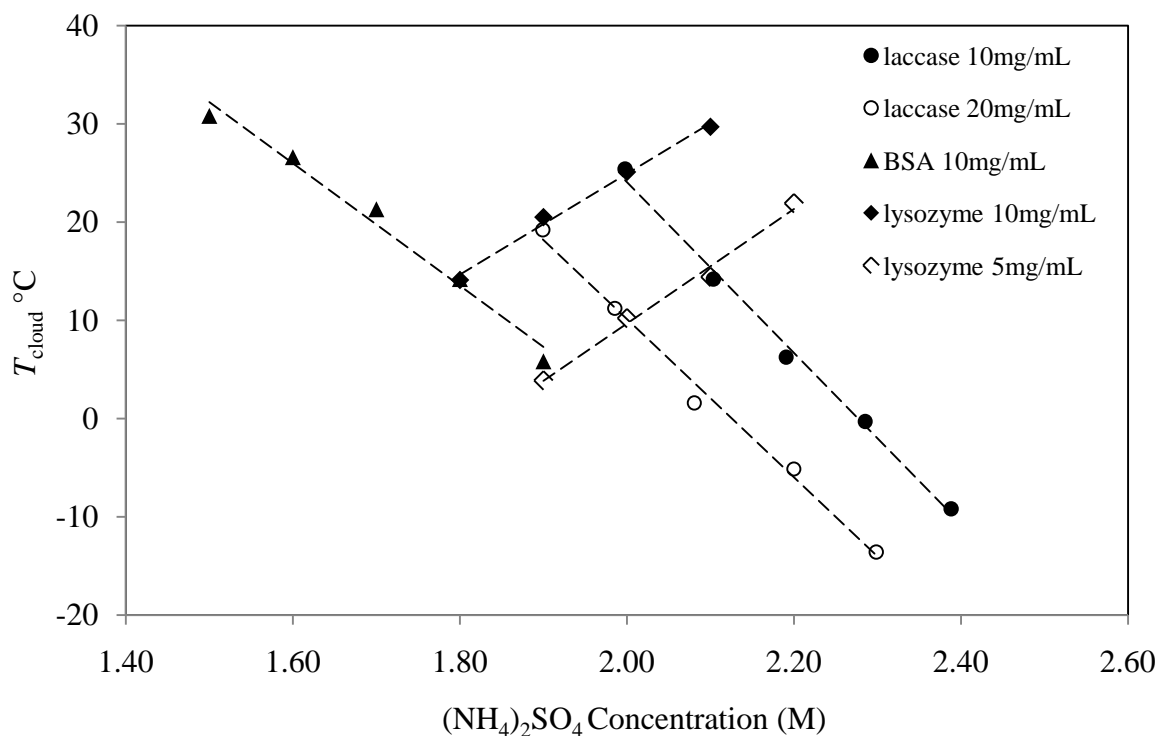
Figure 9. Phase transition temperatures for laccase, BSA and lysozyme in the presence of $(NH_4)_2SO_4$ at pH 4.0. The dashed lines are linear fits to the data.

In the case of lysozyme, we have however found that protein condensation is driven by lowering temperature (Fig 9). Consistently, the slope of the T_{ph} lines in Fig 9 has the opposite sign. This behavior is the opposite of that observed in the case of laccase and BSA. We speculate that the observed difference in the behavior of T_{ph} for the three protein cases may be related to a difference in common-ion effect. At pH 4.0, laccase (pI=3.5±0.5) and BSA (pI=5.0±0.2) have a small net charge. This implies that common-ion effects are weak. On the other hand, lysozyme (pI=10.5±0.5) has a large positive charge at this pH. This implies that the common-ion effect is relative strong in the case of lysozyme. Further experimental and theoretical work is needed to understand the effect of $(NH_4)_2SO_4$ on protein condensation.

5. Phase Boundary for the Laccase-PEG-Buffer System

We have also characterized the phase-separation boundary for the laccase-PEG-buffer ternary system at pH 4.0 by turbidity measurements. In the case of PEG1450 and PEG8000, the spherical laccase aggregates were observed in the presence of PEG concentration of 6% (w/w) at both pH 4.0 and pH 5.9. In the PEG8000 cases and PEG1450 case at pH 4.0, sample turbidity was found to steadily increase when lowering temperature without a sharp change to a well-defined temperature. In the case of PEG1450 at pH 5.9 (with 1.0 M NaCl), turbidity also increased when lowering temperature. Although the change of turbidity was not observed to be sharp as in the case of $(NH_4)_2SO_4$ the value of T_{ph} could be approximately determined in this case as shown in Fig 10. For this case, we have characterized laccase condensation on the phase boundary. As shown in Fig 11, the phase transition temperatures, T_{ph}, increases as PEG1450 concentration increases. This behavior is consistent with that obtained in the case of the LLPS of BSA-PEG-buffer (chapter 8) and lysozyme-PEG-NaCl-buffer systems (chapter 11). However, the turbidity values for the opaque laccase-PEG samples were significantly lower than those observed in the case of the BSA-PEG-buffer system and laccase-$(NH_4)_2SO_4$-buffer systems. This is consistent with the presence of few laccase aggregates as has been observed by light microscopy (Fig 4e, f). Further experimental and theoretical work is needed to understand laccase condensation in the presence of PEG.
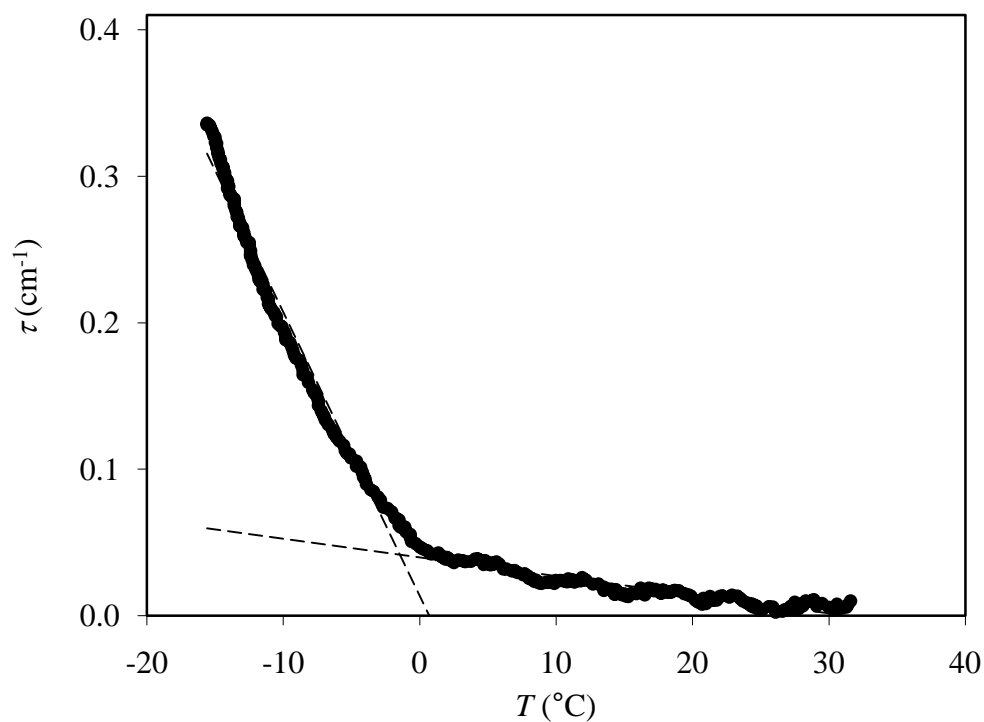
Figure 10. Turbidity of a representative laccase sample (20 mg/mL laccase, 13% (w/w) PEG1450, 1.0-M NaCl, 0.015-M phosphate buffer, pH 5.9) as a function of temperature during a cooling process.
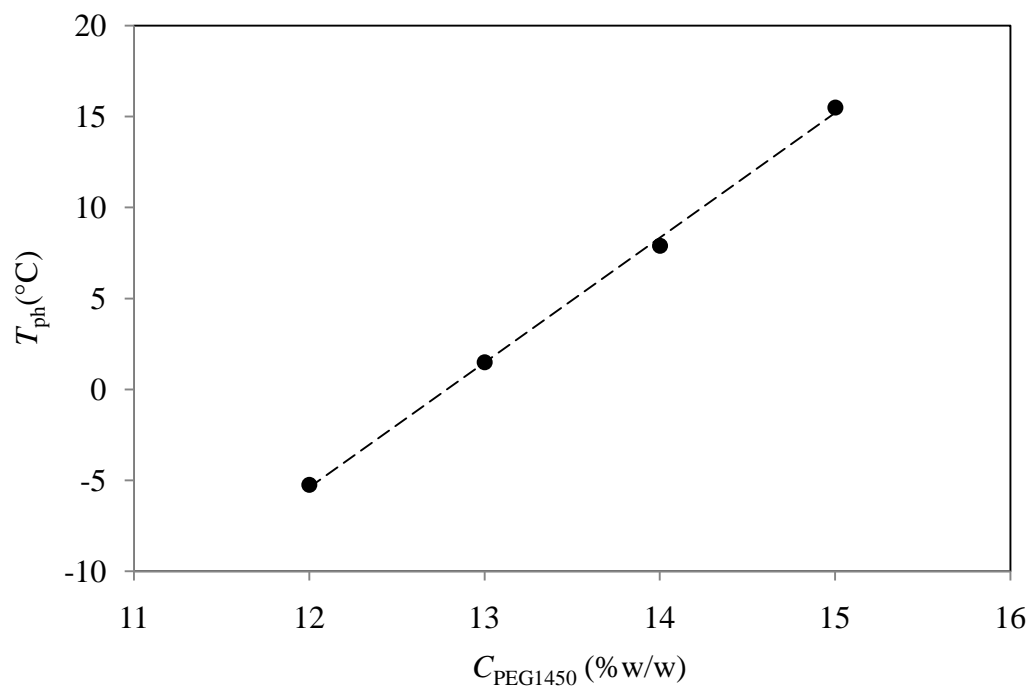
Figure 11. Phase separation temperature of laccase solution as a function of PEG1450 concentration at pH 5.9. The dashed line is a linear fit to the data.

Nonetheless, we can conclude this section by observing that 1) both $(NH_4)_2SO_4$ and PEG induce laccase condensation; 2) cooling enhances laccase condensation in the presence of PEG, while the opposite behavior was observed in the presence of $(NH_4)_2SO_4$; 3) PEG was found to be less effective than $(NH_4)_2SO_4$ in inducing laccase condensation.

6. Laccase condensation in the $(NH_4)_2SO_4$-PEG-buffer system

We have also examine the combined effect of PEG8000 and $(NH_4)_2SO_4$ on laccase condensation. We note that the PEG-$(NH_4)_2SO_4$-water system itself exhibit LLPS and centrifugation of these biphasic systems yield two liquid coexisting phase separated by a clear interface (Fig 12a). The bottom solution is rich in $(NH_4)_2SO_4$ (and with a very low PEG concentration) while the top solution is rich in PEG. Interestingly, the addition of laccase-buffer stock solutions to these biphasic systems forms a white solid film at the interface (Fig 12b). Laccase activity at the interface was found to be higher than that inside the two phases for laccase total concentrations as low as 0.1 mg/mL. Hence, the solid film at the interface is a laccase condensed phase. After extensive stirring and centrifugation, the laccase condensed film reformed at the interface. This indicates that the film density is lower than that of the bottom phase rich in $(NH_4)_2SO_4$. On the other hand, the laccase aggregates produced in the presence of $(NH_4)_2SO_4$ as described in the previous sections are located at the bottom of the samples after centrifugation. The corresponding $(NH_4)_2SO_4$ concentrations (and densities) were even higher than those used to produce the biphasic system. Surprisingly, polarized-light microscopy has shown

that the laccase film is birefringent (Fig 13). This indicates the presence of an ordered structure inside the film, which may be described as a crystalline-like material. Further experimental and theoretical studies will be needed on these biphasic systems. These studies will potentially lead to discovering new ways for controlling the formation of protein ordered structures, preparing cross-linked enzyme aggregates, purifying protein materials and performing enzyme catalysis at the liquid-liquid interface.

(A)   **(B)**
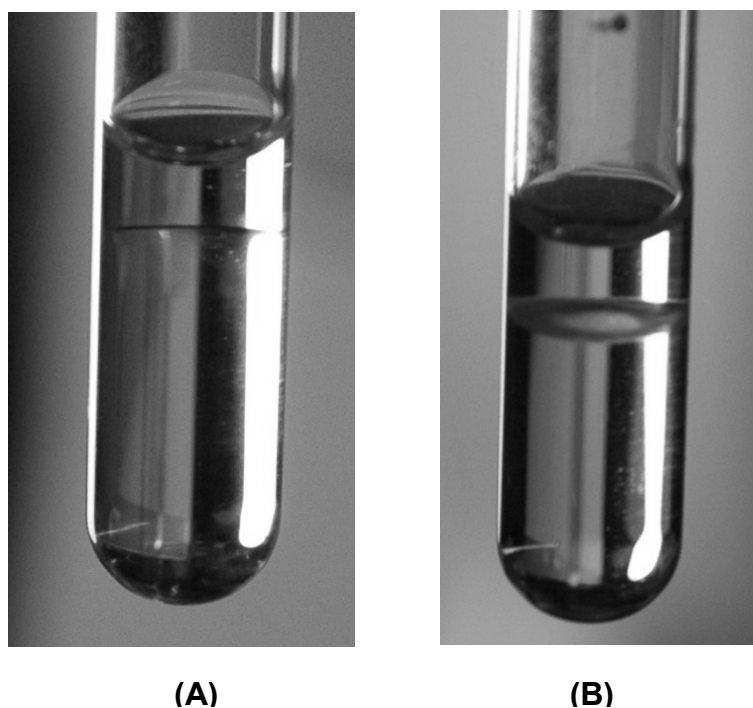
Figure 12. (A) PEG8000-$(NH_4)_2SO_4$-water biphasic system. (B) Laccase condensed film at the interface of two coexisting phase of PEG8000-$(NH_4)_2SO_4$-buffer system. The total concentration of laccase is 1 mg/mL, that of PEG 8000 is 10%(w/w), and that of $(NH_4)_2SO_4$ is 2M.
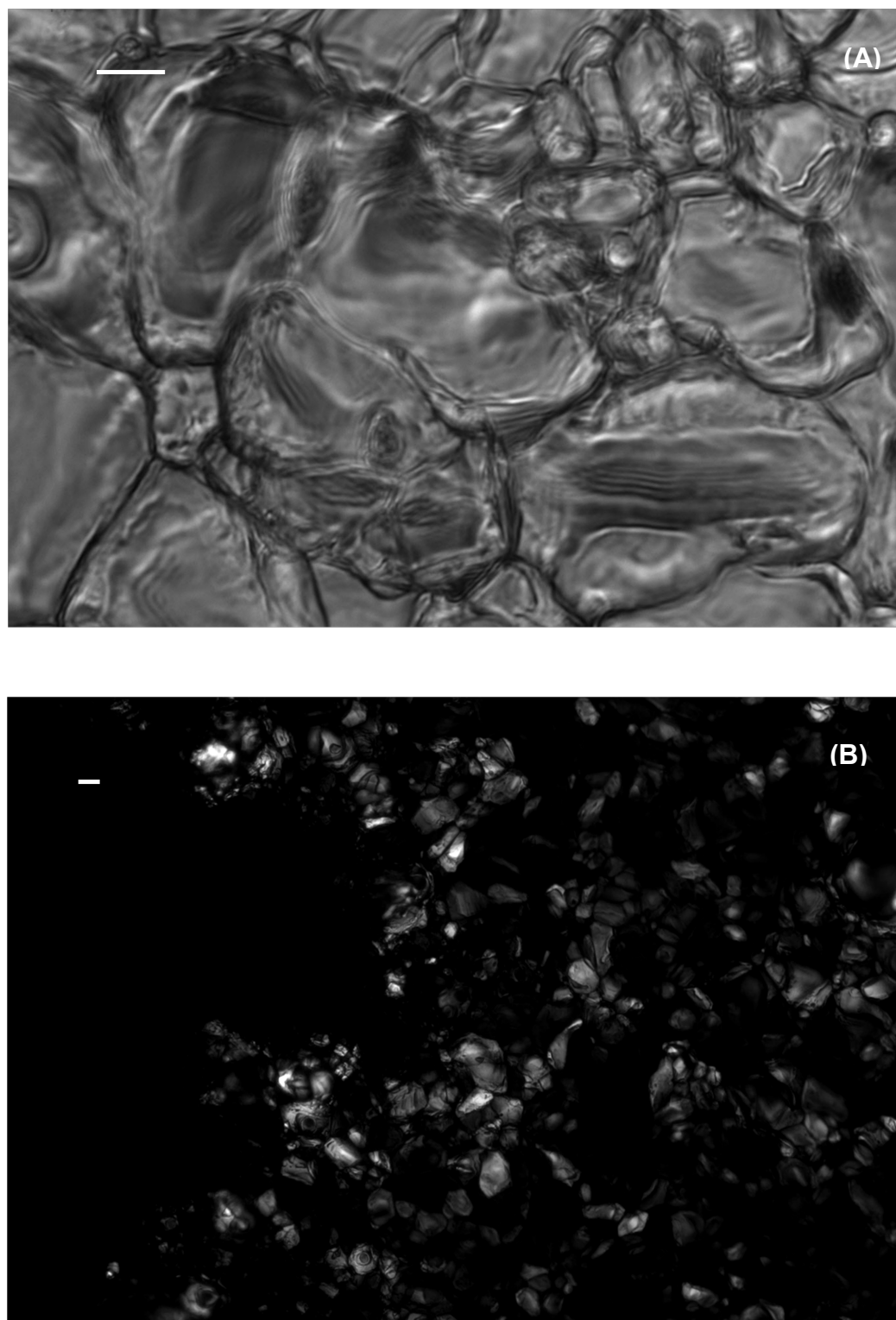
Figure 13. Light microscope images of the laccase condensed film. (A) Light-Microscopy image without polarization (A) and with polarization (B). The length of the horizontal bar is 10 μm.

7.   Coaggregation of Laccase in the presence of BSA

Finally, we have examined laccase-BSA coaggregation in the presence of $(NH_4)_2SO_4$ at pH 4.0. This approach may be useful for inducing enzyme condensation when the enzyme concentration is low. Solutions were prepared by mixing a 80 mg/mL BSA and 0.8 mg/mL laccase stock solution (0.1 M acetate buffer, pH 4.0) with a 4.0 M $(NH_4)_2SO_4$ stock solution. The final concentrations were 2.0 M $(NH_4)_2SO_4$, 10mg/mL BSA and 0.1 mg/mL laccase. Protein condensation was observed immediately after the addition of $(NH_4)_2SO_4$. The morphology of these aggregates characterized by light microscope (Fig 14a, b) is similar to that observed in the cases described above $(NH_4)_2SO_4$. The activity of the supernatant was then investigated as a function of time. The laccase activity was found to be independent of time (Fig 15). We have also verified that BSA has no effect on laccase activity. Our results show that laccase does not coaggregate with BSA under these experimental conditions.
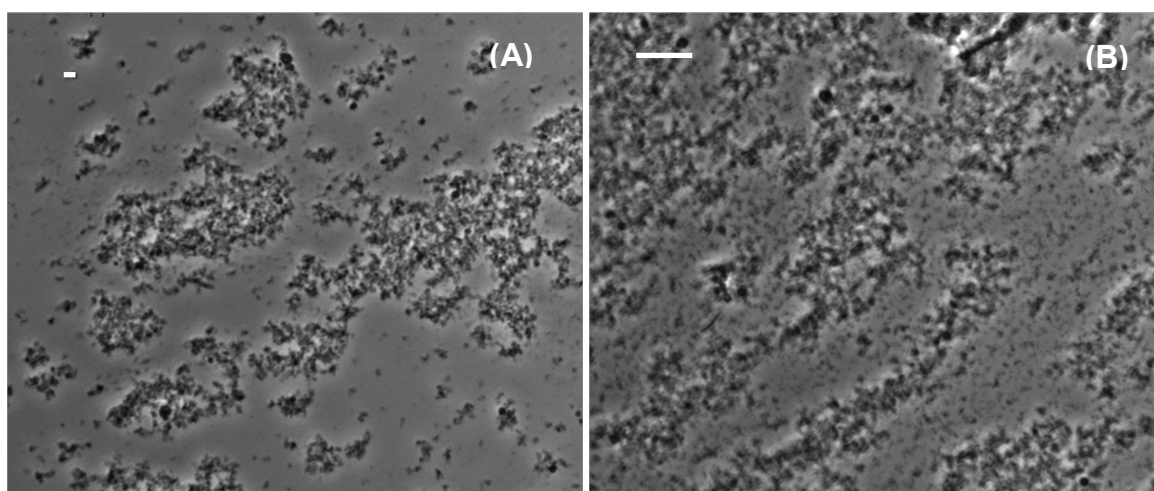
Figure 14. Light microscope images of aggregates precipitated from a BSA-laccase solution by $(NH_4)_2SO_4$. The solution is with 2M $(NH_4)_2SO_4$, 10 mg/mL BSA and 0.1 mg/mL laccase in 0.1M acetate buffer at pH 4. The scale bar indicates 10 μm.
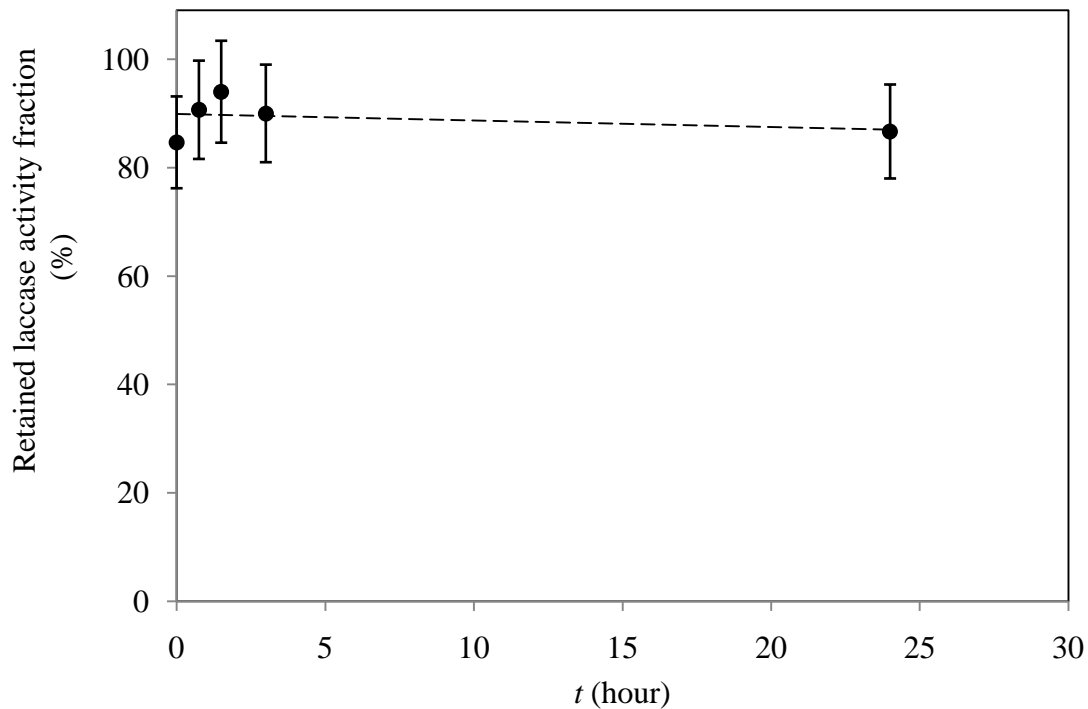
Figure 15. Percentage of retained laccase activity in supernatant as a function of incubation time. Protein samples (2.0 M $(NH_4)_2SO_4$, 10 mg/mL BSA and 0.1 mg/mL laccase in 0.1M acetate buffer, pH 4.0) were incubated at 25°C. The dashed line is a linear fit to the data.

We have also observed that laccase coaggregation did not occur also at 4°C and 35°C (Fig 16). Correspondingly BSA concentrations in the supernatants decreased as temperature increases as expected (Fig 17).
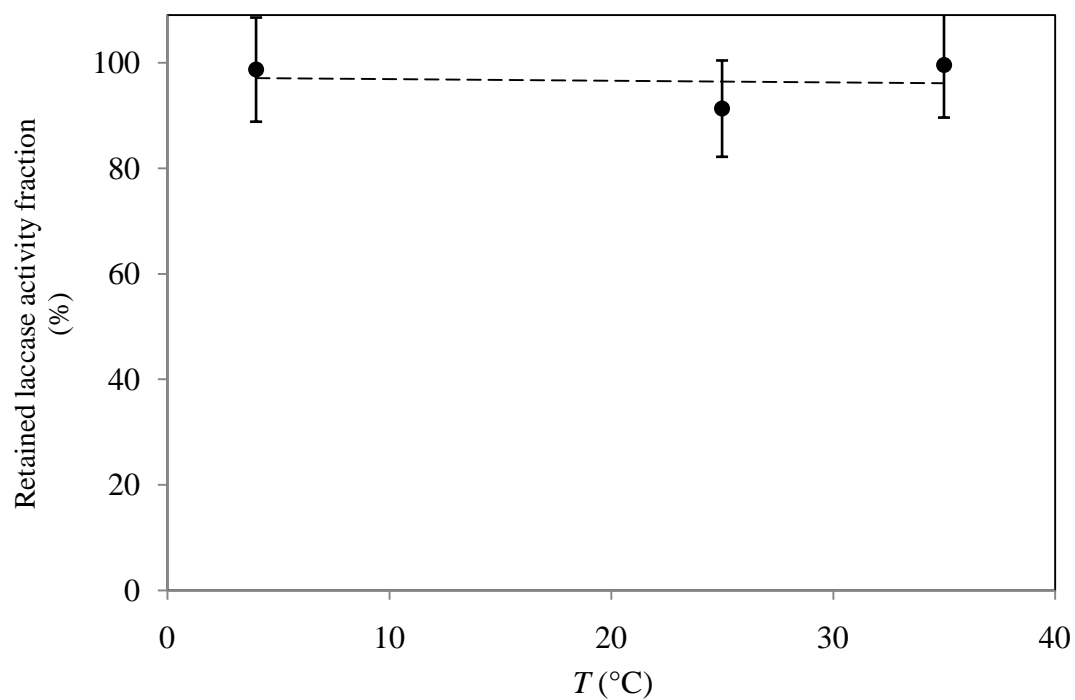
Figure 16. Percentage of laccase activity retained in supernatant as a function of aggregation temperature. Measurements performed after one hour of incubation. The dashed line is a linear fit to the data.
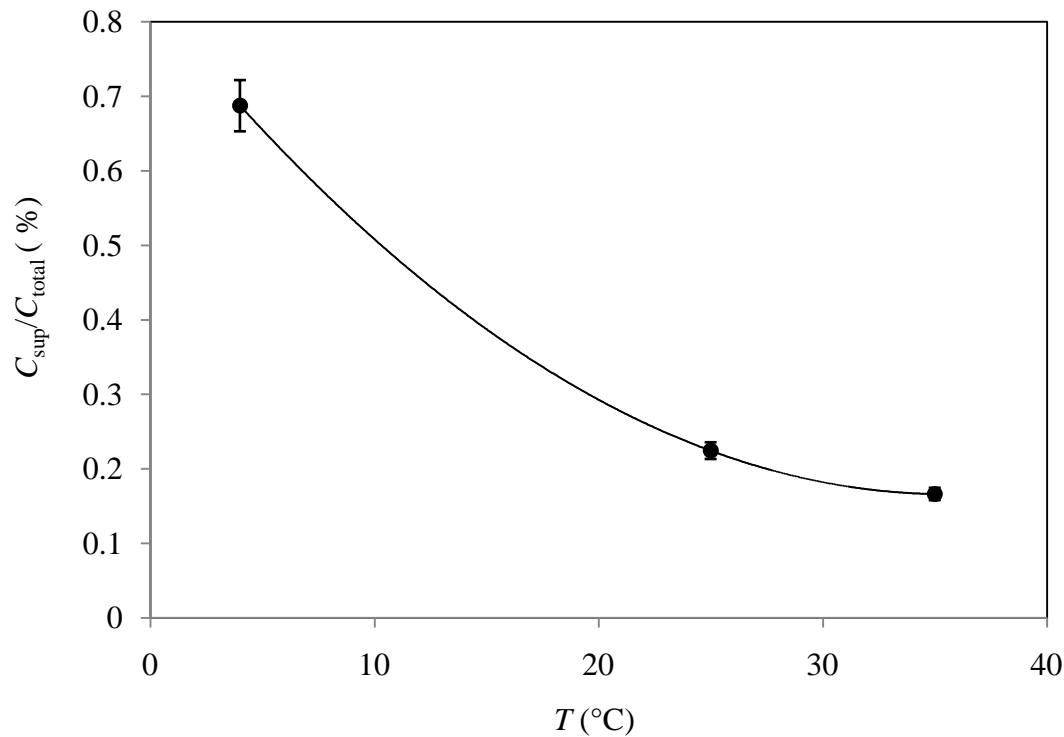
Figure 17. Fraction of BSA retained in supernatant as a function of aggregation temperature, where C_{sup} is the BSA concentration in supernatant, and C_{total} is the total BSA concentration in the total sample. The dashed line is a guide for the eye.

In order to enhance coaggregation of laccase in the presence of BSA, we had added glutaraldehyde as a chemical cross-linker. In a first approach, we have first mixed the BSA-laccase stock solution with $(NH_4)_2SO_4$, induced aggregation, and then added glutaraldehyde. This is the approach typically used to produce cross-linked enzyme aggregates.

The suspension was then incubated at 25 °C. The laccase activities in the supernatant were measured as a function of incubation time, t (Fig 18). At $t = 0$, the activity in the supernatant (8.7 U/mg) is slightly lower than that in the suspension (9.9 U/mg) and in that of the BSA-free and $(NH_4)_2SO_4$-free laccase-glutaraldehyde control samples (9.7 U/mg). The activity in the supernatant decreases with the incubation time, and dropped to 27 % (2.7 U/mg) of its initial value after 24 hours (Fig 18). The activity in the BSA-free and $(NH_4)_2SO_4$-free laccase-glutaraldehyde control samples (10.0 U/mg) did not decrease after 24 hours of incubation.
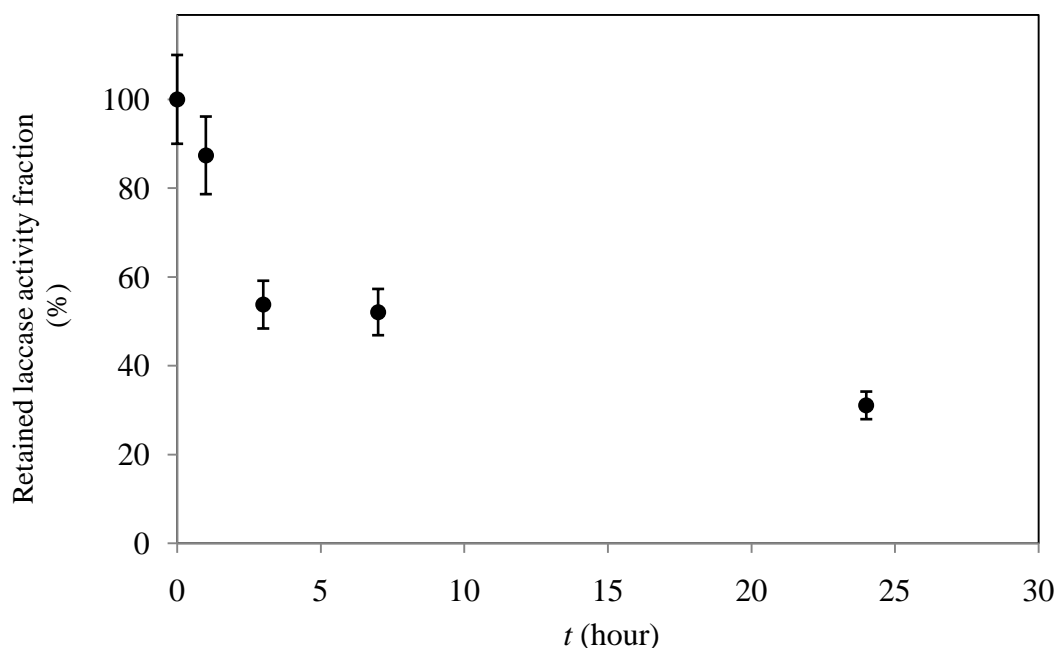
Figure 18. The percentage of retained laccase activity in supernatant as a function of incubation time. The suspension with 2M $(NH_4)_2SO_4$, 10 mg/mL BSA and 0.1 mg/mL laccase in 0.1M acetate buffer at pH 4 is incubated at 25°C.

Our results show that chemical cross-linking facilitates binding of laccase to BSA aggregates. Because laccase coaggregation does not occur in the absence of cross-linking, our results indicate that most laccase binds on the surface of BSA preformed aggregates.

In a second approach, we have first incubated the laccase-BSA homogenous stock solution with glutaraldehyde at 25 °C for 10 minutes, and then added $(NH_4)_2SO_4$ to reach the final concentrations. The suspension was then incubated at 25 °C for 1 hour and insoluble aggregates were obtained (Fig 19). The activity in the supernatant was only 0.3 % (0.033 U/mg) of the control samples. This result can be attributed to the initial formation of BSA-laccase oligomers due to chemical cross-linking. In this case, $(NH_4)_2SO_4$ induces the precipitation of the BSA-laccase oligomers. Thus our results for this second approach indicate that most laccase is located inside the aggregates. Because this second approach represents a novel method to produce cross-linked enzyme aggregates, we have isolated and characterized their enzymatic activity.

The suspension was first centrifuged and the supernatant removed. The cross-linked BSA-laccase aggregates were then washed with fresh acetate buffer three times and lyophilized. The insoluble cross-linked aggregates were placed into the substrate (ABTS) solution. Enzymatic activity was then measured under stirring. The activity of laccase in the aggregate (4.5 U/mg) was found to be 46% of that in the control sample. The decrease of laccase activity may be attributed to substrate diffusion restriction into the aggregates. Nonetheless, these aggregates represent a new type of enzymatically-active material. Moreover, these materials are expected to be robust because laccase is located inside the aggregates and may thus find applications in biotechnological processes.
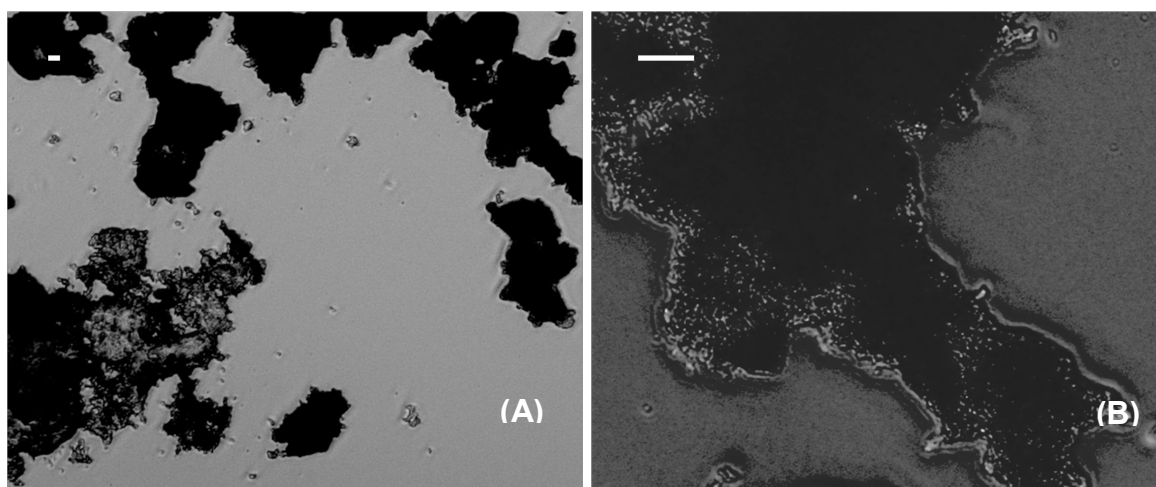
Figure 19. Light microscope images of aggregates precipitated from a BSA-laccase solution by $(NH_4)_2SO_4$ in the presence of 0.6% (w/w) glutaraldehyde. The length of the horizontal bar is 10 μm.

Conclusions

BSA in aqueous buffer solutions does not undergo condensation within the experimental temperature domain. However, condensation into oil-like spherical droplets is observed in the presence of PEG by lowering temperature. We have shown that the crowding effect of polymer on protein condensation can be examined by experimentally characterizing two thermodynamically independent properties of the LLPS phase boundary: (1) the effect of polymer concentration on the LLPS temperature and (2) protein/polymer partitioning in the two liquid coexisting phases (chapter 8). Indeed, the parallel examination of these two thermodynamic properties is a valuable tool for examining the reliability of existing depletion models and for developing more accurate ones. Using a thermodynamic perturbation theory that incorporates the free-volume fraction available to the polymer, the value of the normalized thickness, q, of the depletion layer surrounding the protein molecules could be extracted from the two independent properties of the phase boundary. For the BSA-PEG1450-buffer system, we have found that both properties yield the same value of q within the experimental error. This result represents a necessary condition for the validity of the depletion model. However, the value of q theoretically obtained from the radius of BSA and the radius of gyration of PEG1450 by treating PEG as an ideal chain ($q = 0.57$) was found to be significantly higher than that determined from the LLPS experiments ($q = 0.31$). We have hypothesized that the non-ideal behavior of the PEG chain significantly contributes to the observed discrepancy. Using Monte-Carlo simulations on a model system (chapter 9), we have found that the length of the chain segments, segment-segment bond angle, segment-

segment steric repulsion and trans-gauche conformational transition energy appreciably contribute to the value of q. For PEG, we have built a two-parameter model, a self-avoiding chain, which has produced a value of q significantly closer to the experimental one (q=0.39). We believe that this two-parameter model can be used in describing the crowding effect of PEG in aqueous solutions. Our simulations also predict that q for polymers is not a constant and it is expected to significantly decrease as protein concentration increases. Monte Carlo simulations on a model system also predict that the effect of temperature on depletion interactions is small in the case of the BSA-PEG1450-buffer system within the experimental temperature domain (chapter 10). The effects of non-spherical shape of globular proteins and protein-polymer attractive interactions (e.g. van der Waals interactions) on q are aspects that should be investigated in the future.

The role of protein oligomerization on protein condensation was also investigated (chapter 11). We have demonstrated that the formation of protein-rich droplets can be isothermally induced by protein cross-linking in aqueous solutions. These droplets evolve into cross-linked protein microspheres. We related this phenomenon to the LLPS properties of the protein monomer and to the increase of LLPS temperature during protein oligomerization. When macroscopic aggregation competes with LLPS, a rationale choice of pH, PEG, and salt concentrations may be used to favor LLPS relative to aggregation. This work contributes to the fundamental understanding on both phase transitions of protein solutions and morphology of cross-linked protein precipitates. It also provides guidance for the development of new methods for the stabilization of the protein liquid cluster and the preparation of protein-based materials. The mild conditions

of temperature and chemical environment used in these experiments also can be extended to more labile proteins.

We have investigated laccase condensation from aqueous solutions (chapter 12). We have found that both $(NH_4)_2SO_4$ and PEG induce laccase condensation without loss of enzymatic activity. However the phase behavior of laccase solutions in the presence of $(NH_4)_2SO_4$ is substantially different from that in the presence of PEG. Heating enhances laccase condensation in the presence of aqueous $(NH_4)_2SO_4$, while the opposite behavior was observed in the presence of aqueous PEG. We also observed a distinct condensation of laccase at the interface of a $(NH_4)_2SO_4$–rich liquid phase and PEG-rich liquid phase. The resulting laccase film displays birefringence. Finally, we have shown that laccase condensation can be driven by BSA-laccase cross-linking and BSA condensation when the enzyme concentration is low. This has produced catalytically-active laccase-BSA coaggregates. The reported preliminary results on laccase provide the basis for more systematic experimental and theoretical investigations on mixtures containing laccase. These studies should also be extended to other enzyme cases.

Appendix A

The Scaled Particle Theory

For many fluids, such as gases at high volume fraction, liquid crystals and the mixtures of macromolecules or colloidal particles, the excluded volume effect is the dominant interaction. Scaled particle theory provides a semi-macroscopic route to obtain the equation of state and in turns other thermodynamic properties for these fluids.[93,94] In this appendix, scaled particle theory is used to derive the equation of state for the hard-sphere (HS) simple fluid (see eq 18) and expression of the free-volume fraction for an AO-sphere (see Chapter1 Section 6) in the HS fluid (see eq 25).

First, we consider the general case of a fluid with pair-wise intermolecular potential $v(r)$. According to the virial theorem, the pressure of this fluid is given by,

$$P = \frac{N}{V}k_BT - \frac{1}{6V}\langle \sum_{i=1}^{N}\sum_{j>i}^{N} r_{ij}\,v'(r_{ij})\rangle \tag{1}$$

where N is the number of spherical molecules in the fluid, V is the volume of the fluid, k_B is Boltzmann constant, T is the temperature, r_{ij}'s are the intermolecular distance, $v'(r) \equiv dv(r)/dr$ is the potential gradient in the radial direction of a spherical molecule, and the angular parentheses stands for taking the average over all the spatial configuration of the fluid molecules.

Since the N spheres are identical, eq 1 can be rewritten into,

$$P = \frac{N}{V}k_BT - \frac{N}{6V}\langle \sum_{j=2}^{N} r_{1j}\,v'(r_{1j})\rangle = \rho k_BT\left(1 - \frac{1}{6k_BT}\int rv'(r)\rho g(r)\mathrm{d}\mathbf{r}\right) \tag{2}$$

where $\rho = N/V$ is the average density of the fluid, $g(r) \equiv \rho(r)/\rho$ is the radial distribution function, $\rho(r)$ is the local density around a given sphere, and $\mathbf{r}$ is the radial vector. Therefore, the compressibility factor is given by,

$$Z \equiv \frac{P}{\rho k_B T} = 1 - \frac{2\pi\rho}{3k_B T}\int_0^\infty r^3 v'(r)g(r)\mathrm{d}r \tag{3}$$

For the HS fluid, $v(r)$ has a discontinuity point at $r = \sigma$. In order to apply eq 3 to the HS fluid , we express $g(r)$ in a form of $g(r) = y(r)\exp(-v(r)/k_B T)$, where the $\exp(-v(r)/k_B T)$ describe the exponential decay when the two spheres are near contact at the high density limit, and $y(r)$ reduces to 1 at the low density limit. If we substitute this expression of $g(r)$ into eq 3, the compressibility factor become:

$$Z = 1 + \frac{2\pi\rho}{3}\int_0^{+\infty}\left[\frac{\mathrm{d}}{\mathrm{d}r}\exp(-\beta v(r))\right]y(r)r^3\mathrm{d}r \tag{4}$$

where $\beta \equiv 1/k_B T$. For hard sphere, $v(r) = \infty$ when $r \leq \sigma$, where σ is the diameter of the hard-spheres, and $v(r) = 0$ when $r > \sigma$, thus $\exp(-\beta v(r))$ is the Heaviside step function $\theta(r - \sigma)$. The derivative of θ is the Dirac δ function. According to the property of Dirac δ function,

$$Z = 1 + \frac{2\pi\rho}{3}\int_0^{+\infty}\delta(r - \sigma)y(r)r^3\mathrm{d}r = 1 + \frac{2\pi\rho}{3}y(\sigma)\sigma^3 \tag{5}$$

For hard-spheres, $y(r) = g(r)$ when $r > \sigma$. Due to the continuity of $g(r)$ at the limit of $r \to \sigma^+$, the compressibility factor for hard-sphere fluid is given by:

$$Z = 1 + 4\phi g(\sigma) \tag{6}$$

where $\phi = \pi\rho\sigma^3/6$, and $g(\sigma)$ is $g(r)$ at the $r \to \sigma^+$ limit.

In order to determine the expression for $g(\sigma)$, we investigate the microscopic structure of the fluid by scaled particle theory. We Consider the probability, $p(R)$, of finding(creating) a sperical cavity with radius R in the fluid, i.e. of inserting a hard-sphere with radius $R' = R - \sigma/2$ into the fluid. This probability is $p(R) = Q(R,N,V)/Q(N,V)$, where $Q(N,V)$ is the partitioning function of the fluid and $Q(R,N,V)$ is the partitioning function of the fluid with a cavity. Since the Helmoholtz free energy in the canonical ensemble is given by $F = -k_B T \ln(Q)$, the free energy change caused by creating the cavity is given by,

$$\Delta F(R) = -k_B T \ln\left(\frac{Q(R,N,V)}{Q(N,V)}\right) = -k_B T \ln(p(R)) \tag{7}$$

When $0 \le R \le \sigma/2$, i.e. inserting a sphere with zero or negative radius into fluid, we can write:

$$\Delta F(R) = -k_B T \ln\left(\frac{Q(R,N,V)}{Q(N,V)}\right) = -k_B T \ln\left(1 - \frac{4\pi R^3}{3}\rho\right) \tag{8}$$

When $R \ge \sigma/2$, the work for creating a macroscopic spherical cavity at constant T is $W(R) = W_0(\phi) - P\Delta V + \gamma \Delta A = \Delta F(R)$. The first term, which is independent of R, mantains continuity of $W(R)$ at the point $R = \sigma/2$, the second term takes into account of the volumetric work, and the third term takes into account the increase of surface energy. The change of volume $\Delta V = -4\pi R^3/3$ and the change of surface area is $\Delta A = 4\pi R^2$. Thus, the free energy change caused by creating the cavity is given by,

$$\Delta F(R) = W_0(\phi) + P\frac{4\pi R^3}{3} + \gamma 4\pi R^2 \tag{9}$$

where

$$\gamma = \gamma_0(1 - \chi\sigma/R) \tag{10}$$

where χ is a dimensionless correction factor taking into acount the curvature of the cavity.

The basic assumption of scaled particle theory is that eqs 9-12 for creating a macroscopic spherical cavity also hold for creating the microscopic cavity, i.e, inserting a spherical molecule into the fluid, when $R \geq \sigma/2$.

To study the fluid structure, which is charaterized by the radial distribution function $g(r)$, we consider the probability of creating a spherical cavity with radius between R and $R+$ dR. This probability is equal to $-dp(R) = p(R) - p(R + dR)$. This probability can be also expressed in another way by the product of the probability of creating a cavity of radius R and a conditional probability that there is at least one sphere whose center is located with in the interval $[R, R+$ d$R]$. Since the probability for a specific sphere presenting in the interval is $(\rho_{local}\, 4\pi R^2 dR)/N$, where ρ_{local} is the local density within $[R, R+$ d$R]$, and the N spheres are indistinguishable, the conditional probability can be expressed by $\rho_{local}\, 4\pi R^2 dR$. Therefore, the probability of creating a spherical cavity with radius between R and $R+$ dR can be expressed by

$$-dp(R) = p(R)(4\pi R^2 \rho G(R)\mathrm{d}R), \qquad (11)$$

where $G(R) \equiv \rho_{local}\,/\rho$ is the HS radial distribution function in the presence of the cavity at R, i.e. the inserted new sphere. We comment that, when $R = \sigma/2$, $G(\sigma) = g(\sigma)$. Thus, according to eq 8,we can write:

$$Z = 1 + 4\phi G(\sigma) \qquad (12)$$

From eq 11, we have: $G(R) = [-\mathrm{dln}(p(R))/\mathrm{dR}\ ]/(4\pi R^2 \rho)$. Hence, using eq 7, we obtain:

$$G(R) = \frac{\beta}{4\pi R^2 \rho}\frac{d\Delta F}{dR} \qquad (13)$$

If we substitute eqs 9, 10 into eq 13, we obtain:

$$G(R) = Z + \frac{4\Gamma}{\phi\kappa} - \frac{4\Gamma\chi}{\phi\kappa^2} \qquad (14)$$

where $\Gamma \equiv \pi\beta\gamma_0\sigma^2/6$, and $\kappa \equiv 2R/\sigma$.

Therefore, by substituting eq 14 into eq 12, we obtain:

$$Z = \frac{1 + 8\Gamma - 4\Gamma\chi}{1 - 4\phi} \qquad (15)$$

In eq 15, there are two unknowns Γ and χ both related to the surface tension. These two

unknowns can be determined by applying the continuity conditions for $G(R)$ at $R = \sigma/2$.

When $0 \le R \le \sigma/2$, eq 8 and 13, , we obtain:

$$G(R) = \frac{1}{1 - \phi\kappa^3} \qquad (16)$$

Since $G(R)$ and its first derivative $G'(R)$ are continuous at $R = \sigma/2$, by matching eq 14

and eq 16 at $R = \sigma/2$, we obtain:

$$8\Gamma = \frac{2\phi(2 + \phi)}{(1 - \phi)^2} - 4Z\phi \qquad (17a)$$

$$4\Gamma\chi = \frac{\phi(1 + 2\phi)}{(1 - \phi)^2} - Z\phi \qquad (17b)$$

If we substitute eqs 17a,b into eq 15, an equation of state for the HS fluid is derived:

$$Z = \frac{1 + \phi + \phi^2}{(1 - \phi)^3} \qquad (18)$$

Using eq 18, we can calculate the free-volume fraction for a AO sphere in the HS fluid.

By substituting eq 18 into eqs 17a,b, we have:

$$\Gamma = \frac{-3\phi^2(1 + \phi)}{4(1 - \phi)^3} \qquad (19a)$$

$$\Gamma\chi = \frac{-3\phi^3}{4(1-\phi)^3} \tag{19b}$$

Then, the surface tension can be obtained by substituting eqs 19a,b into eq 10:

$$\gamma = \frac{9\phi^2}{2\pi\beta\sigma^2(1-\phi)^3}\left[\frac{2\phi}{\kappa} - (1+\phi)\right] \tag{20}$$

Thus, according to eq 9, the free energy change by adding a AO sphere with radius $R' = R - \sigma/2$, where $R' > 0$, into the HS fluid is given by

$$\Delta F(R) = W^{(0)} + \frac{\phi(1+\phi+\phi^2)\kappa^3}{\beta(1-\phi)^3} + \frac{9\phi^2\kappa^2}{2\beta(1-\phi)^3}\left[\frac{2\phi}{\kappa} - (1+\phi)\right] \tag{21}$$

If we define $\eta \equiv \phi/(1-\phi)$, eq 21 can be rewritten as

$$\beta\Delta F(R) = \beta W^{(0)} + (\eta + 3\eta^2 + 3\eta^3)\kappa^3 - \frac{9}{2}(\eta^2 + 2\eta^3)\kappa^2 + 9\eta^3\kappa \tag{22}$$

We can nowapply continuity of $W(R)$ at $R= \sigma/2$ by matching eq 8 to eq 22:

$$\beta W^{(0)} = -\ln(1-\phi) - (\eta + 3\eta^2 + 3\eta^3) + \frac{9}{2}(\eta^2 + 2\eta^3) + 9\eta^3 \tag{23}$$

We then introduce the reduced thickness of the depletion layer $q = 2R'/\sigma = \kappa - 1$. This allows us to obtain the expression for the free-volume in terms of ϕ and q using eqs 22, 23:

$$\beta\Delta F(R) = -\ln(1-\phi) + (3q + 3q^2 + q^3)\eta + \left(\frac{9}{2}q^2 + 3q^3\right)\eta^2 + 3q^3\eta^3 \tag{24}$$

Because there is only excluded-volume effect in the HS fluid, the free-volume fraction for the HS solute is equal to the probability of finding a cavity with radius of R' inside the fluid:

$$\alpha(\phi, q) = \exp\left(-\beta\Delta F(R)\right) = (1-\phi)\exp(-A\eta - B\eta^2 - C\eta^3) \tag{25}$$

where $A = 3q + 3q^2 + q^3$, $B = \frac{9}{2}q^2 + 3q^3$ and $C = 3q^3$.

Appendix B

Density Profile of Ideal Chain

Near a Hard-Sphere Surface

In this appendix, we derive the expression of monomer density of an ideal chain at distance r from the center of a hard-sphere (see eq 34) adapted from Einsenriegler[124] theory. Using this expression, the thickness of depletion layer for an ideal chain can be defined (see eqs 29).

We first consider an ideal chain with a characteristic length $\sqrt{\tau}$, where τ is an apparent time, and two fixed ends at coordinates r and r'. The partition function $Z(\tau, r, r')$ of this ideal chain is described by the following diffusion-like equation:

$$\frac{\partial Z}{\partial \tau} = \nabla^2 Z \tag{1}$$

Note that $Z(\tau, r, r')$ is symmetric in r, r', thus the Laplacian operator is acting only on r. The initial condition is:

$$Z(0, r, r') = \delta(r - r') \tag{2}$$

where δ is the Dirac distribution.

The solution of the equation with *free-boundary* conditions is

$$Z = \frac{1}{(4\pi\tau)^{3/2}} e^{-\frac{|r-r'|^2}{4\tau}} \tag{3}$$

We can now appreciate the link between the radius of gyration, R_g, and τ. Indeed the square average end-to-end distance $\langle |r - r'|^2 \rangle$ is given by

$$\langle |r - r'|^2 \rangle = 4\pi \int_0^\infty |r - r'|^4 Z(r, r')\mathrm{d}\,|r - r'|$$

$$= \frac{4\pi}{(4\pi\tau)^{3/2}} \int_0^\infty |r - r'|^4 e^{-\frac{|r-r'|^2}{4\tau}} \mathrm{d}|r - r'|$$

$$= 6\tau \tag{4}$$

Since $R_g^2 = (1/6)\langle |r - r'|^2 \rangle$, we obtain $R_g = \sqrt{\tau}$.

The partitioning function $Z(\tau, r, r')$ refers to an ideal chain with fixed ends. Actually, it denotes the ratio between the partition function of a chain with two fixed ends and the partition function of a chain with one end position integrated. Thus, it has the dimension of an inverse of volume. The diffusion equation describes how $Z(\tau, r, r')$ changes as the chain gets longer starting from r'.

The probability that a given chain starts from r' is uniform throughout the space. We can define the partitioning function for a chain with only one end fixed at r by

$$\hat{Z}(\tau, r) = \int_V Z(\tau, r, r')\mathrm{d}^3 r' \tag{5}$$

We note that the density of chain ends at the point r exhibits a dependence on r that is proportional to the Boltzmann-weighted number of configurations of a single chain with on end fixed at r. For ideal chains, the density of chain ends is thus proportional to the partition function, $\hat{Z}$. We can apply the diffusion equation to $\hat{Z}$. This allows us to obtain:

$$\frac{\partial \hat{Z}}{\partial \tau} = \nabla^2 \hat{Z} \tag{6}$$

The initial condition becomes:

$$\hat{Z}(0, r) = 1 \tag{7}$$

We now consider a spherical box with radius R_∞ infinitely large. At the center of this box (and of the reference system) there is one hard sphere of radius R. Thus, r is defined only for $r > R$. This leads to the boundary condition:

$$\hat{Z}(\tau, R) = 0 \quad \text{and} \quad \hat{Z}(\tau, R_\infty) = 1 \tag{8}$$

The first boundary condition is related to the fact that there is a repulsive potential at the surface, making $\hat{Z}$ vanish.

Since the Laplacian in spherical coordinates is given by

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} \tag{9}$$

we can set $\gamma \equiv r\,\hat{Z}$ and obtain:

$$\frac{\partial \gamma}{\partial \tau} = \frac{\partial^2 \gamma}{\partial r^2} \tag{10}$$

with initial condition $\gamma(0, r) = r$.

This equation can be easily solved performing a Laplace transformation:

$$\bar{\gamma}(s) \equiv L[\gamma(\tau)] = \int_0^\infty \gamma e^{-s\tau} d\tau \tag{11}$$

This allows us to write:

$$L[\partial \gamma / \partial \tau] = s\bar{\gamma}(s) - \gamma(0, r) = s\bar{\gamma}(s) - r \tag{12}$$

and

$$s\bar{\gamma}(s) - r = \frac{\partial^2 \bar{\gamma}}{\partial r^2} \tag{13}$$

The general solution of this equation is

$$\bar{\gamma} = C_1 e^{-\sqrt{s}\,r} + C_2 e^{\sqrt{s}\,r} + \frac{r}{s} \tag{14}$$

We now note that $\hat{Z}(\tau, R_\infty) = 1$ implies that $\gamma(\tau, R_\infty) = R_\infty$ and $\bar{\gamma}(s) = R_\infty/s$. We therefore obtain:

$$C_1 e^{-\sqrt{s}R_\infty} + C_2 e^{\sqrt{s}R_\infty} = 0 \tag{15}$$

Since $e^{\sqrt{s}R_\infty}$ diverges the sum of exponentials is zero only if $C_2 = 0$. If we now consider $\hat{Z}(\tau, R) = 0$, we can write $\gamma(\tau, R) = 0$ and $\bar{\gamma}(s) = 0$ for $r = R$.

We therefore obtain:

$$C_1 e^{-\sqrt{s}R} + \frac{R}{s} = 0 \tag{16}$$

This implies that

$$C_1 = -\frac{R}{s} e^{\sqrt{s}R} \tag{17}$$

We can rewrite $\bar{\gamma}$ as

$$\bar{\gamma} = \frac{r}{s} - \frac{R}{s} e^{-\sqrt{s}(r-R)} \tag{18}$$

The inverse Lapace transform of $\bar{\gamma}$ is

$$\bar{\gamma} = r - R \operatorname{erfc}\left(\frac{r-R}{2\sqrt{\tau}}\right) \tag{19}$$

and

$$\hat{Z} = 1 - \frac{R}{r} \operatorname{erfc}\left(\frac{r-R}{2\sqrt{\tau}}\right) \tag{20}$$

The function $\hat{Z}(\tau, r)$ describes the depletion of polymer concentration with respect to the bulk solution ($\hat{Z} = 1$ at $r = R_\infty$). The excluded volume can be computed by

$$V_{ex} = \frac{4}{3}\pi R^3 + \int_1^\infty 4\pi r^2 (1 - \hat{Z})\,dr \tag{21}$$

This equation can be compared with that for an equivalent hard sphere:

$$V_{ex} = \frac{4}{3}\pi R^3 (1 + q)^3 \tag{22}$$

For the chain, we can write:

$$V_{ex} = \frac{4}{3}\pi R^3 + 4\pi R \int_1^\infty r \, \mathrm{erfc}\left(\frac{r-R}{2\sqrt{\tau}}\right) dr \qquad (23)$$

We now note that

$$\int \mathrm{erfc}(x)dx = x\mathrm{erfc}(x) - \frac{e^{-x^2}}{\sqrt{\pi}} \qquad (24)$$

If we set $x = (r-R)/(2\sqrt{\tau})$, we obtain:

$$\int_1^\infty r \, \mathrm{erfc}\left(\frac{r-R}{2\sqrt{\tau}}\right) dr = 2R\sqrt{\tau} \int_0^\infty \mathrm{erfc}(x)dx + 4\tau \int_0^\infty x \, \mathrm{erfc}(x)dx \qquad (25)$$

Since

$$\int_0^\infty \mathrm{erfc}(x)dx = \frac{1}{\sqrt{\pi}} \qquad (26)$$

$$\int_0^\infty x \, \mathrm{erfc}(x)dx = \left[x\left(x\,\mathrm{erfc}(x) - \frac{e^{-x^2}}{\sqrt{\pi}}\right)\right]_0^\infty - \int_0^\infty \left(x\,\mathrm{erfc}(x) - \frac{e^{-x^2}}{\sqrt{\pi}}\right)dx$$

$$= -\int_0^\infty x \, \mathrm{erfc}(x)dx + \frac{1}{2} \qquad (27)$$

or

$$\int_0^\infty x \, \mathrm{erfc}(x)dx = \frac{1}{4}$$

We therefore obtain

$$\int_1^\infty r \, \mathrm{erfc}\left(\frac{r-R}{2\sqrt{\tau}}\right) dr = 2R\frac{\sqrt{\tau}}{\sqrt{\pi}} + \tau \qquad (28)$$

The excluded volume is then given by

$$V_{ex} = \frac{4}{3}\pi R^3 + 4\pi R \left(2R\frac{\sqrt{\tau}}{\sqrt{\pi}} + \tau\right) \qquad (29)$$

We then determine the normalized profile of monomer density, $f(\tau, r)$. The monomer density at r is

$$f(\tau,r) = \frac{1}{\tau} \int_0^\tau \hat{Z}(\tau - \tau',r)\,\hat{Z}(\tau',r)\mathrm{d}\tau' \tag{30}$$

To understand this expression, we consider a short chain made of three discrete segments (with $\tau = 3$) and examine how we should determine the monomer density. At r, we should consider that the probability of finding the monomer at r has to be related to the following possible configurations shown on the left figure:
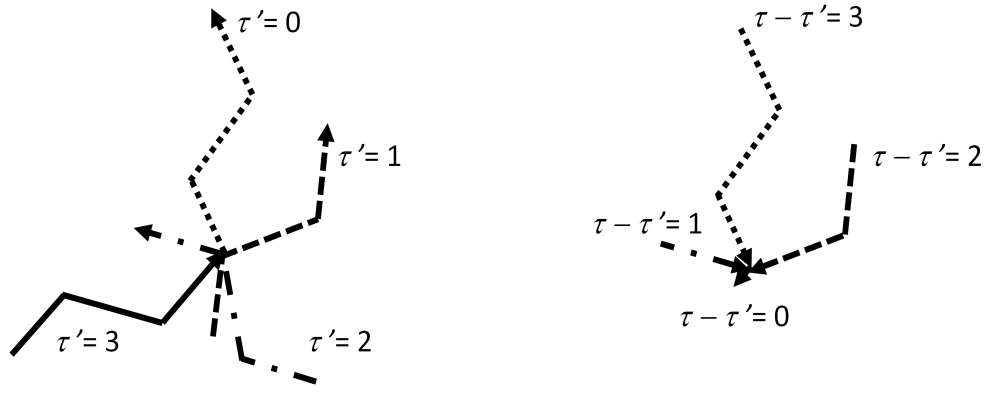
This probability is given by the sum:

$$f(\tau,\tilde{r}) = \sum_{\tau'=0}^{\tau} w_{\tau'}\hat{Z}(\tau',\tilde{r}) \tag{31}$$

where $w_{\tau'}$ are weights that depend on the probability $\hat{Z}(\tau',\tilde{r})$, where $\tilde{r}$ is the end position of a given chain. We can interpret $\hat{Z}(\tau',\tilde{r})$ as the probability of end positions of a shorter chain starting from $\tilde{r}$, $\hat{Z}(\tau - \tau',\tilde{r})$. This is illustrated in the right figure above. We therefore obtain:

$$f(\tau,\tilde{r}) = \sum_{\tau'=0}^{\tau} \hat{Z}(\tau - \tau',\tilde{r})\hat{Z}(\tau',\tilde{r}) \tag{32}$$

The monomer density can be calculated from

$$f(\tau,r) = \frac{1}{\tau}\int_0^\tau \left[1 - \frac{R}{r}\,\text{erfc}\left(\frac{r-R}{2\sqrt{\tau-\tau'}}\right)\right]\left[1 - \frac{R}{r}\,\text{erfc}\left(\frac{r-R}{2\sqrt{\tau'}}\right)\right]d\tau' \qquad (33)$$

The result is

$$f(\tau,r) = \left(\frac{1}{1+2yz}\right)^2 \left\{(2yz)^2 + 4yz\left[\left(\frac{2z}{\sqrt{\pi}}\right)\exp(-z^2) + \text{erf}\,(z) - 2z^2\,\text{erfc}\,(z)\,\right]\right.$$

$$+ 2\text{erf}\,(z) - \text{erf}\,(2z) + \left(\frac{4z}{\sqrt{\pi}}\right)[\exp(-z^2) - \exp(-4z^2)]$$

$$\left.+ 8z^2\left[\text{erfc}\,(2z) - \frac{1}{2}\text{erfc}\,(z)\right]\right\} \qquad (34)$$

where $y \equiv \sqrt{\tau}/R$ and $z \equiv (r-R)/2\sqrt{\tau}$.

Appendix C

Thermodynamic Perturbation Theory

In this appendix, the expression of Helmholtz free energy of protein-water binary system at high temperature (see eq 24) is derived based on the thermodynamic perturbation theory. The incompressible protein-water solution is described by a thermodynamically equivalent compressible one-component fluid (see Chapter 1 Section 2). The protein molecules are modeled using a spherical hard core with pair-wise square-well potentials.

In a N particle system, the potential energy of a pair of proteins, whose centers are separated by a distance of r, is considered as the sum of two parts: a reference potential energy $u_0(r)$ and a perturbation potential energy $u_1(r)$:

$$u(r) = u_0(r) + u_1(r) \tag{1}$$

This corresponds to the following high-temperature expansion of the Helmholtz free energy:

$$\frac{F}{NkT} = \frac{F_0}{NkT} + \frac{F_1}{NkT}\beta + \frac{F_2}{NkT}\beta^2 + \frac{F_3}{NkT}\beta^3 + \cdots \tag{2}$$

where $\beta \equiv 1/k_B T$. For the correct choices of $u_0(r)$ (hard-sphere potential) and $u_1(r)$ (weak attractive potential), $F = F_0$ in the limit of high temperatures.

To determine F_1 and F_2, we define, for the unperturbed system, the probability, $p(N_1, N_2, \ldots, N_i, \ldots)$, of a specific configuration where N_i intermolecular distances are found in the range between r_i and r_{i+1}. If the perturbing potential $u_1(r)$ can be regarded as constant in the small interval $dr = [r_i, r_{i+1}]$, then the partition function can be written as

$$Q = Q_0 \sum_{N_1,N_2,\ldots} \left[p(N_1, N_2, \ldots, N_i, \ldots) \exp(-\beta \sum_i N_i u_1(r_i)) \right]$$

$$= Q_0 \langle \exp(-\beta \sum_i N_i u_1(r_i)) \rangle, \tag{3}$$

where the Q_0 is the partition function of N hard spheres, and the angular parentheses stands for the average over all configurations of the unperturbed system.

The distribution function $p(N_1, N_2, \ldots, N_i, \ldots)$ can be expressed in terms of moments. If we consider only the first and the second moment, we obtain:

$$Q = Q_0[\exp(-\beta \sum_i \langle N_i \rangle u_1(r_i))]\langle \exp(-\beta \sum_i (N_i - \langle N_i \rangle)u_1(r_i)) \rangle$$

$$\frac{F - F_0}{kT} = \beta \sum_i \langle N_i \rangle u_1(r_i) - \frac{1}{2}\beta^2 \sum_{ij} (\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle)u_1(r_i) u_1(r_j) + \cdots \tag{4}$$

The higher order terms are expected to be less important at high densities because the distribution is going to look more like a Gaussian. The first momentum is directly related to the radial distribution function $g_0(r_i)$:

$$\langle N_i \rangle = \frac{1}{2} N \rho g_0(r_i)(4\pi r_i^2)dr \tag{5}$$

This implies that

$$\frac{F_1}{NkT} = \frac{1}{2}\rho \int_0^\infty u_1(r)g_0(r)(4\pi r^2)dr \tag{6}$$

We now want to determine $\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle$. This is based on a macroscopic approximation. We assume that the shells are macroscopic and can be connected to the macroscopic compressibility. For $i = j$:

$$\langle N_i^2 \rangle - \langle N_i \rangle^2 = \frac{\langle N_i \rangle}{Z} = \langle N_i \rangle kT \left(\frac{d\rho}{dP}\right)_0 \tag{7}$$

For $i \neq j$, the $\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle$ expression describes the correlation between particles in different shells. Henderson and Barker assumed that no correlation exists: $\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle = 0$. Later Zhang proposed to estimate the correlation between two consecutive shells by considering the correlation coefficient:

$$\varphi = \frac{\langle N_i N_{i+1} \rangle - \langle N_i \rangle \langle N_{i+1} \rangle}{\sqrt{\langle N_i^2 \rangle - \langle N_i \rangle^2} \sqrt{\langle N_{i+1}^2 \rangle - \langle N_{i+1} \rangle^2}} \tag{8}$$

This correlation coefficient should be close to 1 when the fluid has condensed (i.e. at the volume fraction: $\phi_{con} = 0.493$). We also expect that $\varphi \sim \phi^2$, where ϕ is the volume fraction, because φ must be related to the concentrations of particles in two consecutive layers. These observations allowed Zhang to derive: $\varphi \approx (\phi/\phi_{con})^2$.

It is also a reasonable assumption to say that, for two consecutive shells, we have:

$$\sqrt{\langle N_i^2 \rangle - \langle N_i \rangle^2} \approx \sqrt{\langle N_{i+1}^2 \rangle - \langle N_{i+1} \rangle^2} \text{ and } u_1(r_i) \approx u_1(r_{i+1}).$$

We therefore obtain:

$$\langle N_i N_{i+1} \rangle - \langle N_i \rangle \langle N_{i+1} \rangle \approx (\phi/\phi_{con})^2 (\langle N_i^2 \rangle - \langle N_i \rangle^2) \tag{9}$$

and

$$\sum_{ij} (\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle) u_1(r_i) u_1(r_j)$$

$$= \sum_i (\langle N_i^2 \rangle - \langle N_i \rangle^2) u_1(r_i)^2 + 2 \sum_i (\langle N_i N_{i+1} \rangle - \langle N_i \rangle \langle N_{i+1} \rangle) u_1(r_i) u_1(r_{i+1})$$

$$= \left(1 + 2\left(\frac{\phi}{\phi_{con}}\right)^2\right) \sum_i (\langle N_i^2 \rangle - \langle N_i \rangle^2) u_1(r_i)^2$$

$$= \left(1 + 2\left(\frac{\phi}{\phi_{con}}\right)^2\right) \sum_i \langle N_i \rangle kT \left(\frac{d\rho}{dP}\right)_0 u_1(r_i)^2 \tag{10}$$

Therefore,

$$\sum_{ij}(\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle)u_1(r_i)\,u_1(r_j)$$

$$= \frac{1}{2}\rho N k T \left(1 + 2\left(\frac{\phi}{\phi_{con}}\right)^2\right)\left(\frac{\mathrm{d}\rho}{\mathrm{d}P}\right)_0 \int_0^\infty u_1(r)^2 g_0(r)(4\pi r^2)\mathrm{d}r \tag{11}$$

and

$$\frac{F_2}{NkT} = -\frac{1}{4}\rho k T \left(1 + 2\left(\frac{\phi}{\phi_{con}}\right)^2\right)\left(\frac{\mathrm{d}\rho}{\mathrm{d}P}\right)_0 \int_0^\infty u_1(r)^2 g_0(r)(4\pi r^2)\mathrm{d}r \tag{12}$$

Higher-order terms can be obtained as proposed by Baker and Henderson:

$$F_i = \frac{F_1}{i!}\left(\frac{2F_2}{F_1}\right)^{i-1} \tag{13}$$

By setting $= \beta(1 + 2(\phi/\phi_{con})^2)^{-1}(\mathrm{d}P/\mathrm{d}\rho)_0$, we can summarize the following results:

$$\frac{F_1}{NkT} = 2\pi\rho \int_0^\infty u_1(r)g_0(r)r^2\mathrm{d}r$$

$$\frac{F_2}{NkT} = -\pi\rho\xi^{-1} \int_0^\infty u_1(r)^2 g_0(r)r^2\mathrm{d}r \tag{14}$$

For a square-well potential

$$u_1(r_i) = \begin{cases} -\varepsilon, & for\ \sigma \le r_i \le \lambda\sigma \\ 0, & for\ r_i > \lambda\sigma \end{cases}$$

Define $R \equiv r_i/\sigma$ and $\phi = \pi\sigma^3\rho/6$, we can write:

$$\frac{F_1}{NkT} = -2\pi\sigma^3\rho\varepsilon \int_1^\lambda g_0(R)R^2\mathrm{d}R = -12\phi\varepsilon \int_1^\lambda g_0(R)R^2\mathrm{d}R$$

$$\frac{F_2}{NkT} = -\pi\sigma^3\rho\xi^{-1}\varepsilon^2 \int_1^\lambda g_0(R)R^2\mathrm{d}R = -6\phi\xi^{-1}\varepsilon^2 \int_1^\lambda g_0(R)R^2\mathrm{d}R$$

$$\frac{F_3}{NkT} = -\frac{1}{3}\pi\sigma^3\rho\xi^{-2}\varepsilon^3 \int_1^\lambda g_0(R)R^2\mathrm{d}R = -2\phi\xi^{-2}\varepsilon^3 \int_1^\lambda g_0(R)R^2\mathrm{d}R$$

$$\frac{F_4}{NkT} = -\frac{1}{12}\pi\sigma^3\rho\xi^{-3}\varepsilon^4 \int_1^\lambda g_0(R)R^2 dR = -\frac{1}{2}\phi\xi^{-3}\varepsilon^4 \int_1^\lambda g_0(R)R^2 dR \qquad (15)$$

and so on.

This implies that

$$\frac{F - F_0}{NkT} = -12\phi\xi \sum_{i=1}^{\infty} \frac{1}{i!}\left(\frac{\varepsilon\beta}{\xi}\right)^i \int_1^\lambda g_0(R)R^2 dR$$

which can be compressed into:

$$\frac{F - F_0}{NkT} = -12\phi\xi \left(\exp\left(\frac{\varepsilon\beta}{\xi}\right) - 1\right) \int_1^\lambda g_0(R)R^2 dR \qquad (16)$$

We will now consider the $g_0(R)$ for hard spheres. We will consider the following series expansion:

$$g_0(R) = \frac{1}{\phi}\sum_{i=1}^{\infty} \gamma_i(R)\left(\frac{\phi}{1-\phi}\right)^i \qquad (17)$$

Let $\eta = \phi/(1-\phi)$, we can then write:

$$\int_1^\lambda g_0(R)R^2 dR = \frac{1}{\phi}\sum_{i=1}^{\infty} \eta^i \int_1^\lambda \gamma_i(R)R^2 dR \qquad (18)$$

We shall focus in the range $\lambda = 1 - 2$ because this is the relevant range for proteins. In this range, we just need to retain the first four terms:

$$g_0(R) = \frac{1}{\phi}[\gamma_1(R)\eta + \gamma_2(R)\eta^2 + \gamma_3(R)\eta^3 + \gamma_4(R)\eta^4] \qquad (19)$$

$$\int_1^\lambda g_0(R)R^2\,\mathrm{d}R = \frac{1}{\phi}\left[\eta\int_1^\lambda \gamma_1(R)R^2\,\mathrm{d}R + \eta^2\int_1^\lambda \gamma_2(R)R^2\,\mathrm{d}R + \eta^3\int_1^\lambda \gamma_3(R)R^2\,\mathrm{d}R\right.$$

$$\left. + \eta^4\int_1^\lambda \gamma_4(R)R^2\,\mathrm{d}R\right] \tag{20}$$

where

$$\gamma_1(R) = 1$$

$$\gamma_2(R) = 7 - 6R + 0.5R^3$$

$$\gamma_3(R) = 21.019724 - 60.405915R^{-1} + 39.886191R^{-2}$$

$$\gamma_4(R) = -9.100022 + 13.686797R + 4.586775R^2$$

$$\int_1^\lambda \gamma_1(R)R^2\,\mathrm{d}R = \frac{1}{3}(\lambda^3 - 1)$$

$$\int_1^\lambda \gamma_2(R)R^2\,\mathrm{d}R = \frac{1}{12}\lambda^6 - \frac{3}{2}\lambda^4 + \frac{7}{3}\lambda^3 - \frac{11}{12}$$

$$\int_1^\lambda \gamma_3(R)R^2\,\mathrm{d}R = 7.006575\lambda^3 - 30.202958\lambda^2 + 39.886192\lambda - 16.689809$$

$$\int_1^\lambda \gamma_4(R)R^2\,\mathrm{d}R = -0.917355\lambda^5 + 3.421699\lambda^4 - 3.033341\lambda^3 + 0.528997$$

Clearly those integrals can be directly determined provided that λ is know. We shall refer to them as Γ_1, Γ_2, Γ_3, Γ_4 and write:

$$\frac{F - F_0}{NkT} = -12\xi\left(\exp\left(\frac{\varepsilon\beta}{\xi}\right) - 1\right)(\Gamma_1\eta + \Gamma_2\eta^2 + \Gamma_3\eta^3 + \Gamma_4\eta^4 \tag{21}$$

We also observe that since $P/kT = \rho(1 + \phi + \phi^2 - \phi^3)/(1 - \phi)^3$, we have

$$\frac{\partial}{\partial\rho}\frac{P}{kT} = \frac{1 + 4\phi + 4\phi^2 - 4\phi^3 + \phi^4}{(1 - \phi)^4} \tag{22}$$

This implies that

$$\xi = \frac{1 + 4\phi + 4\phi^2 - 4\phi^3 + \phi^4}{(1 + 2(\phi/\phi_{con})^2)(1 - \phi)^4} \tag{23}$$

In terms of the reduced expression employed in our model:

$$\hat{f}(c_1, T) - \hat{f}^{(0)}(c_1) = -12c_1\xi \left(\exp\left(\frac{\varepsilon\beta}{\xi}\right) - 1 \right) (\Gamma_1\eta + \Gamma_2\eta^2 + \Gamma_3\eta^3 + \Gamma_4\eta^4) \tag{24}$$

Appendix D

Numerical Method for Constructing Phase Diagram

The following program is written with Matlab R2007a for numerically solving the equilibrium conditions for LLPS of protein-polymer-buffer ternary system. At a given phase separation temperature $T=T_{ph}$ and polymer chemical potential $\hat{\mu}_2$, the protein concentrations in the two coexisting phases c_1^I and c_1^{II} are numerically solved from the equilibrium conditions $\hat{\mu}_1^I(c_1^I, \hat{\mu}_2, T) = \hat{\mu}_1^{II}(c_1^{II}, \hat{\mu}_2, T)$ and $\widehat{\Pi}(c_1^I, \hat{\mu}_2, T) = \widehat{\Pi}(c_1^{II}, \hat{\mu}_2, T)$. Then the c_2^I and c_2^{II} are calculated from $c_2 = \alpha(c_1)\exp(\hat{\mu}_2)$, and the LLPS phase diagram is constructed.

1. Flowchart

At given T and $\hat{\mu}_2$, calculate $\widehat{\Pi}$ (c_1). Was a minimum $\widehat{\Pi}$ can be found? — No → Decrease T

Yes ↓

Assign a seed value to c_1^I

Calculate $\hat{\mu}_1$ with c_1^I and solve c_1^{II} from $\hat{\mu}_1$ ← $c_1^I = c_1^{I'}$

Calculate $\widehat{\Pi}$ with c_1^{II} and solve $c_1^{I'}$ from $\widehat{\Pi}$

If $(c_1^{I'} - c_1^I)/c_1^I < 10^{-5}$? — No →

Yes ↓

Report c_1^I, c_1^{II}, T, $\hat{\mu}_2$
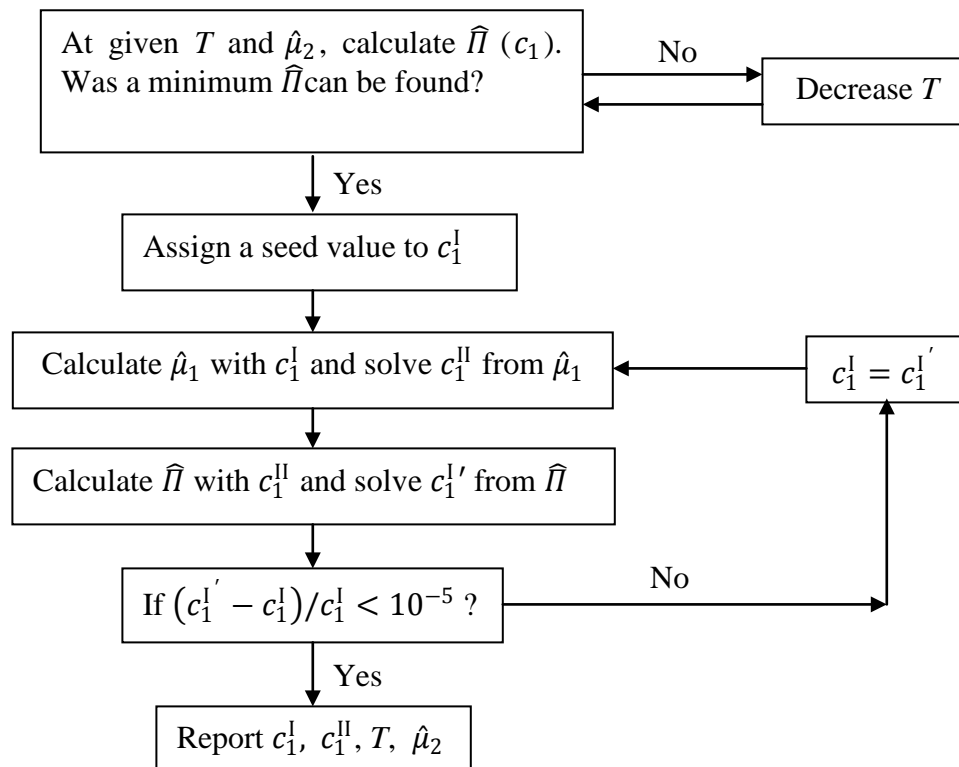
2. Main Program "PHASED.M"

```
% calculate phase diagram protein-polymer mixtures
T=[]; C2=[]; C1=[]; C2=[]; M=[]; B=[]; sC1=[]; F=[]; ddU=[]; dp=[];
pC1I=[]; pC1II=[]; pC1a=[]; pC1b=[]; uC1a=[]; uC1b=[]; fu1=[];
U1=[]; P=[]; a=[]; F=[]; MF=[]; UHS=[]; Uid=[]; Ue=[]; fe=[]; fHS=[];
U1b=[]; Pb=[]; T=[]; C1I=[]; C1II=[]; rC2=[]; alpha=[]; fp=[]; Tph=[];
h=[]; dh=[]; g=[]; dg=[]; en=[]; den=[]; fep=[]; fep2=[]; R=[]; eC1=[];
%_____
%Input data
R=8.31447;
MW1=69000;
MW2=1450;
v=0.00071;      % (L/g) protein specific volume
q=0.6;       % ratio polymer/protein size
e=975;
de=0/R;      % (J/mol) depth of well
l=1.7;       % range of interactions

Tmin=(e/R)/0.45;          %initial temperature

Nmax=10000;
toll=0.00001;

eC1=[50 222 222 222 222];

nC2=5;
C2min=40;
C2max=90;       %g/L
rC2=linspace(C2min,C2max,nC2)

q1=3*q+3*q^2+q^3;
q2=4.5*q^2+3*q^3;
q3=3*q^3;

g1=(l^3-1)/3;
g2=2.333333333333*l^3-1.50000*l^4+0.0833333333333*l^6-
0.9166666666666667;
g3=39.8861910000*l-30.202957500*l^2+7.006574666666667*l^3-
16.68980816666667;
g4=-3.033340666666667*l^3+3.421699250000000*l^4-
0.9173550000000000*l^5+.5289964166666667;
kc=0.493^(-2);

%F=0.15*pi;
%MF=F/(1-F);
%R1=linspace(1,1.3,1000);
%ig1=1;
%ig2=7-6*R1+0.5*R1.^3;
%ig3=21.019724-60.405915./R1+39.886191./(R1.^2);
%ig4=-9.100022+13.686797*R1-4.586775*R1.^2;
%ig=(1/F)*(ig1*MF+ig2*MF^2+ig3*MF^3+ig4*MF^4);
%plot(R1,ig),grid,zoom
%pause
```

```
%R=linspace(1,l,8100);
%ig1=sum(R.^2)*(l-1)/8100
%ig2=sum((7-6*R+0.5*R.^3).*R.^2)*(l-1)/8100
%ig3=sum((21.019724-60.405915./R+39.886191./R.^2).*R.^2)*(l-1)/8100
%ig4=sum((-9.100022+13.686797*R-4.866775*R.^2).*R.^2)*(l-1)/8100


%pause
%_____


%_____
% functions to evaluate. They depend on iC1, iT, iU2
%F='v*iC1*MW1';
%MF='eval(F)/(1-eval(F))';
%a='(1-eval(F))*exp(-q1*eval(MF)-q2*eval(MF)^2-q3*eval(MF)^3)';
%da='(    -eval(a)/(1-eval(F))      *     ( 1+
(q1+2*q2*eval(MF)+3*q3*eval(MF)^2)    / (1-eval(F))   )   )*(v*MW1)';

% modeling hard-core part: Carnahan-Staling HS

%UHS='eval(F)*(8-9*eval(F)+3*eval(F)^2)/(1-eval(F))^3';
%Uid='log(iC1)';
%fHS='eval(F)*(4-3*eval(F))/(1-eval(F))^2';

% Ue='-2*eval(F)*(4*e*l^3)/(R*iT)';      %this is the mean-field part to
be changed
%fe2='-eval(F)*(4*e*l^3)/(R*iT)';


% modeling attraction part: perturbation theory

%h=' (1+4*eval(F)+4*eval(F)^2-4*eval(F)^3+eval(F)^4) / (1-eval(F))^4  /
(1+2*kc*eval(F)^2)';
%dh='4* ( (2+5*eval(F)-eval(F)^2) - kc*eval(F)*(1-eval(F)-10*eval(F)^2-
6*eval(F)^3+5*eval(F)^4-eval(F)^5))  /   ( (1+2*kc*eval(F)^2)^2 * (1-
eval(F))^5 )';

%en='exp( e/ ((R*iT)*eval(h)) )-1';
%den='( -e /((R*iT)*eval(h)^2) ) * exp( e/((R*iT)*eval(h))) *eval(dh)';

%g='g1*eval(MF)+g2*eval(MF)^2+g3*eval(MF)^3+g4*eval(MF)^4';
%dg='(g1+2*g2*eval(MF)+3*g3*eval(MF)^2+4*g4*eval(MF)^3) / (1-
eval(F))^2';

%fe='-12*eval(h)*eval(en)*eval(g)';
%Ue='-12* (  eval(h)*eval(en)*eval(g)   +eval(F) *
(eval(dh)*eval(en)*eval(g)
+eval(h)*eval(den)*eval(g)+eval(h)*eval(en)*eval(dg) ))';



%P='iC1*(1+eval(UHS)+eval(Ue)-eval(fHS)-eval(fe)-
eval(da)*exp(iU2))+eval(a)*exp(iU2)';
%U1='eval(Uid)+eval(UHS)+eval(Ue)-eval(da)*exp(iU2)';
%_____
```

```
%  _____  plot stuff (to remove later)
    points=1000;
    iU2=-inf;
    C1=linspace(0.0001,0.45/v,points);
    iT=Tmin;
    for i=1:points
    iT=Tmin;
    iC1=C1(i)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
%
%    vh(i)=h;
%    ven(i)=en;
%    vg(i)=g;
%    fem(i)=eval(fe2);
%    fep(i)=fe;
%    free(i)=(log(iC1/exp(1))+eval(fHS)+eval(fe));
    fu1(i)=U1;
%    fp(i)=P/iC1;
%    iT=Tmin*(1+.0001);
%
[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
%    fep2(i)=fe;
%    uen(i)=-(fep2(i)-fep(i))/(iT-Tmin)*Tmin;
    end
%df=diff(free)./diff(C1)*MW1;

    plot(C1*v*6/pi,fu1),zoom,grid
    pause
%    plot(C1*v*6/pi,fp),zoom,grid
%    pause
%    plot(C1*v,uen),zoom,grid
%    pause

%_____

for m=1:nC2
iU2=log(rC2(m)/MW2);
%for m=1:nC2
% This cycle is responsible for the increments of C2

T(1)=Tmin;
for k=1:Nmax
% This cycle is responsible for the increments of T

%_____
% find initial seed for C1II and check T
% we look for a minimum in the pressure behavior starting from high
concentrations and gradually lowering the value
% if the minimum is not found we need to decrease Tmin

for it1=1:Nmax
pC1I(1)=0;
pC1II(1)=0.45/v;
```

```
dC=pC1II(1)/100;
for it2=1:100
    iT=T(k);
    vC1II=pC1II(1);
    iC1=pC1II(1)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    p2=P;
    iC1=(pC1II(1)-dC)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    p1=P;
    dp=p2-p1;
    if dp<0 break
    end
    pC1II(1)=iC1*MW1;
end
    if it2==100 & k==1
        'minimum not found! lower Tmin!'
    end
    if dp<0
        break
    end
    T(k)=(T(k)+T(k-1))/2;
end
% if the minimu is found the program chooses the C1 immediately to the
right of the minimum. This is a seed for C1II.
%Note that if the minimum it is not found and k>1, then the program
automatically lowers T.
%_____

for i=1:Nmax
% This cycles starts from the C1II seed and compute the chemical
potential at C1II, which should be the same at C1I.
% Starting from two very low concentrations seeds of C1I, it determines
the value of C1I from the chemical potential curve.
    iC1=pC1II(i)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    vU1=U1;
    uC1a(1)=dC/10000;
    uC1b(1)=2*dC;
    iC1=uC1a(1)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    U1a(1)=U1;
    iC1=uC1b(1)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    U1b(1)=U1;
```

```
    for j=1:Nmax
% this is the actual cycle responsible for C1I obtainment
        slope=(U1b(j)-U1a(j))/(uC1b(j)-uC1a(j));
        uC1a(j+1)=uC1b(j);
        uC1b(j+1)=uC1b(j)+(vU1-U1b(j))/slope;
        U1a(j+1)=U1b(j);
        iC1=uC1b(j+1)/MW1;

        err=abs(uC1b(j+1)-uC1b(j))/abs(uC1b(j+1));
        if  err<toll
            pC1I(i+1)=uC1b(j+1);
            break
        end
        if  uC1b(j+1)<0
            'negative'
            break
        end

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
        U1b(j+1)=U1;
    end
% the obtained value of C1I is then plugged in the pressure equation.
The value of the pressure is then used to obtain a new estimation of
C1II.
    iC1=pC1I(i+1)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    vP=P;
    pC1a(1)=0.4/v;
    pC1b(1)=0.45/v;
    iC1=pC1a(1)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    Pa(1)=P;
    iC1=pC1b(1)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    Pb(1)=P;

    for j=1:Nmax

        slope=(Pb(j)-Pa(j))/(pC1b(j)-pC1a(j));
        pC1a(j+1)=pC1b(j);
        pC1b(j+1)=pC1b(j)+(vP-Pb(j))/slope;
        Pa(j+1)=Pb(j);
        iC1=pC1b(j+1)/MW1;

        err=abs(pC1b(j+1)-pC1b(j))/pC1b(j+1);
        if  err<toll
            pC1II(i+1)=pC1b(j+1);
            break
        end
```

```
[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
        Pb(j+1)=P;
    end

    err=max([abs(pC1I(i+1)-pC1I(i))/pC1I(i+1) abs(pC1II(i+1)-
pC1II(i))/pC1II(i+1)]);
    if  err<toll
            C1I(k)=pC1I(i+1);
            C1II(k)=pC1II(i+1);
            break
    end
end
    if k>1 & (T(k)-T(k-1))/T(k)<1e-3
        Tc(m)=T(k);
        Cc(m)=(C1I(k)+C1II(k))/2;
        break
    end
T(k+1)=(1+.05/(k^0.5))*T(k);
end

Y=[T, Tc(m), fliplr(T)];
X=[C1I, Cc(m), fliplr(C1II)];
for i=1:length(eC1)
    iC1=eC1(i)/MW1;

[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2);
    C2(m,i)=MW2*a*exp(iU2)
    Tph(m,i)=interp1(X,Y,eC1(i))
end
m

T=[]; X=[]; Y=[]; C1I=[]; C1II=[];

end

%plot(C2(:,1),Tph(:,1),C2(:,2),Tph(:,2),C2(:,3),Tph(:,3),C2(:,4),Tph(:,
4),C2(:,5),Tph(:,5),[79  75  72  68  64  58  52], [285.1 281.05 277.4
274.15  270.55  265.65  262.15], 'o',[79    74  70  65  61  54],
[290.15   284.8   280.25  275.05  271.15  266], 'o',[79   74  69  63
57  52  47], [308.85   301.85  292.6   285 277.25  271.2   267.3],
'o',[45 43  40  35  31], [299.15   294.2   287.75  279.55  270.2],
'o',[36 30  26  23  20], [298.8 286.95  274.55  268.3   264.2],'o')

plot(C2(:,1),Tph(:,1)), grid, zoom

C2(:,1)
Tph(:,1)

%e/(R*Tc)
%Cc*v
%plot(v*C1I,T,v*C1II,T, v*Cc, Tc,'o'), zoom, grid
```

3. Subfunction "PHASED.M"

```
%Subfunction for calculating the chemical potential of protein

function
[F,MF,a,da,Uid,UHS,fHS,h,dh,en,den,g,dg,fe,Ue,P,U1]=pchem(iC1,iT,v,e,de
,l,kc,R,MW1,MW2,g1,g2,g3,g4,q1,q2,q3,iU2)

%functions_____

F=v*iC1*MW1;
MF=F/(1-F);
a=(1-F)*exp(-q1*MF-q2*MF^2-q3*MF^3);
da=(-a/(1-F) * (  1+ (q1+2*q2*MF+3*q3*MF^2)  / (1-F) ))*(v*MW1);
Uid=log(iC1);
UHS=F*(8-9*F+3*F^2)/(1-F)^3;
fHS=F*(4-3*F)/(1-F)^2;
h=(1+4*F+4*F^2-4*F^3+F^4) / (1-F)^4  / (1+2*kc*F^2)';
dh=4* ( (2+5*F-F^2) - kc*F*(1-F-10*F^2-6*F^3+5*F^4-F^5))  /
( (1+2*kc*F^2)^2 * (1-F)^5 );
en=exp( (e+de/iT)/ ((R*iT)*h) )-1;
den=( -(e+de/iT) /((R*iT)*h^2) ) * exp( (e+de/iT)/((R*iT)*h)) *dh;
g=g1*MF+g2*MF^2+g3*MF^3+g4*MF^4;
dg=(g1+2*g2*MF+3*g3*MF^2+4*g4*MF^3) / (1-F)^2;
fe=-12*h*en*g;
Ue=-12* (  h*en*g   +F * (dh*en*g  +h*den*g+h*en*dg ));
P=iC1*(1+UHS+Ue-fHS-fe-da*exp(iU2))+a*exp(iU2);
U1=Uid+UHS+Ue-da*exp(iU2);
%_____
```

Appendix E

The Scaled Particle Theory on the Fluid

with Square-Well Potential

In this appendix, we will use scaled particle theory to derive the free energy change of inserting a solute particle into a fluid where the fluid particles interact with a square-well potential. This yields an expression for the free-volume fraction that depends on temperature (see eq 13).

In appendix A, we have shown the work, W, for creating a cavity with radius R, i.e, inserting a AO sphere with radius R, in a fluid:

$$\beta W(q) = -\ln(1 - \phi) + 3(q + q^2)\eta + \frac{9}{2}q\eta^2 + Z\phi q^3 \tag{1}$$

We now need to consider the expression of Z in the case of a fluid with attractive interactions. Because we consider a fluid with weak attraction, we can apply the first order perturbation theory as described in appendix E. The free energy of the attractive fluid is expressed by sum of the hard-sphere term and a perturbation term,

$$\frac{\hat{f}}{\rho} = \frac{\hat{f}_0}{\rho} - 12\beta\phi\varepsilon \int_1^\lambda g_0(r)r^2 \mathrm{d}r \tag{2}$$

where $g_0(r)$ is the radial distribution function of hard sphere given by eq 19 in appendix E.

From the basic thermodynamic relation: $F = N\mu - PV$, we can obtain: $\hat{f} = \frac{\rho \partial \hat{f}}{\partial \rho} - \beta P$. This allows us to express compressibility Z as,

$$Z = \rho \frac{\partial(\hat{f}/\rho)}{\partial \rho} \tag{3}$$

Therefore, by combining eq 2 and eq 3, we obtain:

$$Z = Z_0 - 12\beta\varepsilon\phi \frac{\partial\left(\phi \int_1^\lambda g_0(r)r^2 dr\right)}{\partial\phi} \tag{4}$$

where Z_0 is the compressibility of hard spheres(see eq D-20).

Then substitute eq 4 into eq 1, we obtain:

$$\beta W(q) = -\ln(1-\phi) + (3q + 3q^2 + q^3)\eta + \left(\frac{9}{2}q + 3q^3\right)\eta^2 + 3q^3\eta^3$$

$$- 12\beta\varepsilon\phi^2 q^3 \frac{\partial\left(\phi \int_1^\lambda g_0(r)r^2 dr\right)}{\partial\phi} \tag{5}$$

We now use W_0 to denote the work for creating the cavity in hard-sphere fluid, i.e. the first four terms in eq 10, and write:

$$\beta W = \beta W_0 - 12\beta\varepsilon\phi^2 q^3 \frac{\partial\left(\phi \int_1^\lambda g_0(r)r^2 dr\right)}{\partial\phi} \tag{6}$$

Within scaled particle theory, the work for creating the cavity is equal to the free energy change of inserting the second solute hard sphere. We thus relate the work to free volume fraction of the AO sphere by,

$$W = \Delta F = -kT\ln(\alpha) \tag{7}$$

In conclusion, we obtain:

$$\frac{\ln(\alpha/\alpha_0)}{\varepsilon} = 12\beta\phi^2 q^3 \frac{\partial\left(\phi \int_1^\lambda g_0(r)r^2 dr\right)}{\partial\phi} \tag{8}$$

Appendix F

LLPS of the Fluid of Square-Well Chain

The increase of T_{ph} with the degree of polymerization has been theoretically investigated using a perturbation method by Banaszak. Banaszak et al. have reported a perturbation theory based on Wertheim's theoretical framework to determine the effect of hard-sphere connectivity on the LLPS phase boundary. We derive the expressions of the Helmholtz free energy and chemical potential for a fluid of square-well chain (see eq 8 and eq 9).

In this model, chain connectivity is introduced by applying thermodynamic perturbation theory to the hard-sphere monomer with diameter σ. The chain-chain attractive interactions, which are responsible for LLPS, were introduced using the square-well potential for the monomer with a well width λ of 1.5 σ, according to the Barker-Henderson perturbation theory.

Based on Wertheim's first-order themodynamic perturbation theory, the compressiblility factor Z for chains consisting of n spheres, the n-mer, can be expressed as,

$$Z_n = nZ_1 + Z_{chain} \tag{1}$$

Where Z_1 is the reference compressibility of the non-bonded sphere (monomer), and Z_{chain} is the perturbation caused by the connectivity(bonding) of the spheres. Z_{chain} is given by

$$Z_{chain} = (1-n)\left(1 + \phi \frac{\partial}{\partial \phi} \ln g(\sigma, \phi)\right) \tag{2}$$

where ϕ is the volume fraction of reference spheres, and $g(\sigma,\phi)$ is the contact number of the spheres, i.e., the value of the radial distribution function of the reference sphere fluid at the distance $r = \sigma$.

The free energy, $\hat{f}_n$, of the n-sphere chain can be expressed in terms of Z_n by combining the basic thermodynamic relations $\hat{f}_n = c_n\hat{\mu}_n - \hat{\Pi}_n$ and $\hat{\mu}_n = \left(\partial\hat{f}_n/\partial c_n\right)_{V,T}$.

$$Z_n = \frac{\hat{\Pi}_n}{c_n} = c_n\frac{\partial\left(\hat{f}_n/c_n\right)}{\partial c_n} \tag{3}$$

$$\hat{f}_n/c_n = \ln(c_n/e) + \int_0^{c_n}\frac{Z_n - 1}{c_n}dc_n \tag{4}$$

where c_n is the molarity of the chain (oligomer). The molarity of monomer is denoted by c_1. According to mass conservation, $c_n=c_1/n$.

From eq 1 and eq 2, we find:

$$\frac{Z_n - 1}{c_n} = n^2\frac{Z_1 - 1}{c_1} + n(1 - n)\frac{\partial}{\partial c_1}\ln g(\sigma,\phi) \tag{5}$$

$$\int_0^{c_n}\frac{Z_n - 1}{c_n}dc_n = n\int_0^{c_1}\frac{Z_1 - 1}{c_1}dc_1 + (1 - n)\int_0^{c_1}\frac{\partial}{\partial c_1}\ln g(\sigma,\phi)\,dc_1 \tag{6}$$

Substituting eq 4 into eq 6, we obtain:

$$\frac{\hat{f}_n}{c_n} - \ln(c_n/e) = n\left(\frac{\hat{f}_1}{c_1} - \ln(c_1/e)\right) + (1 - n)\ln g(\sigma,\phi) \tag{7}$$

Therefore, we finally find:

$$\hat{f}_n = c_n\ln(c_n/e) + \left(\hat{f}_1 - c_1\ln(c_1/e)\right) + \frac{(1 - n)}{n}c_1\ln g(\sigma,\phi) \tag{8}$$

$$\hat{\mu}_n = \ln c_n + n(\hat{\mu}_1 - \ln c_1) + (1 - n)\frac{\partial}{\partial\phi}(\phi\ln g(\sigma,\phi)) \tag{9}$$

According the first-order thermodynamic perturbation theory for monomer (see appendix C), the free energy for monomer is given by:

$$\hat{f}_1 = \hat{f}_1^{(0)} + \frac{\hat{f}_1^{(1)}}{kT}\varepsilon + \frac{\hat{f}_1^{(2)}}{kT}\varepsilon^2 + \mathcal{O}(\varepsilon^2) \tag{10}$$

Since ε is small for weak attractive interactions, only the first three terms of free energy were retained. At $\lambda=1.5$, we have

$$\hat{f}_1^{(0)}/c_1 = \ln\left(\frac{c_1}{e}\right) + \frac{\phi(4-3\phi)}{(1-\phi)^2} \tag{11}$$

$$\hat{f}_1^{(1)}/c_1 = 3.173136\left[1 - \exp\left(\frac{4.5\phi}{\sqrt{2}-\phi}\right) - \frac{4.5}{\sqrt{2}}\phi\right] - 4.974192\phi + 5.134186\phi^2$$

$$\hat{f}_1^{(2)}/c_1 = -0.384466\left[1 - \exp\left(\frac{9.75\phi}{\sqrt{2}-\phi}\right) - \frac{9.75}{\sqrt{2}}\phi\right] - 2.487096\phi - 0.047652\phi^2$$

The monomer chemical potential $\hat{\mu}_1$ is calculated by differentiating eq 10 with respect to c_1.

The radial distribution function of the spheres with square well potential expressed by,

$$g(\sigma,\phi) = g^{(0)}(\sigma,\phi) + \frac{\varepsilon\lambda^3}{kT}g^{(0)}(\lambda\sigma,\phi) + \frac{\varepsilon}{4kT}\frac{\partial\hat{f}_1^{(1)}/c_1}{\partial\phi} \tag{12}$$

where

$$g^{(0)}(\sigma,\phi) = \frac{1-\phi/2}{(1-\phi)^3}$$

$$g^{(0)}(\lambda\sigma,\phi) = 1 + \gamma_1\phi + \gamma_2\phi^2 + \gamma_3\phi^3 + \gamma_4\phi^4$$

At $\lambda=1.5$, we can use: $\gamma_1 = 0.653305$, $\gamma_2 = -1.38146$, $\gamma_3 = -7.58844$, and $\gamma_1 = 8.40166$.

Appendix G

Flowchart of the Monte Carlo Algorithm

In this Appendix we present the flowcharts describing the Monte Carlo algorithm used to calculate free-volume and average gyration radius of polymer in a sphere suspension.

Fig 1 presents the main procedure of the algorithm as follows:

1. The initial close-packing configuration of spheres is set up. The number of spheres is fixed to $N = 1000$. The diameter of sphere, σ, depends on the sphere volume fraction associated with the simulation ($\phi = 0.001\text{-}0.4$). The potential-energy parameters were the normalized well depth $\hat{\varepsilon} = \varepsilon/kT$ and range λ. The variables P and *Sum* respectively accumulate the number of successful addition of polymer into the hard-sphere suspension and the sum of the square of gyration radius for individual polymer configurations. The variable K keeps track of the number of attempts to change protein configuration.

2. A (i^{th}) sphere is chosen. The number of contacts η_1 made by this sphere is calculated by counting the proteins whose centers are within $\lambda\sigma$ from the center of ith sphere. This calculation is done under periodic condition, i.e., the spheres in 26 neighboring unit cube with identical configuration are also taken into account. In order to save computational time, the periodic condition is only applied to the spheres for which the three spatial coordinates are in the ranges from 0 to $\lambda\sigma$ and from 1- $\lambda\sigma$ to 1.

3-5. A position (X, Y, Z) in the box is selected at the distance a (the step length of sphere displacement) away from the center of the ith sphere in a random direction. This position is first selected in the local spherical coordinate system, whose origin is at the center of ith sphere. The radial coordinate equals to a fixed step length of sphere displacement, a, assigned according to the acceptance rate. The longitude angular coordinate φ are randomly selected between 0 and 2π. The latitude angular coordinate θ are randomly selected between 0 and π. The spherical coordinates are converted to the Cartesian coordinates (dX, dY, dZ) with respect to i^{th} sphere. Then the coordinates of the new position in the box, (X, Y, Z) is calculated from (dX, dY, dZ).

6. An attempt is made to move i^{th} sphere to the new position. If there is an overlap with the hard core of any other spheres (under periodic condition), i.e. the distance between their center is smaller than σ , the attempt is considered as unsuccessful and the program returns to step 2 via step 12 shown below.

7.   The number of contacts η_2 made by the sphere at the new position is calculated.

8-10. The Metropolis equilibration algorithm. The potential energy difference, Δ, made by moving the ith sphere is calculated by $\Delta = (\eta_2 - \eta_1)\varepsilon$. If Δ is negative, the condition in step 7 will always be true and the ith sphere is moved to the new position. If Δ is positive, the move is accepted with a probability of exp($-\Delta /kT$).

11. An attempt is made to generate a set of coordinates (x_M, y_M, z_M) for a polymer chain (or AO sphere). The chain contains M_{max} points (atoms). These points are generated in sequence by a cycle described below and shown in Fig 2. A successful attempt leads to the variable P increase by one. The square gyration

radius $R_g{}^2$ of this polymer chain configuration is calculated and added to the variable *Sum*.

12. If the total number of attempts K to change protein configuration is greater than a specified maximum $K_{max}= 10^8$, the program report the free-volume, v, for the polymer in the solution as $v = P/(K- K_0)$, where K_0 is the cut-off for equilibration of the Metropolis algorithm, and the quadratic mean of gyration radius as $R_g^{RMS} = \sqrt{Sum/P}$. The standard deviation of v is given by $d_v = \sqrt{\frac{v(1-v)}{K_p}}$ where K_p is the total number of attempts to generate polymer chain. The standard deviation of R_g^{RMS} is given by $d_{R_g^{RMS}} = \sqrt{\frac{<R_g^4>-<R_g^2>^2}{P}}$.

Figure 1: The main cycle of the Monte Carlo algorithm

1 | Place N spheres in a $1\times1\times1$ box with a close-packing configuration. P=0; Sum=0; Number of attempts K=0.

12

Increase K by one. If K > Kmax the report $v = P/(K\text{-}K_0)$ and $R_g^{RMS} = \sqrt{Sum/P}$

2 | Choose a random integer i between 1 and N. Calculate the number of contacts η_1 of ith spheres.

3 | Generate two random number ϕ between 0 and π and θ between 0 and 2π. (l,θ,φ) define the spherical coordinates of the new position with respect to the center of ith sphere.

4 | Calculate the Cartesian coordinates (dx,dy,dz) from (l,θ,φ). These define the new position with respect to the ith sphere.

5 | Calculate the (X, Y, Z) from (dx,dy,dz). These define the new position in the box.

6 | Does hard core corresponding to position (X, Y, Z) overlap with the hard core of any other sphere (except i^{th}) in the box?

Yes

No

7 | Calculate number of contacts η_2 at position (X, Y, Z), (except i^{th}) in the box

8 | Calculate $\Delta = -\varepsilon(\eta_2\text{-}\eta_1)$. Generate a random number W between 0 and 1

9 | Is exp(-Δ/kT) > W?

No

Yes

10 | Move i^{th} sphere to position X, Y, Z

11 | Generate a set of coordinates for a polymer molecule. Calculate P and $R_g{}^2$

Fig 2-4 illustrates the generation of polymer chain in step 11. In our study, four kinds of polymer models are used, i.e. hard-sphere, linear chain of points with fixed bond length and arbitrary bond angle (Fig 2), linear chain with fixed bond length and bond angle (Fig 3), and linear self-avoiding chain of spheres with two parameters α, the effective atomic diameter, and ε, the trans-gauche transition energy (Fig 4).

The procedures in Fig 2 are as follows:

1. The program first checks whether the number of attemps K to change sphere configuration is greater than a specified minimum $K_0 = 10^6$. If $K \geq K_0$, the program starts to generate polymer chains. If $K < K_0$, the program does not generate polymer chains and continues with the Metropolis equilibration algorithm (steps 8-10). This ensures that the determination of physical properties by simulation is only done after equilibration.

2. The position (x_1, y_1, z_1) is selected at random. The index of points of polymer chain is initialed at one.

3. An attempt is made to generate the first point of polymer chain at (x_1, y_1, z_1). If there is an overlap with the hard core of spheres, the attempt is consider as unsuccessful and the program continues to step 12, leaving P unchanged.

4-7. Increase the index, M, of the points on polymer chain by one. A position (x_M, y_M, z_M) in the box is selected at a distance of l (bond length) away from the preceding point. This is done through selecting the spherical coordinate with respect to last point, and then converting them the Cartesian coordinates in the coordinates system of the box (similar to Step3-5).

6. An attempt is made to generate the next point of the polymer chain at (x_M, y_M, z_M). If there is an overlap with the spheres, the attempt is unsuccessful and the program continues to step 12 in main program, leaving P unchanged.

7. The program checks if the generation of polymer chain is completed. If the index M is smaller than the number of points on the polymer chain M_{max}, the program continues to steps 4-7 to generate more points. If $M \geq M_{max}$, the program finishes generation polymer chain and precedes to step 10.

8. The variable P, which tracks the number of successful polymer configuration, is increased by one. The R_g^2 for this chain is calculated by

$$R_g^2 = \frac{\sum_{M=1}^{M_{max}} (x_M^2 + y_M^2 + z_M^2)}{M_{max}} - \frac{\left(\sum_{M=1}^{M_{max}} x_M\right)^2 + \left(\sum_{M=1}^{M_{max}} y_M\right)^2 + \left(\sum_{M=1}^{M_{max}} y_M\right)^2}{M_{max}^2}$$

and added to the variable *sum*, and return to step 12 in main program.

Figure 2, Step 11: Generate a polymer chain with fixed bond length and arbitrary angle

From step 10 (main cycle)

1 — If $K > K_0$? — No

Yes

2 — $M = 1$; Generate three random number (x_1,y_1,z_1), between 0 and 1. These define the position of a terminal atom of the polymer.

3 — Does the position (x_1,y_1,z_1) overlap with the hard core of any sphere in the box? — Yes

No

4 — Increase M by one.

5 — Generate two random number ϕ between 0 and π and θ between 0 and 2π. (l,θ,φ) define the spherical coordinates of the successive atom of the polymer, using last atom as the origin.

6 — Calculate the Cartesian coordinates (dx,dy,dz) from (l,θ,φ). These define the position of the atom with respect to last atom.

7 — Calculate the (x_M,y_M,z_M) from (dx,dy,dz). These define the position of the atom in the box.

8 — Does the position (x_M,y_M,z_M) overlap with the hard core of any sphere in the box? — Yes

No

9 — If $M < M_{max}$? — No

Yes

10 — Increase P by one. Calculate R_g^2 for this polymer chain. Add R_g^2 to Sum.

To step 12 (main cycle)

The procedures in Fig 3 are as follows:

1-8. The same operations as in Fig A2 are done to generate the first and second points of the polymer chain.

9-12. A position (x_M, y_M, z_M) in the box is selected at a distance of l (the bond length) away from the preceding point. This position is first selected in the local spherical coordinate system, whose origin is at last point and z' axis is in the direction of last segment. The radial coordinate equals to a specified bond length, l. The latitude angular coordinate θ is with a specified value of bond angle, θ_0. The longitude angular coordinate φ is randomly selected between 0 and 2π. The spherical coordinates are converted to the Cartesian coordinates (dx, dy, dz) with respect to last point. Then the coordinates of this position in the coordinate system of the box, (x_M, y_M, z_M) is calculated from (dx, dy, dz).

13. An attempt is made to generate the next point of the polymer chain at (x_M, y_M, z_M). If there is an overlap with the hard core of spheres, the attempt is consider as unsuccessful and the program continues to step 12 in the main program, leaving P unchanged.

14. The program checks if the generation of polymer chain is completed. If the index M is smaller than the number of points on the polymer chain M_{max}, the program continues to steps 9-12 to generate more points. If $M \geq M_{max}$, the program finishes generation polymer chain and precedes to 15.

15. The variable P, which tracks the number of successful polymer configuration, is increased by one. The $R_g{}^2$ for this chain is calculated and added to the variable *sum*, and return to step 12 in main program.

Figure 3, Step 11: Generate a polymer chain with fixed bond length and bond angle

From step 10 (main cycle)

No $\quad$ | 1 |

Yes

| 2 | Generate the first atom

Yes

| 3 |

No

| 4-7 | Generate the second atom

Yes

| 8 |

No

9 | Increase M by one.

10 | Generate one random number ϕ between 0 and 2π. (l,θ_0,φ) define the spherical coordinates of the successive atom of the polymer, using last atom as the origin and last bond as z axis.

11 | Calculate the Cartesian coordinates (dx,dy,dz) from (l,θ_0,φ). These define the position of the atom with respect to last atom.

12 | Calculate the (x_M,y_M,z_M) from (dx,dy,dz). These define the position of the atom in the box.

Yes

13 | Does the position (x_M,y_M,z_M) overlap with the hard core of any sphere in the box?

No

14 | If $M < M_{max}$? $\quad$ No

Yes

15 | Increase P by one. Calculate R_g^2 for this polymer chain. Add R_g^2 to Sum.

To step 12 (main cycle)

The procedures in Fig 4 are as follows:

1-10. The same operations as in Fig 2 are done to generate the first and second atoms of the polymer chain. The only difference is, in this case, the atoms are represented by hard-sphere with diameter a, instead of points.

11-14. A position (x_M, y_M, z_M) in the box is selected at a distance of l (the bond length) away from the center of preceding sphere. This position is first selected in the local spherical coordinate system, whose origin is at center of last sphere and z' axis is in the direction of last segment. The radial coordinate equals to a specified bond length, l. The latitude angular coordinate θ is with a specified value of bond angle, θ_0. The longitude angular coordinate φ' for the second to the last segment is calculated. The longitude angular coordinate φ for the new position is selected between 0 and 2π with a probability of $\exp(-\hat{\epsilon})/(1+\exp(-\hat{\epsilon}))$ within $\varphi' \pm \frac{\pi}{2}$ and of $1/(1+\exp(-\hat{\epsilon}))$ otherwise. The spherical coordinates (l, ϕ_0, φ) are converted to the Cartesian coordinates (dx, dy, dz) with respect to the last sphere. Then the coordinates of this position in the coordinate system of the box, (x_M, y_M, z_M) is calculated from (dx, dy, dz).

15. An attempt is made to generate the next point of the polymer chain at (x_M, y_M, z_M). If there is an overlap with the hard core of any preceding atoms of polymer chain, the attempt is consider as unsuccessful and the program continues to step 2, to regenerate the polymer chain.

16. The program checks if the generation of polymer chain is completed. If the index M is smaller than the number of spheres on the polymer chain M_{max}, the

program continues to steps 9-12 to generate more points. If $M \geq M_{max}$, the program finishes generation polymer chain, and precedes to step 17.

17. Check whether the atoms of the polymer chain overlap with the hard cores of spheres in the box. If there is an overlap, the attempt is consider as unsuccessful and the program continues to step 12 in main program, leaving P unchanged.

18. The variable P, which tracks the number of successful polymer configuration, is increased by one. The R_g^2 for this chain is calculated by and added to the variable *sum*, and return to step 12 in main program.

Note:

(1) In step 15, if the last generated atom overlap with previous atoms in polymer, the generation of polymer chain needs to start over to take into account the self exclusion effect of polymer. However, the sphere configuration should not be changed until a whole polymer chain is generated, because the generation of polymer chain is an independent process of the change of sphere configuration.

(2) We check the overlap of polymer chain with spheres after the whole chain is generated. If we check the overlap after each atom is generated as described in the other polymer models, the simulation will be faster, but the self exclusion of polymer chain will be coupled with the mutual exclusion between polymer and sphere. This would distort the simulation result and lead to a free-volume fraction value larger than the real value.

Figure 4, Step 11: Generate a self-avoiding polymer chain with two parameters (α, ε)

From step 10 (main cycle)

No		

1

Yes

2

3-6

7-10

11 | Increase M by one.

12 | Calculate the local longitude of the second to the last bond, φ'. Generate one number φ between 0 and 2π with a probability of $\exp(-\varepsilon/kT)/(1+\exp(-\varepsilon/kT))$ within $\varphi' \pm \frac{\pi}{2}$ and of $1/(1+\exp(-\varepsilon/kT))$ otherwise. (l, θ_0, φ) define the spherical coordinates of the successive atom of the polymer, using last atom as the origin and last bond as z axis.

13 | Calculate the Cartesian coordinates (dx,dy,dz) from (l, θ_0, φ). These define the position of the atom with respect to the preceding atom.

14 | Calculate the (x_M, y_M, z_M) from (dx,dy,dz). These define the position of the atom in the box.

15 | Does the position (x_M, y_M, z_M) overlap with any existing polymer atoms?   Yes

No

16 | If $M < M_{max}$?   No

Yes

17 | Does the position (x_i, y_i, z_i), where i={1, 2,··· M_{max}}, overlap with the hard core of any sphere or existing polymer atoms?

Yes

18 | Increase P by one. Calculate $R_g{}^2$ for this polymer chain. Add $R_g{}^2$ to Sum.

To step 12 (main cycle)

Appendix H

C Program for Monte Carlo Simulation on Free-Volume

The following program is for determination of free-volume and average gyration radius for polymer in the protein solution. The polymer is modeled by a self-avoiding chain with a trans-gauche conformation transition potential energy. The programs for other polymer models are similar and not shown. The random number generator used in the program is a Mersenne Twister random number generator based on the code written by Makoto Matsumoto, Takuji Nishimura, and Shawn Cokus.

1. The main function "freevolume.cpp"

```
-----------------------------------------------------------------------

#include "defines.h"

//function declaration
   Matrix matrix(int n,int m);
   int  edge(Matrix q, Matrix p, int n, int m, double l);
   void PrintMatrix(Matrix q, int n, int m , FILE* fid),
        initialcon(Matrix q, int n, double l),
        newpostion (Matrix q,int i,int n,Matrix p2,Matrix pc2, double*
ne,double
                    lt,double la),
        contact(Matrix q, Matrix p, int n, int i, int m, double l),
        contact2(double* o, Matrix q, Matrix p, int n, int i, int m,
double l),
        polymer(int Max, double a, double b, Matrix o, Matrix oo,
Matrix q, int n,
                double l, MTRand mt);

//extern variables declaration
        int i3, ic2, I, Contact, Contact2, sigma;
        double P, rr, alpha, beta, c, gama;

int main(void)
{
//declare and initialize a mersenne twister random number generator
        MTRand mt;
        MTRand();
```

```c
//open or create a file for recording results
     FILE* fid;
     char filename[21];
     printf("\n Please input filename:");
     scanf("%s", & filename);
     fid=fopen(filename, "wb+");

//ask for user's inputs
     int N;
     printf("\n Number of particals: N= ");
     scanf("%d",&N);
     double theta;
     printf("\n diameter of particle: theta= ");
     scanf("%lf",&theta);
     double PHI;
     PHI= theta*theta*theta*N*PI/6;
     printf("\n volume fraction,PHI= %lf ",PHI);
     double lamda, Lamda;
     printf("\n normalized range of attraction: lamda/R0= ");
     scanf("%lf",&Lamda);
     lamda=theta*Lamda;
     double energy;
     printf("\n reduced depth of attraction: e/kT= ");
     scanf("%lf",&energy);
     double step, STEP;
     printf("\n normalized steplength: step/(average distance -
theta)= ");
     scanf("%lf",&STEP);
     double a=pow(1.0/N,1.0/3.0);
     step=(a-theta)*STEP;
     int Mmax;
     printf("\n Number of segments of polymer chain: M= ");
     scanf("%d",&Mmax);
     double length, Length;
     printf("\n normalized bond length of polymer : length= ");
     scanf("%lf",&Length);
      length=Length*theta;
     double angle,Angle;
     printf("\n bond angle of polymer (degree): angle= ");
     scanf("%lf",&Angle);
     angle=Angle/180*PI;
     double radii;
     printf("\n normalized atomic excluded radii with respect to bond
length :
                radii= ");
     scanf("%lf",&radii);
     c= radii*length;
     double T;
     printf("\n Temperature(K)= ");
     scanf("%lf",&T);
     double en, En;
      printf("\n Trans-Gauche energy (Cal/mol)= ");
     scanf("%lf",&En);
      en= En/1.987/T;
     double K0;
     printf("\n equibibrium cut-off K, K0= ");
```

```
      scanf("%lf",&K0);
      double KMAX;
      printf("\n Maximum Kp, Kmax= ");
      scanf("%lf",&KMAX);


//Variable declaration
      Matrix Q = matrix( N, 3); int i=0;
      Matrix Edge = matrix( N, 4); int i2=0;
      Matrix Edge2 = matrix( N, 4); i3=0;
      Matrix Edgec = matrix( N, 4); int ic=0;
      Matrix Edgec2 = matrix( N, 4); ic2=0;
      Matrix Poly = matrix( Mmax+1, 3);
      Matrix PolyN = matrix( Mmax+1, 3);
      double New[3];
      int j, k;
      double K, Kp, R, RR, R4;
      double dx, dy, dz, Phi, Theta, deltaE, W, acceptance, v, stdv,
stdR, stdRR;

//Generate the initial close-packing configuration
      initialcon(Q, N, theta);
       /*count the particles in the steric edge*/
      i2= edge(Q, Edge, N, i2, theta);
       /*count the particles in the contact edge*/
      ic= edge(Q, Edgec, N, ic, lamda);

      K=0; Kp=0; P=0; RR=0; R4=0; alpha= length*sin(PI-angle);
       beta= length*cos(PI-angle);
       sigma= floor(360/(180-Angle))-1; gama= 1/exp(en);

//Main cycle to generate statistical ensemble of protein configurations
      while (Kp< KMAX)
            { //random change of protein configuration.
              I=1;
              while (I==1)
               {
                  /*randomly select a particle*/
                     i= floor(N*mt.rand53());
                  Contact = 0;
                  contact(Q, Edgec, N, i, ic, lamda);

                     /*attempt to move the selected particle randomly*/
                     Theta= 2*PI*mt.rand53();
                     Phi= PI*mt.rand53();
                     dx=step*sin(Phi)*cos(Theta);
                     dy=step*sin(Phi)*sin(Theta);;
                     dz=step*cos(Phi);

                  New[0]= Q[i][0]+ dx;
                  New[1]= Q[i][1]+ dy;
                  New[2]= Q[i][2]+ dz;

                     /*restrict the destination within the cube*/
                  for (k = 0; k < 3; k++) {
                        if (New[k]>1) New[k]--;
                        else if (New[k]<0) New[k]++;
```

```
                }
                K++;

                //Check whether the new configuration is accepted or
not
                i3=i2; ic2=ic;
                    /*temporarily store Edge in Edge2*/
                for (j = 0; j < i3; j++) {
                   Edge2[j][0]=Edge[j][0];
                   Edge2[j][1]=Edge[j][1];
                   Edge2[j][2]=Edge[j][2];
                   Edge2[j][3]=Edge[j][3];
                }
                    /*temporarily store Edgec in Edgec2*/
                for (j = 0; j < ic2; j++) {
                   Edgec2[j][0]=Edgec[j][0];
                   Edgec2[j][1]=Edgec[j][1];
                   Edgec2[j][2]=Edgec[j][2];
                   Edgec2[j][3]=Edgec[j][3];
                }
                newpostion (Q,i,N,Edge2,Edgec2,New,theta,lamda);
             }


             /*calculate the No. of contacts if move to the new
position*/
             Contact2=0;
             contact2(New, Q, Edgec2, N, i, ic, lamda);
             /*Calculate the energy change due to move ith particles,
             and decide whether to accept the movement of ith
particle*/
             deltaE= energy*(Contact2- Contact);
             W= mt.rand53();
             if ( exp(deltaE)>=W) {
                    /*move particle*/
                 Q[i][0]=New[0]; Q[i][1]=New[1]; Q[i][2]=New[2];
                    /*renew Edge*/
                 for (j = 0; j < i3; j++) {
                   Edge[j][0]=Edge2[j][0];
                   Edge[j][1]=Edge2[j][1];
                   Edge[j][2]=Edge2[j][2];
                   Edge[j][3]=Edge2[j][3];
                 }
                 i2=i3;
                    /*renew Edge2*/
                 for (j = 0; j < ic2; j++) {
                   Edgec[j][0]=Edgec2[j][0];
                   Edgec[j][1]=Edgec2[j][1];
                   Edgec[j][2]=Edgec2[j][2];
                   Edgec[j][3]=Edgec2[j][3];
                 }
                 ic=ic2;
             }


             //Generate a polymer chain after equilibrium being
reached
```

```
            if (K > K0) {
                Kp++;
                polymer(Mmax, length, angle, Poly, PolyN, Q, N, theta,
mt);
                RR=RR+rr; R4=R4+square(rr);
                unsigned long int fla= (unsigned long int)Kp;
                  /*cycle counter, and result monitor*/
                if (fla%5000==0) printf("Kp= %lf; P/Kp= %lf;
Contact= %d;
                                        Contact2= %d\n",
Kp,P/Kp,Contact,Contact2);
                    if (fla%1000000==0) fprintf(fid,"Kp= %lf; P/Kp= %lf;
Contact= %d;
                                        Rg= %lf \n",Kp,P/Kp,Contact2,
sqrt(rr));
            }
        }

    //Calculate free-volume and the average gyration radius of
polymer chain
    acceptance=Kp/(K-K0);
    v=P/Kp;
     stdv= sqrt(v*(1-v)/Kp);
    RR= RR/P;
     stdRR= sqrt((R4/P-square((RR/P)))/P);
     R= sqrt(RR);
     stdR= stdRR/2/R;
    /*print out conditions and results in the file*/
    fprintf(fid,"Number of particals, N= %d. \n",N);
    fprintf(fid,"volume fraction, phi= %lf. \n",PHI);
    fprintf(fid,"diameter of particle, theta= %lf. \n",theta);
    fprintf(fid,"normalized range of attraction, lamda= %lf.
\n",lamda/theta);
    fprintf(fid,"reduced depth of attraction, e/kT= %lf. \n",energy);
    fprintf(fid,"Normalized bond length of Polymer length= %lf*theta.
\n",
            Length);
     fprintf(fid,"Bond Angle = %lf degree. \n",Angle);
     fprintf(fid,"Number of segments of polymer chain, M= %d.
\n",Mmax);
     fprintf(fid,"Atomic excluded radii= %lf. \n", c);
     fprintf(fid,"Temperature= %f K. \n", T);
     fprintf(fid,"Effective trans-Gauche energy = %lf cal/mol. \n",
En);
    fprintf(fid,"K0 %lf ; Kmax %lf ; \n",K0,KMAX);
    fprintf(fid,"Kp %lf ; P %lf ; acceptance %lf.
\n",Kp,P,acceptance);
    fprintf(fid,"Free volume for polymer is %lf, error(3stdv)= %lf.
\n",v,
            3*stdv);
    fprintf(fid,"mean radius of gyration of polymer is %lf,
error(3stdR)= %lf.
            \n \n",R, 3*stdR);
    fprintf(fid,"last configuration of colloidal particles \n");
    PrintMatrix(Q, N, 3 , fid);
    fprintf(fid,"\n\n last configuration of polymer random coil \n");
    PrintMatrix(Poly, Mmax+1, 3 , fid);
```

```
        fclose(fid);
    return 0;
}
//------------------------------------------------------------------------
```

2. Subfunction "matrix.cpp"

This subfunction is used for allocating addresses for a matrix

```
//------------------------------------------------------------------------
#include "defines.h"

#include "matrix.h"

Matrix matrix(int n,int m)
{
  void freematrix(Matrix q, int n);
  Matrix q;
  int i;
  q=(double**) malloc(n*sizeof(double*));
  if (q==NULL)
        return NULL;
  for (i=0; i < n; i++) {
        q[i]=(double*) malloc(m*sizeof(double));
        if (q[i]== NULL) {
              freematrix(q,i);/*avoid garbage*/
              return NULL;
        }
  }
  return q;
}

void freematrix(Matrix q, int n)/*Free the memory if it is full*/
{
  int i;
  for (i = 0; i < n; i++)free (q[i]);
  free (q);
}
//------------------------------------------------------------------------
```

3. Subfunction "edge.cpp"

This subfunction is used for labeling the spheres in the edge of the box, where the

periodic condition is applied.

```
//------------------------------------------------------------------------
#include "defines.h"
#include "edge.h"

int edge(Matrix q, Matrix p, int n, int m, double l)
 {  /*Count the particles in the edge of the box with thickness l*/
```

```
    int i,k;
      for (i = 0; i < n; i++) {
            for (k = 0; k < 3; k++) {
                  if (q[i][k]<l||q[i][k]>(1-l)) {
                        p[m][0]=q[i][0];
                        p[m][1]=q[i][1];
                        p[m][2]=q[i][2];
                        p[m][3]=i;
                        m++;
                  }
                  break;
            }
      }
      return m;
  }
//----------------------------------------------------------------
```

4. Subfunction "PrintMatrix.cpp"

This subfunction is used for saving a matrix to a text file.

```
//----------------------------------------------------------------
#include "defines.h"
#include "PrintMatrix.h"

void PrintMatrix(Matrix q, int n, int m , FILE* fid )
{     int i,j;
      for (i = 0; i < n; i++){
        fprintf(fid,"\n");
        for (j = 0; j < m; j++)
              fprintf(fid," %lf  ", q[i][j]);
      }
}

//----------------------------------------------------------------
```

5. Subfunction "initialcon.cpp"

This subfunction is used for assigning the close-packing initial configuration for N

spheres in the unit cube.

```
//----------------------------------------------------------------
#include "defines.h"
#include "initialcon.h"

void initialcon(Matrix q, int n, double l)
{
      double a, k1, k2, k3, Rsqu;
      int i, j;
      i=0; a= pow(4.0/n,1.0/3.0);
      for (k1 = 0; k1 < 1; k1+=a) {
```

```
                   for (k2 = 0; k2 < 1; k2+=a) {
                        for (k3 = 0; k3 < 1; k3+=a){
                             q[i][0]=k1; q[i][1]=k2; q[i][2]=k3;
                             i++;
                        }
                   }
              }
         for (k1 = a/2; k1 < 1; k1+=a) {
              for (k2 = a/2; k2 < 1; k2+=a) {
                   for (k3 = 0; k3 < 1; k3+=a){
                             q[i][0]=k1; q[i][1]=k2; q[i][2]=k3;
                             i++;
                        }
                   }
              }
         for (k1 = a/2; k1 < 1; k1+=a) {
              for (k2 = 0; k2 < 1; k2+=a) {
                   for (k3 = a/2; k3 < 1; k3+=a){
                             q[i][0]=k1; q[i][1]=k2; q[i][2]=k3;
                             i++;
                        }
                   }
              }
         for (k1 = 0; k1 < 1; k1+=a) {
              for (k2 = a/2; k2 < 1; k2+=a) {
                   for (k3 = a/2; k3 < 1; k3+=a){
                             q[i][0]=k1; q[i][1]=k2; q[i][2]=k3;
                             i++;
                        }
                   }
              }
         /*check whether there is overlap*/
         for (i = 0; i < (n-1); i++) {
              for (j =(i+1); j < n; j++) {
                   Rsqu= dist(q,q,i,j);
                   if (Rsqu <= square(l)){
                   perror("Large n leads to overlap betwe en particles");
                   exit(1);}
              }
         }
    }
//----------------------------------------------------------------------
```

6. Subfunction "newposition.cpp"

This subfunction is used for moving the sphere by Metropolis equilibration algorithm.

```
//----------------------------------------------------------------------
#include "defines.h"
#include "newposition.h"

void newpostion (Matrix q, int i, int n, Matrix p2, Matrix pc2, double*
                ne, double lt, double la)
{
    extern int i3, ic2, I;
```

```
int j,h,k,k2; double Rsqu;

/*check overlap between particles in the box*/
for ( j=0; j< n; j++){
    if (j==i) continue;
    Rsqu=square((ne[0]-q[j][0]))+square((ne[1]-q[j][1]))+
        square((ne[2]-q[j][2]));
    if (Rsqu< square(lt)) return;
}


/*update matrix of sphere in the edge*/
/*check whether the old position in the edge*/
for (j = 0; j < i3; j++) {
    /*if the selected particle in the edge2, remove it from the p2
      matrix*/
    if ((int)p2[j][3]==i) {
        for (h = j; h < (i3-1); h++) {
          p2[h][0]=p2[h+1][0];
          p2[h][1]=p2[h+1][1];
          p2[h][2]=p2[h+1][2];
          p2[h][3]=p2[h+1][3];
        }
        i3--;
        break;
    }
}
/*check whether the new position in the edge*/
/*if the new position in the edge2, add it in the end of the p2
  matrix */
if (ne[0]<lt||ne[0]>(1-lt)||ne[1]<lt||ne[1]>(1-lt)||ne[2]<lt
    ||ne[2]>(1-lt)) {
    p2[i3][0]=ne[0];
    p2[i3][1]=ne[1];
    p2[i3][2]=ne[2];
    p2[i3][3]=i;
    i3++;}
else {I=0; return;}

/*Check overlap in the edge under periodic boundary condition*/
for (j = 0; j < (i3-1); j++) {
    Rsqu=square((ne[0]-p2[j][0]))+square((ne[1]-p2[j][1]))
        +square((ne[2]-p2[j][2]));
    /*6 neighbors on the faces*/
    for (k = 0; k < 3; k++) {
        if ((Rsqu+2*(ne[k]-p2[j][k])+1)<square(lt)||(Rsqu-2*(ne[k]-
            p2[j][k])+1)<square(lt))
            return;
    /*12 neighbors on the sides*/
    for (k2 = 1; k2 < 3; k2++) {
      if(k2<=k) continue;
      if((Rsqu+2*(ne[k]-p2[j][k])+2*(ne[k2]-p2[j][k2])+2)<square(lt)
        ||(Rsqu-2*(ne[k]-p2[j][k])-2*(ne[k2]-p2[j][k2])+2)<square(lt)
        ||(Rsqu+2*(ne[k]-p2[j][k])-2*(ne[k2]-p2[j][k2])+2)<square(lt)
        ||(Rsqu-2*(ne[k]-p2[j][k])+2*(ne[k2]-p2[j][k2])+2)<square(lt)
          ) return;
    }
  }
```

```
        /* 8 neighbors at the corners*/
        if ((Rsqu+2*(ne[0]-p2[j][0])+2*(ne[1]-p2[j][1])+2*(ne[2]-
            p2[j][2])+3)<square(lt)||
            (Rsqu-2*(ne[0]-p2[j][0])-2*(ne[1]-p2[j][1])+2*(ne[2]-
            p2[j][2])+3)<square(lt)||
            (Rsqu+2*(ne[0]-p2[j][0])-2*(ne[1]-p2[j][1])+2*(ne[2]-
            p2[j][2])+3)<square(lt)||
            (Rsqu-2*(ne[0]-p2[j][0])+2*(ne[1]-p2[j][1])+2*(ne[2]-
            p2[j][2])+3)<square(lt)||
            (Rsqu+2*(ne[0]-p2[j][0])+2*(ne[1]-p2[j][1])-2*(ne[2]-
            p2[j][2])+3)<square(lt)||
            (Rsqu-2*(ne[0]-p2[j][0])-2*(ne[1]-p2[j][1])-2*(ne[2]-
            p2[j][2])+3)<square(lt)||
            (Rsqu+2*(ne[0]-p2[j][0])-2*(ne[1]-p2[j][1])-2*(ne[2]-
            p2[j][2])+3)<square(lt)||
            (Rsqu-2*(ne[0]-p2[j][0])+2*(ne[1]-p2[j][1])-2*(ne[2]-
            p2[j][2])+3)<square(lt))
            return;
    }
    I=0;

    /*no overlap for moving the sphere*/
     /*renew the matrix pc2*/
     for (j = 0; j < ic2; j++) {
       if ((int)pc2[j][3]==i) {
          for (h = j; h < (ic2-1); h++) {
            pc2[h][0]=pc2[h+1][0];
            pc2[h][1]=pc2[h+1][1];
            pc2[h][2]=pc2[h+1][2];
            pc2[h][3]=pc2[h+1][3];
          }
          ic2--; break;
       }
     }
     if (ne[0]<la||ne[0]>(1-la)||ne[1]<la||ne[1]>(1-la)||ne[2]<la
         ||ne[2]>(1-la)) {
         pc2[ic2][0]=ne[0];
         pc2[ic2][1]=ne[1];
         pc2[ic2][2]=ne[2];
         pc2[ic2][3]=i;
         ic2++;
     }
}
//-------------------------------------------------------------------
```

7. Subfunction "contact.cpp"

This subfunction is used for calculating the number of contacts made by the selected

sphere.

```
//-------------------------------------------------------------------
#include "defines.h"
#include "contact.h"
```

```
void contact(Matrix q, Matrix p, int n, int i, int m, double l)
    {
     extern int Contact;
     int j, h, k, k2; double Rsqu;

     /*contacts made by the selected sphere inside the box.*/
     for (j = 0; j < n; j++) {
       if (j==i) continue;
       Rsqu= dist(q,q,i,j);
       if (Rsqu<square(l)) Contact++;
     }

     /*check if the selected particle in the edge*/
     for (h = 0; h < m; h++) {
     /*if it is in the edge, calculate the contacts made by the
      selected sphere in the edge under periodic boundary condition.*/
       if ((int)p[h][3]==i) {
          for (j = 0; j < m; j++) {
            if ((int)p[j][3]==i) continue;
            Rsqu= dist(q,p,i,j);
            for (k = 0; k < 3; k++) {
                /*6 neighbors on the faces*/
                if (Rsqu+2*(q[i][k]-p[j][k])+1<square(l)||
                    Rsqu-2*(q[i][k]-p[j][k])+1<square(l))
                    Contact++;
                /*12 neighbors on the sides*/
                for (k2 = 1; k2 < 3; k2++) {
                    if (k2<=k) continue;
                    if ((Rsqu+2*(q[i][k]-p[j][k])+2*(q[i][k2]-
                         p[j][k2])+2) <square(l)||
                        (Rsqu-2*(q[i][k]-p[j][k])-2*(q[i][k2]-
                         p[j][k2])+2) <square(l)||
                        (Rsqu+2*(q[i][k]-p[j][k])-2*(q[i][k2]-
                         p[j][k2])+2)<square(l)||
                        (Rsqu-2*(q[i][k]-p[j][k])+2*(q[i][k2]-
                         p[j][k2])+2)<square(l))
                         Contact++;
                }
          }
           /*8 neighbors at the corners*/
          if ((Rsqu+2*( q[i][0]-p[j][0])+2*(q[i][1]-p[j][1])+2*(q[i][2]
               -p[j][2])+3)<square(l)||
              (Rsqu-2*(q[i][0]-p[j][0])-2*(q[i][1]-p[j][1])+2*(q[i][2]
               -p[j][2])+3)<square(l)||
              (Rsqu+2*(q[i][0]-p[j][0])-2*(q[i][1]-p[j][1])+2*(q[i][2]
               -p[j][2])+3)<square(l)||
              (Rsqu-2*(q[i][0]-p[j][0])+2*(q[i][1]-p[j][1])+2*(q[i][2]
               -p[j][2])+3)<square(l)||
               Rsqu+2*(q[i][0]-p[j][0])+2*(q[i][1]-p[j][1])-2*(q[i][2]
               -p[j][2])+3)<square(l)||
              (Rsqu-2*(q[i][0]-p[j][0])-2*(q[i][1]-p[j][1])-2*(q[i][2]
               -p[j][2])+3)<square(l)||
              (Rsqu+2*(q[i][0]-p[j][0])-2*(q[i][1]-p[j][1])-2*(q[i][2]
               -p[j][2])+3)<square(l)||
              (Rsqu-2*(q[i][0]-p[j][0])+2*(q[i][1]-p[j][1])-2*(q[i][2]
               -p[j][2])+3)<square(l))
```

```
                Contact++;
            }
        }
      break;
    }
  }
//-----------------------------------------------------------------------
```

8. Subfunction "contact2.cpp"

This subfunction is used for calculating the number of contacts made by the new position.

```cpp
//-----------------------------------------------------------------------
#include "defines.h"
#include "contact2.h"

void contact2(double* o, Matrix q, Matrix p, int n, int i, int m,
              double l)
{
    extern int Contact2;
    int j, k, k2; double Rsqu;

    /*contacts made by the new position inside the box.*/
    for (j = 0; j < n; j++) {
      if (j==i) continue;
      Rsqu= square((o[0]-q[j][0]))+ square((o[1]-q[j][1]))+
            square((o[2]-q[j][2]));
      if (Rsqu<square(l)) Contact2++;
    }

    /*check if the new position in the edge*/
    if ( m>0 && (int)p[m-1][3]==i) {
    /*if it is in the edge, calculate the contacts made by the
     new position in the edge under periodic boundary condition.*/
      for (j = 0; j < m; j++) {
        if ((int)p[j][3]==i) continue;
        Rsqu= square((o[0]-p[j][0]))+ square((o[1]-p[j][1]))+
              square((o[2]-p[j][2]));
        /*6 neighbors on the faces*/
        for (k = 0; k < 3; k++) {
            if (Rsqu+2*(o[k]-p[j][k])+1<square(l)||Rsqu-2*(o[k]-
               p[j][k])+1<square(l))
              Contact2++;
        /*12 neighbors on the sides*/
        for (k2 = 1; k2 < 3; k2++) {
          if  (k2<=k) continue;
          if  ((Rsqu+2*(o[k]-p[j][k])+2*(o[k2]-p[j][k2])+2)<square(l)
              ||(Rsqu-2*(o[k]-p[j][k])-2*(o[k2]-p[j][k2])+2)<square(l)
              ||(Rsqu+2*(o[k]-p[j][k])-2*(o[k2]-p[j][k2])+2)<square(l)
             ||(Rsqu-2*(o[k]-p[j][k])+2*(o[k2]-p[j][k2])+2)<square(l))
              Contact2++;
        }
      }
        /*8 neighbors at the corners*/
```

```
        if ((Rsqu+2*(o[0]-p[j][0])+2*(o[1]-p[j][1])+2*(o[2]-p[j][2])+3)
            <square(l)||
            (Rsqu-2*(o[0]-p[j][0])-2*(o[1]-p[j][1])+2*(o[2]-p[j][2])+3)
            <square(l)||
            (Rsqu+2*(o[0]-p[j][0])-2*(o[1]-p[j][1])+2*(o[2]-p[j][2])+3)
            <square(l)||
            (Rsqu-2*(o[0]-p[j][0])+2*(o[1]-p[j][1])+2*(o[2]-p[j][2])+3)
            <square(l)||
            (Rsqu+2*(o[0]-p[j][0])+2*(o[1]-p[j][1])-2*(o[2]-p[j][2])+3)
            <square(l)||
            (Rsqu-2*(o[0]-p[j][0])-2*(o[1]-p[j][1])-2*(o[2]-p[j][2])+3)
            <square(l)||
            (Rsqu+2*(o[0]-p[j][0])-2*(o[1]-p[j][1])-2*(o[2]-p[j][2])+3)
            <square(l)||
            (Rsqu-2*(o[0]-p[j][0])+2*(o[1]-p[j][1])-2*(o[2]-p[j][2])+3)
            <square(l))
            Contact2++;
        }
    }
}
//---------------------------------------------------------------------
```

9. Subfunction "polymer.cpp"

This subfunction is used for generating a polymer chain.

```
//---------------------------------------------------------------------
#include "defines.h"
#include "polymer.h"

void polymer(int Max, double a, double b , Matrix o, Matrix oo, Matrix q,
int n, double l, MTRand mt)
{
   extern double P, rr, alpha, beta, c, gama;
   extern int sigma;

   int j, k, M, m, w, Sign; double Rsqu, x, y, z, dx, dy, dz, Theta,
Phi, av, avv, zeta, delta, Theta2, check;
       rr= 0;check=1;

   while (check==1)
   {   check=0;
       M= 0;
       /*generation of a terminal of polymer chain*/
       o[0][0]= mt.rand53(); o[0][1]= mt.rand53(); o[0][2]= mt.rand53();
       M++;

       /*generation of the second point of polymer*/

       Theta= 2*PI*mt.rand53(); Phi= PI*mt.rand53();
       dx=a*sin(Phi)*cos(Theta);//changes in 3 directions
       dy=a*sin(Phi)*sin(Theta);
       dz=a*cos(Phi);
```

```
       o[1][0]= o[0][0]+dx;
        o[1][1]= o[0][1]+dy;
        o[1][2]= o[0][2]+dz;
       M++;

    /*generation of the third point*/
           /*coordinates in the local coordinate system*/
           delta= 2*PI*mt.rand53();
           x= alpha*cos(delta);
           y= alpha*sin(delta);
           z= beta;
           /*convert the local coordinates to universal coordinates*/
                  Theta= atan(dy/dx)+PI*(dx<0);
           Phi= acos(dz/a);
           dx = cos(Phi)*cos(Theta)*x -sin(Theta)*y
+sin(Phi)*cos(Theta)*z;
           dy = cos(Phi)*sin(Theta)*x +cos(Theta)*y
+sin(Phi)*sin(Theta)*z;
           dz = -sin(Phi)*x +cos(Phi)*z;

           o[M][0]= dx + o[M-1][0];
           o[M][1]= dy + o[M-1][1];
           o[M][2]= dz + o[M-1][2];
       M++;

       /*generation of the other succesive points of polymer chain*/
       while (M<= Max) {
              /*coordinates in the local coordinate system*/
           Theta= atan(dy/dx)+PI*(dx<0);
           Phi= acos(dz/a);
           x= cos(Phi)*cos(Theta)*(o[M-3][0] -o[M-2][0])
+cos(Phi)*sin(Theta)*(o[M-3][1] -o[M-2][1]) -sin(Phi)*(o[M-3][2] -o[M-
2][2]);
           y= -sin(Theta)*(o[M-3][0] -o[M-2][0]) +cos(Theta)*(o[M-3][1]
-o[M-2][1]);
           z= cos(Theta)*sin(Phi)*(o[M-3][0] -o[M-2][0])
+sin(Theta)*sin(Phi)*(o[M-3][1] -o[M-2][1]) +cos(Phi)*(o[M-3][2] -o[M-
2][2]);
           Theta2= atan(y/x)+PI*(x<0);//theta angle of second last
bond
           zeta= mt.rand53();
          if (zeta< 1/(1+gama)) delta= PI*(1+gama) *zeta+ Theta2+ PI/2;
           else delta= PI*((1+gama)*zeta-1+gama)/gama+ Theta2+ PI/2;
           x= alpha*cos(delta);
           y= alpha*sin(delta);
           z= beta;
           /*convert the local coordinates to universal coordinates*/
           dx = cos(Phi)*cos(Theta)*x -sin(Theta)*y
+sin(Phi)*cos(Theta)*z;
           dy = cos(Phi)*sin(Theta)*x +cos(Theta)*y
+sin(Phi)*sin(Theta)*z;
           dz = -sin(Phi)*x +cos(Phi)*z;

           o[M][0]= dx + o[M-1][0];
           o[M][1]= dy + o[M-1][1];
           o[M][2]= dz + o[M-1][2];
```

```
            /*Check whether the point is exclued by the preceding
points*/
                for (j= 0; j<= (M-sigma); j++) {
                    Rsqu=square((o[M][0]-o[j][0]))+ square((o[M][1]-
o[j][1]))+ square((o[M][2]-o[j][2]));
                    if (Rsqu<=square((2*c))) {check=1; break;};
                        }
                if (check==1) break;//if the last point is steric rejected
generate another Polymer
                M++;
            }
    }
    M--;

    /*check whether the polymer configuration is sterically allowed in
the presence of spheres*/
    for (m = 0; m <= M; m++){
        /*In the box*/
        for (j = 0; j < n; j++) {
            Rsqu=square((o[m][0]-q[j][0]))+ square((o[m][1]-q[j][1]))+
square((o[m][2]-q[j][2]));
            if (Rsqu<square((l/2))) return;
        }
        /*Under periodic condition*/
        for (k= 0; k < 3; k++) {
            oo[m][k]=o[m][k]; w=0;
            if (o[m][k]<l/2) {oo[m][k]=o[m][k]+1; w=1;}
            else if (o[m][k]>1-l/2) {oo[m][k]=o[m][k]-1; w=1;}
        }
        if (w==1) {
         for (j = 0; j < n; j++) {
            Rsqu=square((oo[m][0]-q[j][0]))+ square((oo[m][1]-q[j][1]))+
square((oo[m][2]-q[j][2]));
            if (Rsqu<square((l/2))) return;
            }
        }
    }

  /*Accumulation of gyration radius of the polymer*/
        for (k = 0; k < 3; k++) {
            av= 0; avv=0;
            for (j= 0; j<= M; j++) {
                avv= avv+ square((o[j][k]));
                av= av+ o[j][k];
            }
            rr= rr+ avv/(M+1)- square((av/(M+1)));
        }
        P++;
}
//-----------------------------------------------------------------
```

10. Head file "defines.h"

```
//-----------------------------------------------------------------

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <fcntl.h>
#include "MersenneTwister.h"
#define PI 3.14159265358979
#define square(x) (x*x)
#define dist(A,B,i,j) (square((A[i][0]-B[j][0]))+square((A[i][1]-
B[j][1]))+square((A[i][2]-B[j][2])))
typedef double** Matrix; /*hide the int** data type*/

//-----------------------------------------------------------------
```

Appendix I Data

In this appendix we report the results of Monte Carlo simulations.

Table 1. The free-volume fraction, α, for a AO-sphere in the presence of hard-spheres at different volume fraction of hard-sphere, ϕ, and ratio of AO-sphere radius and hard-sphere radius.

	$q = 0$		$q = 0.15$	
ϕ	α	error	α	error
0.05	0.95000	0.00002	0.92401	0.00003
0.15	0.85005	0.00004	0.77267	0.00004
0.30	0.70005	0.00005	0.54858	0.00005
0.40	0.60008	0.00005	0.40361	0.00005
0.45	0.55008		0.33345	0.00005

	$q = 0.25$		$q = 0.50$	
ϕ	α	error	α	error
0.05	0.90268	0.00003	0.83485	0.00004
0.15	0.71101	0.00005	0.53259	0.00005
0.30	0.43764	0.00005	0.18913	0.00004
0.40	0.27390	0.00005	0.05996	0.00002
0.45	0.20147	0.00004	0.02626	0.00002

	$q = 1$	
ϕ	α	error
0.05	0.64329	0.00005
0.15	0.19255	0.00004
0.30	0.00730	0.00001
0.40	0.00010	0.00000
0.45	0.00000	0.00000

Table 2. The gyration radius, R_g, and free volume fraction, α, of a segment chain at different ratio of segment length to radius of hard-sphere.

		$l\,/\,R$= 0.028519		
ϕ	R_g/R	error	α	error
0.01	0.23505	0.00004	0.98216	0.00001
0.02	0.23499	0.00003	0.96461	0.00002
0.03	0.23491	0.00005	0.94720	0.00002
0.05	0.23466	0.00004	0.91268	0.00003

		$l\,/\,R$= 0.067704		
ϕ	R_g/R	error	α	error
0.01	0.55776	0.00011	0.96822	0.00002
0.02	0.55698	0.00009	0.93723	0.00002
0.03	0.55629	0.00010	0.90685	0.00003
0.05	0.55507	0.00011	0.84766	0.00004

		$l\,/\,R$= 0.121867		
ϕ	R_g/R	error	α	error
0.01	1.00277	0.00019	0.94422	0.00002
0.02	1.00068	0.00018	0.89139	0.00003
0.03	0.99879	0.00018	0.84075	0.00004
0.05	0.99453	0.00020	0.74579	0.00004

		$l\,/\,R$= 0.182800		
ϕ	R_g/R	error	α	error
0.01	1.50262	0.00026	0.91287	0.00003
0.02	1.49834	0.00027	0.83339	0.00004
0.03	1.49411	0.00029	0.76021	0.00004
0.05	1.48554	0.00033	0.63015	0.00005

Table 3. The gyration radius, R_g, and free volume fraction, α, of a fixed-angle segment chain at different ratio of segment length to radius of hard-sphere.

Fixed-angle 30°

l/R= 0.087995				
ϕ	R_g/R	error	α	error
0.01	0.21324	0.00000	0.98136	0.00001
0.02	0.21317	0.00000	0.96295	0.00002
0.03	0.21309	0.00000	0.94470	0.00002
0.05	0.21294	0.00002	0.90867	0.00003

l/R= 0.246385				
ϕ	R_g/R	error	α	error
0.01	0.59667	0.00004	0.95840	0.00002
0.02	0.59588	0.00003	0.91833	0.00003
0.03	0.59518	0.00003	0.87923	0.00003
0.05	0.59392	0.00004	0.80404	0.00004

l/R= 0.439973				
ϕ	R_g/R	error	α	error
0.01	1.06390	0.00007	0.91845	0.00003
0.02	1.06141	0.00006	0.84331	0.00004
0.03	1.05884	0.00008	0.77328	0.00000
0.05	1.05371	0.00007	0.64727	0.00005

l/R= 0.659959				
ϕ	R_g/R	error	α	error
0.01	1.59293	0.00011	0.85979	0.00004
0.02	1.58629	0.00009	0.74009	0.00004
0.03	1.58054	0.00010	0.63646	0.00005
0.05	1.56929	0.00013	0.46868	0.00005

$l / R=$ 0.879945

ϕ	R_g/R	error	α	error
0.01	2.11949	0.00011	0.79188	0.00004
0.02	2.10850	0.00015	0.63061	0.00005
0.03	2.09847	0.00016	0.50323	0.00005
0.05	2.08344	0.00020	0.32278	0.00005

$l / R=$ 1.099932

ϕ	R_g/R	error	α	error
0.01	2.64617	0.00019	0.72183	0.00005
0.02	2.63059	0.00021	0.52882	0.00005
0.03	2.61847	0.00023	0.39143	0.00005
0.05	2.60644	0.00031	0.22213	0.00004

Fixed-angle 60°

$l / R=$ 0.041940

ϕ	R_g/R	error	α	error
0.01	0.20348	0.00000	0.98240	0.00001
0.02	0.20340	0.00000	0.96504	0.00002
0.03	0.20334	0.00000	0.94784	0.00002
0.05	0.20320	0.00002	0.91378	0.00003

$l / R=$ 0.117432

ϕ	R_g/R	error	α	error
0.01	0.56936	0.00004	0.96347	0.00002
0.02	0.56859	0.00003	0.92807	0.00003
0.03	0.56797	0.00003	0.89349	0.00003
0.05	0.56671	0.00004	0.82654	0.00004

		$l/R= 0.209700$			
ϕ	R_g/R	error	α	error
0.01	1.01549	0.00007	0.93232	0.00003
0.02	1.01301	0.00006	0.86914	0.00003
0.03	1.01083	0.00005	0.80944	0.00004
0.05	1.00683	0.00007	0.69945	0.00005

		$l/R= 0.314549$			
ϕ	R_g/R	error	α	error
0.01	1.52103	0.00007	0.88853	0.00003
0.02	1.51600	0.00009	0.79012	0.00004
0.03	1.51107	0.00010	0.70205	0.00005
0.05	1.50201	0.00011	0.55203	0.00005

		$l/R= 0.419399$			
ϕ	R_g/R	error	α	error
0.01	2.02540	0.00011	0.83839	0.00004
0.02	2.01707	0.00015	0.70530	0.00005
0.03	2.00955	0.00008	0.59386	0.00005
0.05	1.99724	0.00018	0.42148	0.00005

		$l/R= 0.524249$			
ϕ	R_g/R	error	α	error
0.01	2.52955	0.00015	0.78500	0.00004
0.02	2.51805	0.00018	0.62169	0.00005
0.03	2.50883	0.00021	0.49467	0.00005
0.05	2.49698	0.00024	0.31784	0.00005

Fixed-angle 90°

ϕ	R_g/R	error	α	error
		$l/R = 0.024373$		
0.01	0.20094	0.00000	0.98294	0.00001
0.02	0.20084	0.00000	0.96609	0.00002
0.03	0.20082	0.00000	0.94933	0.00002
0.05	0.20071	0.00002	0.91617	0.00003

ϕ	R_g/R	error	α	error
		$l/R = 0.068245$		
0.01	0.56210	0.00004	0.96577	0.00002
0.02	0.56164	0.00003	0.93259	0.00003
0.03	0.56102	0.00003	0.90008	0.00003
0.05	0.55979	0.00004	0.83705	0.00004

ϕ	R_g/R	error	α	error
		$l/R = 0.121867$		
0.01	1.00288	0.00007	0.93857	0.00002
0.02	1.00089	0.00006	0.88079	0.00003
0.03	0.99898	0.00005	0.82594	0.00004
0.05	0.99514	0.00007	0.72397	0.00005

ϕ	R_g/R	error	α	error
		$l/R = 0.182800$		
0.01	1.50277	0.00007	0.90132	0.00003
0.02	1.49887	0.00009	0.81294	0.00004
0.03	1.49434	0.00010	0.73271	0.00004
0.05	1.48679	0.00011	0.59331	0.00005

ϕ	R_g/R	error	α	error

l / R= 0.243733

ϕ	R_g/R	error	α	error
0.01	2.00228	0.00011	0.85957	0.00004
0.02	1.99552	0.00012	0.74097	0.00004
0.03	1.98986	0.00013	0.63908	0.00005
0.05	1.97965	0.00015	0.47531	0.00005

l / R= 0.304667

ϕ	R_g/R	error	α	error
0.01	2.50120	0.00015	0.81548	0.00004
0.02	2.49308	0.00018	0.66951	0.00005
0.03	2.48551	0.00018	0.55172	0.00005
0.05	2.47696	0.00022	0.37744	0.00005

Fixed-angle 109.5°

l / R= 0.01729

ϕ	R_g/R	error	α	error
0.01	0.19948	0.00000	0.98323	0.00001
0.02	0.19942	0.00000	0.96672	0.00002
0.03	0.19937	0.00000	0.95027	0.00002
0.05	0.19922	0.00002	0.91777	0.00003

l / R= 0.028519

ϕ	R_g/R	error	α	error
0.01	0.32899	0.00000	0.97798	0.00002
0.02	0.32876	0.00003	0.95636	0.00002
0.03	0.32860	0.00003	0.93497	0.00003
0.05	0.32824	0.00002	0.89278	0.00003

		l / R= 0.048411		
ϕ	R_g/R	error	α	error
0.01	0.55817	0.00004	0.96726	0.00002
0.02	0.55763	0.00003	0.93536	0.00003
0.03	0.55705	0.00003	0.90418	0.00003
0.05	0.55590	0.00004	0.84333	0.00004

		l / R= 0.086449		
ϕ	R_g/R	error	α	error
0.01	0.99588	0.00007	0.94235	0.00002
0.02	0.99409	0.00006	0.88794	0.00003
0.03	0.99233	0.00005	0.83606	0.00004
0.05	0.98902	0.00007	0.73904	0.00004

		l / R= 0.129673		
ϕ	R_g/R	error	α	error
0.01	1.49233	0.00007	0.90915	0.00003
0.02	1.48913	0.00009	0.82701	0.00004
0.03	1.48565	0.00010	0.75179	0.00004
0.05	1.47913	0.00011	0.61943	0.00005

		l / R= 0.172898		
ϕ	R_g/R	error	α	error
0.01	1.98896	0.00011	0.87254	0.00003
0.02	1.98364	0.00012	0.76333	0.00004
0.03	1.97852	0.00013	0.66789	0.00005
0.05	1.97125	0.00015	0.51094	0.00005

		l / R= 0.216122		
ϕ	R_g/R	error	α	error
0.01	2.48586	0.00015	0.83451	0.00004
0.02	2.47921	0.00018	0.70055	0.00005
0.03	2.47373	0.00018	0.58942	0.00005
0.05	2.46681	0.00022	0.41976	0.00005

Fixed-angle 120°

$l/R= 0.014166$				
ϕ	R_g/R	error	α	error
0.01	0.19839	0.00000	0.98344	0.00001
0.02	0.19832	0.00000	0.96712	0.00002
0.03	0.19828	0.00000	0.95087	0.00002
0.05	0.19817	0.00002	0.91879	0.00003

$l/R= 0.039664$				
ϕ	R_g/R	error	α	error
0.01	0.55511	0.00004	0.96813	0.00002
0.02	0.55469	0.00003	0.93713	0.00002
0.03	0.55401	0.00003	0.90669	0.00003
0.05	0.55303	0.00004	0.84756	0.00004

$l/R= 0.070829$				
ϕ	R_g/R	error	α	error
0.01	0.99065	0.00007	0.94471	0.00002
0.02	0.98881	0.00006	0.89243	0.00003
0.03	0.98730	0.00005	0.80944	0.00004
0.05	0.98423	0.00007	0.74871	0.00004

$l/R= 0.106243$				
ϕ	R_g/R	error	α	error
0.01	1.48496	0.00007	0.91392	0.00003
0.02	1.48153	0.00009	0.83577	0.00004
0.03	1.47859	0.00010	0.76379	0.00005
0.05	1.47305	0.00011	0.63616	0.00005

		$l/R=$ 0.141658		
ϕ	R_g/R	error	α	error
0.01	1.97886	0.00011	0.88046	0.00003
0.02	1.97435	0.00012	0.77704	0.00004
0.03	1.97022	0.00013	0.68592	0.00005
0.05	1.96370	0.00015	0.53401	0.00005

		$l/R=$ 0.177072		
ϕ	R_g/R	error	α	error
0.01	2.47310	0.00015	0.84613	0.00004
0.02	2.46778	0.00015	0.71974	0.00005
0.03	2.46369	0.00018	0.61340	0.00005
0.05	2.45813	0.00020	0.44750	0.00005

Fixed-angle 150°

		$l/R=$ 0.006767		
ϕ	R_g/R	error	α	error
0.01	0.18878	0.00000	0.98447	0.00001
0.02	0.18873	0.00000	0.96914	0.00002
0.03	0.18871	0.00000	0.95380	0.00002
0.05	0.18863	0.00000	0.92362	0.00003

		$l/R=$ 0.018948		
ϕ	R_g/R	error	α	error
0.01	0.52836	0.00004	0.97235	0.00002
0.02	0.52815	0.00003	0.94538	0.00002
0.03	0.52792	0.00003	0.91878	0.00003
0.05	0.52733	0.00002	0.86687	0.00003

		$l\,/\,R$= 0.050753		
ϕ	R_g/R	error	α	error
0.01	1.41496	0.00007	0.93450	0.00003
0.02	1.41350	0.00009	0.87376	0.00003
0.03	1.41247	0.00008	0.81664	0.00004
0.05	1.41056	0.00009	0.71174	0.00005

		$l\,/\,R$= 0.067670		
ϕ	R_g/R	error	α	error
0.01	1.88631	0.00011	0.91347	0.00003
0.02	1.88524	0.00012	0.83581	0.00004
0.03	1.88428	0.00013	0.76478	0.00004
0.05	1.88249	0.00013	0.63969	0.00005

		$l\,/\,R$= 0.084588		
ϕ	R_g/R	error	α	error
0.01	2.35892	0.00015	0.89302	0.00003
0.02	2.35771	0.00015	0.80013	0.00004
0.03	2.35703	0.00016	0.71752	0.00005
0.05	2.35731	0.00018	0.57792	0.00005

		$l\,/\,R$= 0.121867		
ϕ	R_g/R	error	α	error
0.01	3.40037	0.00019	0.85143	0.00004
0.02	3.40178	0.00021	0.73127	0.00004
0.03	3.40436	0.00023	0.63090	0.00005
0.05	3.41198	0.00026	0.47477	0.00005

Table 4. The gyration radius, R_g, normalized by the length of segment, l, for self-avoiding chain with two parameters, $\hat{r}$ and $\hat{\epsilon}$. Simulations have been performed on self-avoiding chain with M segments (101, 137 and 206. The scaling exponents, a, are calculated from R_g's.

$\hat{r} = 0.00$					
$\hat{\epsilon}$	$N_A\epsilon$ (cal/mole)	R_g/l (M=101)	R_g/l (M=137)	R_g/l (M=206)	a
0.000	0	5.77	6.73	8.26	0.501
0.343	200	6.38	7.45	9.16	0.506
0.858	500	7.34	8.61	10.61	0.515
1.373	800	8.29	9.75	12.07	0.524
1.888	1100	9.13	10.77	13.37	0.532
2.575	1500	10.00	11.84	14.75	0.541
3.434	2000	10.67	12.67	15.83	0.550
5.150	3000	11.16	13.27	16.63	0.557
8.584	5000	11.28	13.42	16.83	0.558
$\hat{r} = 0.10$					
$\hat{\epsilon}$	$N_A\epsilon$ (cal/mole)	R_g/l (M=101)	R_g/l (M=137)	R_g/l (M=206)	a
0.000	0	5.86	6.86	8.47	0.515
0.343	200	6.44	7.55	9.32	0.516
0.858	500	7.39	8.68	10.73	0.520
1.373	800	8.32	9.80	12.15	0.528
1.888	1100	9.16	10.81	13.44	0.535
2.575	1500	10.02	11.87	14.80	0.544
3.434	2000	10.68	12.69	15.86	0.551
5.150	3000	11.17	13.30	16.66	0.557
8.584	5000	11.29	13.44	16.85	0.559
$\hat{r} = 0.20$					
$\hat{\epsilon}$	$N_A\epsilon$ (cal/mole)	R_g/l (M=101)	R_g/l (M=137)	R_g/l (M=206)	a
0.000	0	6.32	7.49	9.41	0.557
0.343	200	6.82	8.08	10.14	0.552
0.858	500	7.66	9.07	11.35	0.548
1.373	800	8.52	10.09	12.62	0.547
1.888	1100	9.30	11.04	13.81	0.551
2.575	1500	10.12	12.04	15.09	0.556
3.434	2000	10.76	12.83	16.11	0.562

| 5.150 | 3000 | 11.24 | 13.41 | 16.88 | 0.567 |
| 8.584 | 5000 | 11.35 | 13.56 | 17.06 | 0.567 |

$$\hat{r} = 0.30$$

$\hat{\epsilon}$	$N_A\epsilon$ (cal/mole)	R_g/l (M=101)	R_g/l (M=137)	R_g/l (M=206)	a
0.000	0	6.95	8.32	10.59	0.588
0.343	200	7.37	8.81	11.18	0.581
0.858	500	8.08	9.65	12.22	0.577
1.373	800	8.83	10.54	13.33	0.574
1.888	1100	9.55	11.40	14.40	0.573
2.575	1500	10.31	12.32	15.56	0.573
3.434	2000	10.92	13.06	16.51	0.577
5.150	3000	11.37	13.62	17.22	0.579
8.584	5000	11.47	13.75	17.39	0.580

$$\hat{r} = 0.50$$

$\hat{\epsilon}$	$N_A\epsilon$ (cal/mole)	R_g/l (M=101)	R_g/l (M=137)	R_g/l (M=206)	a
0.000	0	7.67	9.25	11.91	0.615
0.343	200	8.00	9.64	12.35	0.605
0.858	500	8.60	10.34	13.22	0.600
1.373	800	9.25	11.10	14.15	0.594
1.888	1100	9.87	11.85	15.09	0.591
2.575	1500	10.56	12.69	16.14	0.590
3.434	2000	11.12	13.37	16.99	0.591
5.150	3000	11.54	13.89	17.68	0.595
8.584	5000	11.65	14.01	17.85	0.595

$$\hat{r} = 0.15$$

$\hat{\epsilon}$	$N_A\epsilon$ (cal/mole)	R_g/l (M=101)	R_g/l (M=137)	R_g/l (M=206)	a
2.747	1600	10.22	12.13	15.16	0.550

Table 5. The gyration radius, R_g, and free volume fraction, α, of a two-parameter self-avoiding chain at different ratio of segment length to radius of hard-sphere. The angle between two segments is fixed at 109.5°. The values of two parameters are chosen as $\hat{r} = 0.15$ and $\hat{\epsilon} = 2.747$.

		$l\,/\,R=$ 0.009800		
ϕ	R_g/R	error	α	error
0.01	0.20026	0.00000	0.98403	0.00001
0.02	0.20028	0.00000	0.96830	0.00002
0.03	0.20020	0.00000	0.95269	0.00002
0.05	0.20007	0.00002	0.92185	0.00003

		$l\,/\,R=$ 0.019377		
ϕ	R_g/R	error	α	error
0.01	0.39596	0.00004	0.97693	0.00002
0.02	0.39572	0.00003	0.95432	0.00002
0.03	0.39555	0.00003	0.93201	0.00003
0.05	0.39512	0.00002	0.88813	0.00003

		$l\,/\,R=$ 0.02744		
ϕ	R_g/R	error	α	error
0.01	0.56053	0.00004	0.97035	0.00002
0.02	0.56007	0.00003	0.94141	0.00002
0.03	0.55993	0.00003	0.91303	0.00003
0.05	0.55909	0.00004	0.85780	0.00004
0.10	0.55701	0.00003	0.72659	0.00005
0.20	0.55270	0.00004	0.49529	0.00005
0.30	0.54837	0.00005	0.31058	0.00005
0.40	0.54527	0.00008	0.17835	0.00004

$l / R=$ 0.049000				
ϕ	R_g/R	error	α	error
0.01	1.00056	0.00007	0.95046	0.00002
0.02	0.99944	0.00006	0.90341	0.00003
0.03	0.99846	0.00005	0.85822	0.00004
0.05	0.99654	0.00007	0.77273	0.00004

$l / R=$ 0.073500				
ϕ	R_g/R	error	α	error
0.01	1.50034	0.00007	0.92567	0.00003
0.02	1.49896	0.00009	0.85755	0.00004
0.03	1.49696	0.00010	0.79396	0.00004
0.05	1.49374	0.00011	0.67914	0.00005

$l / R=$ 0.098000				
ϕ	R_g/R	error	α	error
0.01	2.00075	0.00011	0.90000	0.00003
0.02	1.99860	0.00012	0.81176	0.00004
0.03	1.99647	0.00013	0.73230	0.00004
0.05	1.99418	0.00013	0.59554	0.00005

$l / R=$ 0.122500				
ϕ	R_g/R	error	α	error
0.01	2.50086	0.00015	0.87444	0.00003
0.02	2.49926	0.00015	0.76803	0.00004
0.03	2.49816	0.00016	0.67553	0.00005
0.05	2.49772	0.00020	0.52419	0.00005

Table 6. The free volume fraction, α, of a AO-sphere in the presence of hard-spheres with square-well attractive potential. q is the ratio between AO-sphere radius and hard-sphere radius. The reduced range of the potential well is λ and the reduced depth of the potential well is $\hat{\varepsilon}$.

			$q = 0.25$		
λ	$\hat{\varepsilon}$	$\phi = 0.1$	$\phi = 0.2$	$\phi = 0.3$	$\phi = 0.4$
		α	α	α	α
1.05	0.00	0.80633	0.61748	0.43769	0.27398
1.05	0.20	0.80632	0.61780	0.43861	0.27513
1.05	0.40	0.80642	0.61817	0.43937	0.27681
1.05	0.80	0.80665	0.61909	0.44149	0.28027
1.05	1.20	0.80706	0.62046	0.44440	0.28467
1.05	1.60	0.80747	0.62228	0.44788	0.28958
1.05	2.00	0.80808	0.62466	0.45215	0.29532
1.05	2.50	0.80943	0.62857	0.45907	0.30408
1.15	0.00	0.80625	0.61746	0.43764	0.27400
1.15	0.25	0.80652	0.61839	0.43976	0.27687
1.15	0.50	0.80681	0.61947	0.44188	0.27992
1.15	0.75	0.80717	0.62069	0.44436	0.28316
1.15	1.00	0.80754	0.62226	0.44692	0.28651
1.15	1.25	0.80806	0.62392	0.44985	0.29027
1.15	1.50	0.80864	0.62581		0.29408
1.25	0.00	0.80628	0.61745	0.43771	0.27391
1.25	0.20	0.80645	0.61838	0.43936	0.27601
1.25	0.40	0.80683	0.61929	0.44107	0.27808
1.25	0.60	0.80706	0.62042	0.44304	0.28018
1.25	0.80	0.80748	0.62159	0.44505	0.28251
1.25	1.00	0.80788	0.62293	0.44719	0.28476
1.25	1.25	0.80853	0.62486	0.45014	0.28807
1.35	0.00	0.80618	0.61749	0.43774	0.27397
1.35	0.20	0.80658	0.61830	0.43911	0.27527
1.35	0.40	0.80684	0.61923	0.44048	0.27643
1.35	0.60	0.80710	0.62021	0.44204	0.27775
1.35	0.80	0.80753	0.62135	0.44367	0.27920
1.35	1.00	0.80802	0.62263	0.44561	0.28069
1.35	1.20	0.80852	0.62425	0.44760	0.28227
1.50	0.00	0.80624	0.61741	0.43780	0.27400
1.50	0.12	0.80642	0.61789	0.43815	0.27392
1.50	0.24	0.80660	0.61822	0.43849	0.27386

1.50	0.36	0.80668	0.61867	0.43892	0.27373
1.50	0.48	0.80688	0.61916	0.43943	0.27361
1.50	0.60	0.80702	0.61962	0.43980	0.27364
1.50	0.75	0.80736	0.62032	0.44063	0.27358
1.70	0.00	0.80621	0.61743	0.43770	0.27392
1.70	0.10	0.80638	0.61767	0.43764	0.27345
1.70	0.20	0.80641	0.61789	0.43761	0.27289
1.70	0.30	0.80659	0.61804	0.43758	0.27220
1.70	0.40	0.80664	0.61828	0.43761	0.27156
1.70	0.50	0.80682	0.61851	0.43769	0.27116
1.70	0.55	0.80691	0.61861	0.43760	0.27076
2.00	0.00	0.80632	0.61744	0.43765	0.27394
2.00	0.06	0.80635	0.61755	0.43757	0.27357
2.00	0.12	0.80643	0.61759	0.43740	0.27308
2.00	0.18	0.80641	0.61759	0.43705	0.27283
2.00	0.24	0.80641	0.61761	0.43692	0.27243
2.00	0.30	0.80650	0.61758	0.43671	0.27200
2.00	0.35	0.80659	0.61762	0.43660	0.27180

$q = 0.50$					
λ	$\hat{\varepsilon}$	$\phi = 0.1$	$\phi = 0.2$	$\phi = 0.3$	$\phi = 0.4$
		α	α	α	α
1.05	0.00	0.67834	0.40106	0.18938	0.06011
1.05	0.20	0.67872	0.40219	0.19103	0.06165
1.05	0.40	0.67913	0.40335	0.19307	0.06334
1.05	0.80	0.67998	0.40657	0.19809	0.06767
1.05	1.20	0.68148	0.41096	0.20448	0.07330
1.05	1.60	0.68350	0.41707	0.21296	0.08039
1.05	2.00	0.68616	0.42492	0.22387	0.08955
1.05	2.50	0.69125	0.43846	0.24192	0.10430
1.25	0.00	0.67840	0.40103	0.18932	0.06012
1.25	0.20	0.67969	0.40492	0.19446	0.06373
1.25	0.40	0.68114	0.40912	0.20005	0.06762
1.25	0.60	0.68276	0.41387	0.20618	0.07185
1.25	0.80	0.68480	0.41919	0.21277	0.07651
1.25	1.00	0.68713	0.42512	0.22045	0.08189
1.25	1.25	0.69059	0.43417	0.23124	0.08934
1.35	0.00	0.67831	0.40101	0.18938	0.06017
1.35	0.20	0.67997	0.40556	0.19489	0.06329
1.35	0.40	0.68175	0.41033	0.20075	0.06692
1.35	0.60	0.68391	0.41586	0.20713	0.07087
1.35	0.80	0.68632	0.42193	0.21444	0.07517

1.35	1.00	0.68912	0.42922	0.22256	0.08018
1.35	1.20	0.69266	0.43796	0.23248	0.08583
1.50	0.00	0.67842	0.40105	0.18938	0.06008
1.50	0.12	0.67934	0.40373	0.19224	0.06124
1.50	0.24	0.68055	0.40667	0.19521	0.06255
1.50	0.36	0.68193	0.40985	0.19823	0.06382
1.50	0.48	0.68332	0.41304	0.20160	0.06516
1.50	0.60	0.68475	0.41656	0.20528	0.06669
1.50	0.75	0.68701	0.42175	0.21028	0.06879
1.70	0.00	0.67834	0.40105	0.18934	0.06015
1.70	0.10	0.67928	0.40297	0.19055	0.05997
1.70	0.20	0.68032	0.40495	0.19194	0.05981
1.70	0.30	0.68129	0.40720	0.19345	0.05974
1.70	0.40	0.68254	0.40948	0.19498	0.05963
1.70	0.50	0.68375	0.41197	0.19661	0.05938
1.70	0.55	0.68442	0.41343	0.19765	0.05930
2.00	0.00	0.67843	0.40102	0.18931	0.06018
2.00	0.06	0.67878	0.40151	0.18895	0.05928
2.00	0.12	0.67929	0.40207	0.18855	0.05852
2.00	0.18	0.67975	0.40261	0.18820	0.05764
2.00	0.24	0.68028	0.40324	0.18783	0.05679
2.00	0.30	0.68093	0.40406	0.18751	0.05598
2.00	0.35	0.68135	0.40463	0.18727	0.05531

$q = 1.00$					
λ	$\hat{\varepsilon}$	$\phi = 0.1$	$\phi = 0.2$	$\phi = 0.3$	$\phi = 0.4$
		α	α	α	α
1.05	0.00000	0.3757	0.08422	0.00752	0.00012
1.05	0.20000	0.37673	0.08542	0.00792	0.00013
1.05	0.40000	0.37785	0.08688	0.0084	0.00016
1.05	0.60000	0.37916	0.08862	0.00894	0.00019
1.05	0.80000	0.38064	0.09066	0.00966	0.00023
1.05	1.00000	0.38258	0.09317	0.01053	0.00029
1.05	1.20000	0.38472	0.097	0.0116	0.00036
1.15	0.00000	0.37578	0.08421	0.00752	0.00012
1.15	0.30000	0.37982	0.08944	0.00911	0.00019
1.15	0.60000	0.38504	0.09613	0.01131	0.00032
1.15	0.90000	0.39184	0.10458	0.01435	0.00054
1.15	1.20000	0.40045	0.11561	0.01886	0.00097
1.15	1.50000	0.41172	0.13053	0.02533	0.00183

1.25	0.00000	0.37578	0.08419	0.00751	0.00011
1.25	0.20000	0.37979	0.08924	0.00893	0.00017
1.25	0.40000	0.38462	0.09523	0.01077	0.00026
1.25	0.60000	0.39054	0.10208	0.01307	0.00040
1.25	0.80000	0.39705	0.11042	0.01604	
1.25	1.00000	0.40505	0.12048	0.01990	0.00098
1.25	1.25000	0.41735	0.13638	0.02698	0.00174
1.50	0.00000	0.37578	0.08418	0.00753	0.00012
1.50	0.12000	0.38008	0.08883	0.00859	0.00015
1.50	0.24000	0.38461	0.09386	0.00985	0.00019
1.50	0.36000	0.38985	0.09953	0.01132	0.00024
1.50	0.48000	0.3955	0.10588	0.01313	0.0003
1.50	0.60000	0.40165	0.11304	0.0153	0.00039
1.50	0.75000	0.41057	0.12351	0.01878	0.00055
1.70	0.00000	0.37574	0.08419	0.00752	0.00012
1.70	0.10000	0.38015	0.08851	0.00832	0.00013
1.70	0.20000	0.38492	0.09313	0.00925	0.00014
1.70	0.30000	0.39023	0.09828	0.01037	0.00016
1.70	0.40000	0.39572	0.10409	0.01165	0.00019
1.70	0.50000	0.4019	0.11055	0.01319	0.00022
1.70	0.55000	0.40527	0.1142	0.01406	0.00023
2.00	0.00000	0.37565	0.08417	0.00751	0.00012
2.00	0.06000	0.37877	0.08662		
2.00	0.12000	0.38194	0.08909	0.0081	0.00011
2.00	0.18000	0.38542	0.09181		
2.00	0.24000	0.38894	0.0947	0.00879	0.00011
2.00	0.30000	0.39296	0.09796		
2.00	0.35000	0.39629	0.10097	0.00963	0.00011

List of References:

1. Dumetz, A. C.; Chockla, A. M.; Kaler, E. W.; Lenhoff, A. M. *Biophys. J.* **2008**, *94*, 570-583.

2. McPherson, A. *Crystallization of Biological Macromolecules*; CSHL Press: Cold Spring Harbor, **1999**.

3. Thomson, J. A.; Schurtenberger, P; Thurston, G. M.; Benedek, G. B. *Natl. Acad. Sci. USA* **1987**, *84*, 7079-7083.

4. Annunziata, O.; Ogun, O.; Benedek, G. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 970-974.

5. Galkin, O.; Chem, K.; Nagel, R. L.; Hirsch, R. E.; Vekilov, P. G. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8479-8483.

6. Liu, C.; Pande, J.; Lomakin, A.; Ogun, O.; Benedek, G. B. *Invest. Ophthal. and Vision Sci.*, **1998**, *39*,1609-1619.

7. Family, F.; Landau, D.P. *Kinetics of Aggregation and Gelation* **1984**, Elsevier Science Publishers.

8. Pande, A.; Annunziata, O.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. *Biochemistry* **2005**, *99*, 2491-2500.

9. Stradner, A..; Sedgwick, H.; Cardinaux, F.; Poon, W.C.K.; Egelhaaf, S.U.; Schurtenberger, P. *Nature* **2004**, *432*, 492-496.

10. Patil, G.V. *Drug Dev. Res.* **2003**, *58*, 219-247.

11. Gunton, J. D.; Shiryayev, A.; Pagan, D. L. *Protein Condensation: Kinetic Pathways to Crystallization and Disease* 2007, Cambridge University Press.

12. Ferdinger, E. A.; Monien, B. H.; Urbanc, B.; Lomakin, A.; Tan, M.; Li, H.; Spring, S. M.; Condron, M. M.; Cruz, L.; Xie, C.; Benedek, G. B.; Bitan, G. *Proc. Natl. Acad. Sci. USA* **2008**, *37*, 14175–14180.

13. Vekilov, P. G. *AIChE J.* **2008**, *54*, 2508-2515.

14. Stegemann, J. P.; Kaszuba, S. N.; Rowe, S. L. *Tissue Eng.* **2007**, *13*, 2601-2613.

15. Margolin, A. L.; Navia, M. A. *Angew. Chem. Int. Ed.* **2001**, 40, 2204-2222.

16. Freytes, D. O.; Martin, J.; Velankar, S. S.; Lee, A. S.; Badylak, S. F. *Biomaterials* **2008**, *29*, 1630-1637.

17. Laurent, N.; Haddoub, R.; Flitsch, S. L. *Trends in Biotechnology* **2008**, *26*, 328-337.

18. Kaplan, D.; McGrath, K. Protein-Based Materials (Bioengineering of Materials) **1997**, Birkhäuser Boston.

19. Ahuja, S. K.; Ferreira, G. M.; Moreira, A. R. *Critial Rev. in Biot.* **2004**, *24* (2-3), 125-154.

20. Park, H.; Nam, S.; Lee, J. K.; Yoon, C. N.; Mannervik, B.; Benkovic, S. J.; Kim, H. *Science* **2006**, *311*, 535-538.

21. Jermutus, L.; Honegger, A.; Schwesinger, F.; Hanes, J.; Plückthun, A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 75-80.

22. De Wolf, F. A.; Brett, G. M. *Pharmacol. Rev.* **2000**, *52*, 207-236.

23. Ayala, M.; Vasquez-Duhalt, R. Enzymatic Catalysis on Petroleum Products, in Studies in Surface Science and Catalysis, *Vol. 151*; Vasquez-Duhalt, R., Quintero-Ramirez, R., Eds.; Elsevier: Amsterdam, **2004**; pp 67-111.

24. Abraham, T. E.; Roy, J. J. *Chem. Rev.* **2004**, *104*, 3705-3721.

25. Schoevaart, R.; Wolbers, M. W.; Golubovic, M.; Ottens, M.; Kieboom, A. P. G.; van Rantwijk, F.; van der Wielen, L. A. M.; Sheldon, R. A. *Biotechnol. Bioeng.* **2004**, *87*, 754−762.

26. Langer, K.; Balthasar, S.; Vogel, V.; Dinauer, N.; von Briesen, H.; Schubert, D. *Int. J. Pharm.* **2003**, *257*, 169−180.

27. Wong, S. S. *Chemistry of Protein Conjugation and Cross-Linking*; CRC Press: Boca Raton, FL, **1993**.

28. Kluger, R.; Alagic, A. *Bioorg. Chem.* **2004**, *32*, 451−472.

29. Oosawa, F.; Asakura, S. *Thermodynamics of the Polymerization of Protein*; Academic Press: New York, **1997**.

30. Aprelev, A.; Weng W. J.; Zakharov, M.; Yosmanovich, D.; Kwong, S.; Briehl, R. W.; Ferrone, F. A. *J. Mol. Biol.* **2007**, *369*, 1170−1174.

31. Galkin, O.; Nagel, R. L.; Vekilov, P. G. *J. Mol. Biol.* **2007**, *365*, 425-439.

32. Chen, Q.; Vekilov, P. G.; Nagel, R. L.; Hirsch, R. E. *Biophys. J.* **2004**, *86*, 1702−1712.

33. Giurleo, J. T.; He, X. L.; Talaga, D. S. *J. Mol. Biol.* **2008**, *381*, 1332-1348.

34. Chang, S. C.; Lee, M. J.; Lin, H. M. *Chem. Eng. J.* **2005**, *139*, 416−425.

35. Meijer, J. T; Roeters, M.; Viola, V.; Lowik, D. W. P. M.; Vriend, G.; van Hest, J. C. M. *Langmuir* **2007**, *23*, 2058−2063.

36. Colaco, M.; Park, J.; Blanch, H. *Biophys.* **2008**, *136*, 74-76.

37. Alfred, C. *Separation Science and Technology* **1988**, *23* (8-9), 785-817.

38. Dubin, P. L.; Gao, J.; Mattison, K. Separation and Purification Methods **1994**, *23* (1), 1-16.

39. Mee, C.; Banki, M. R.; Wood, D. W. *Chem. Eng. J.* **2008**, *135* (1-2), 56-62.

40. Benedek, G. B. *Invest. Ophthalmol. Visual Sci.* **1997***, 38,* 1911–1921.

41. Pande, A.; Pande. J.; Asherie, N.; Lomakin, A.; Ogun, O., King, J.; Benedek, G. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6116–6120.

42. Steinberg, M. H.; Forget, B. G.; Higgs, D. R.; Nagel, R. L. *Disorders of Hemoglobin: Genetics, Pathology, Clinical Management,* Cambridge Univ. Press*, Cambridge, U.K.,* **2000**.

43. Selkoe, D. J. *J. Neuropathol. Exp. Neurol.* **1994**, *53,*438–447.

44. Benedek, G. B.; Pande, J.; Thurston, G. M.; Clark, J. I. *Prog. Ret. Eye. Res.* **1999**, *18,* 391–402.

45. Arakawa, T.; Kita, Y.; Niikura, T. *Current Med. Chem.* **2008**, *15*, 2086-2098.

46. Fernandez, A.; Alejandro, C. *Chem. Soc. Rev.* **2008**, *37*, 2373-2382.

47. Ducruix, A.; Giege, R. *Crystallization of Nucleic Acids and Proteins: A Practical Approach (Practical Approach Series)*; Oxford University Press: Oxford, 1999.

48. Finet, S.; Vivares, D.; Bonnete, F.; Tardieu, A. *Methods Enzymol.* **2003**, *368*, 105−129.

49. Chayen, N. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 577−583.]

50. Yau, S. T.; Vekilov, P. G. *Nature (London, U.K.)* **2000**, *406*, 494−497.

51. Galkin, O.; Vekilov, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6277−6281.

52. Gliko, O.; Neumaier, N.; Pan, W.; Haase, I.; Fischer, M.; Bacher, A.; Weinkauf, S.; Vekilov, P. G. *J. Am. Chem. Soc.* **2005**, *127*, 3433−3438.

53. Tanaka, H.; Nishikawa, Y. *Phys. Rev. Lett.* **2005**, *95*, 78103.

54. Taratuta, V. G.; Holschbach, A.; Thurston, G. M.; Blankschtein, D.; Benedek, G. B. *J. Phys. Chem.* **1990**, *94*, 2140−2144.

55. Annunziata, O.; Asherie, N.; Lomakin, A.; Pande, J.; Ogun, O.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14165−14170.

56. Bloustine, J.; Virmani, T.; Thurston, G. M.; Fraden, S. *Phys. Rev. Lett.* **2006**, *96*, 87803.

57. Wang, Y.; Annunziata, O. *J. Phys. Chem. B* **2007**, *111*, 1222−1230.

58. Annunziata, O.; Payne, A.; Wang, Y. *J. Am. Chem. Soc.* **2008**, *130*, 13347-13352.

59. Petsev, D. N.; Wu, X. X.; Galkin, O.; Vekilov, P. G. *J. Phys. Chem. B* **2003**, *107*, 3921−3926.

60. Lomakin, A.; Asherie, N.; Benedek, G. B. *J. Chem. Phys.* **1996**, *104*, 1646−1656.

61. Stradner, A.; Foffi, G.; Dorsaz, N.; Thurston, G.; Schurtenberger, P. *Phys. Rev. Lett.* **2007**, *99*, 198103.

62. Cuendet, M. A.; Michielin, O. *Biophys. J.* **2008**, *95*, 3575-3590.

63. Fasman, G. D. *Prediction of Protein Structure and the Principles of Protein Conformation,* Springer, **1989**.

64. Stites*, Chem. Rev.*, **1997**, *97*, 1233–1250.

65. Norel, R.; Sheinerman, F.; Petrey, D.; Honig, B. *Prot. Sci.* **2001**, *10*, 2147-2161.

66. Lomakin, A.; Asherie, N.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9465−9468.

67. Timasheff, S. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9721−9726.

68. McManus, J. J.; Lomakin, A.; Ogun, O.; Pande, A.; Basan, M.; Pande, J.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16856−16863.

69. Asherie, N.; Lomakin, A.; Benedek, G. B. *Phys. Rev. Lett.* **1996**, *77*, 4832-4835.

70. Ru, M. T.; Hirokane, S. Y.; Lo, A. S.; Dordick, J. S.; Reimer, J. A.; Clark, D. S. *J. Am. Chem. Soc.* **2000**, *122*, 1565-1582.

71. Abbott, N. L.; Blankschtein, D.; Hatton, T. A. *Macromolecules* **1991**, *24*, 4334-4341.

72. Arakawa, T.; Bhat, R.; Timasheff, S. N. *Biochemistry* **1990**, *29*, 1914-1923.

73. Annunziata, O.; Paduano, L.; Pearlstein, A. J.; Miller, D. G.; Albright, J. G. *J. Phys. Chem. B* **2006**, *110*, 1405-1415.

74. Curtis, R. A.; Ulrich, J.; Montaser, A.; Prausnitz, J. M.; Blanch, H. W. *Biotechnol. Bioeng.* **2002**, *79*, 367-380.

75. Tardieu, A.; Bonnete, F.; Finet, S.; Vivares, D. *Acta Cryst., Sect. D: Biol. Crystallogr.* **2002**, *58*, 1549-1553.

76. Tessier, P. M.; Lenhoff, A. M.; Sandler, S. I. *Biophys. J.* **2002**, *82*, 1620-1631.

77. Tardieu, A.; Le Verge, A.; Malfois, M.; Bonnete, F.; Finet, S.; Ries-Kautt, M.; Belloni, L. *J. Cryst. Growth* **1999**, *196*, 193-203.

78. Adams, M.; Fraden, S. *Biophys. J.* **1998**, *74*, 669-677.

79. Annunziata, O.; Paduano, L.; D. G.; Albright, J. G. *J. Phys. Chem. B* **2006**, *110*, 16139-16147.

80. Albertsson, P. Å. *Partition of Cell Particles and Macromolecules*; Wiley: New York, **1986**.

81. Roberts, M. J.; Bentley, M. D.; Harris, J. M. *Adv. Drug. Delivery Rev.* **2002**, *54*, 459-476.

82.   Kulkarni, A. M.; Chatterjee, A. P.; Schweizer, K. S.; Zukoski, C. F. *Phys. Rev. Lett.* **1999**, *83*, 4554.

83.   Hall, D.; Minton, A. P. *Biochim. Biophys. Acta* **2003**, *1649*, 127-139.

84.   Asakura, S.; Oosawa, F. *J. Chem. Phys.* **1954**, *22*, 1255-1256.

85.   Lekkerkerker, H. N. W.; Poon, W. C. K.; Pusey, P. N.; Stroobants, A.; Warren, P. B. *Europhys. Lett.* **1992**, *20*, 559-564.

86.   Ilett, S. M.; Orrock, A.; Poon, W. C. K.; Pusey, P. N. *Phys. Rev. E* **1995**, *51*, 1344-1352.

87.   Vivares, D.; Belloni, L; Tardieu, A.; Bonnete, F. *Eur. Phys. J. E* **2002**, *9*, 15-25.

88.   Ten Wolde, P. R.; Frenkel, D. *Science* **1997**, *277*, 1975-1978.

89.   Bhat, R.; Timasheff, S. N. *Protein Sci.* **1992**, *1*, 1133-1143.

90.   Anderson, V. J.; Lekkerkerker, H. N. W. *Nature* **2002**, *416*, 811-815.

91.   Broide, M. L.; Berland, C. R.; Pande, J.; Ogun, O.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5660-5664.

92.   Pande, A.; Pande, J.; Asherie, N.; Lomakin, A.; Ogun, O.; King, J. A.; Lubsen, N. H.; Walton, D.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1993-1998.

93.   Asakura, S.; Oosawa, F. *J. Poly. Sci.* **1958**, *XXXIII*, 183-192.

94.   Vrij, A. *Pure & Appl. Chem.* **1976**, *48*, 471-483.

95.   Tuinier, R.; Vliegenthart, G. A.; Lekkerkerker, H. N. W. J. Chem. Phys. 2000, 113, 10768-10775.

96.   Eisenriegler, E.; Hanke, A.; Dietrich, S. *Phys. Rev. E* **1996**, *54*, 1134.

97.   Kulkarni, A. M.; Chatterjee, A. P.; Schweizer, K. S. **2000**, *113*, 9863-9872.

98.   Meijer, E. J.; Frenkel, D. *J. Chem. Phys.* **1994**, *100*, 6873-6887.

99.   Renth, F.; Poon, W. C. K.; Evans, R. M. L. *Phys. Rev. E* **2001**, *64*, 031402.

100. Nephew, J. B.; Nihei, T. C.; Carter, S. A. *Phys. Rev. Lett.* **1998**, *80*, 3276−3279.

101. Tran-Cong, Q.; Harada, A. *Phys. Rev. Lett.* **1996**, *76*, 1162−1165.

102. Kyu, T.; Lee, J. H. *Phys. Rev. Lett.* **1996**, *76*, 3746−3749.

103. Williams, R. J. J.; Rozenberg, B. A.; Pascault, J. P. *Adv. Polym. Sci.* **1997**, *128*, 95−156.

104. Luo, K. *Eur. Polym. J.* **2006**, *42*, 1499−1505.

105. Wang, X.; Okada, M.; Matsushita, Y.; Furukawa, H.; Han, C. C. *Macromolecules* **2005**, *38*, 7127−7133.

106. Kimura, K.; Kohama, S.-I.; Yamazaki, S. *Polym. J.* **2006**, *38*, 1005−1022.

107. Sperling, L. H. *Polymeric Multicomponent Materials: An Introduction*; John Wiley and Sons: New York, 1997.

108. Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 330−335.

109. Annunziata, O.; Pande, A.; Pande, J.; Ogun, O.; Lubsen, N. H.; Benedek, G. B. *Biochemistry* **2005**, *44*, 1316−1328.

110. Asherie, N. *The phase diagram of globular protein solution: the role of the range of interaction*, Ph. D. Dissertation, MIT, **1998**.

111. Frenkel, D. *Science* **2006**, *314*, 768-769.

112. Gasser, U.; Weeks, E. R.; Schofield, A.; Pusey, P. N.; Weitz, D. A. *Science* **2001**, *292*, 258-265.

113. Kwon, H. J.; Tanaka, Y.; Kakugo, A.; Shikinaka, K.; Furukawa, H.; Osada, Y.; Gong, J. P. *Biochemistry* **2006**, *45*, 10313-10318.

114.   Flachmann, R.; Kühlbrandt, W. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14966.

115.   Frenkel, D. *Nature Materials* **2006,** *5*, 1.

116.   Tanford, C. *Physical Chemistry of Macromolecules*; Wiley: New York, **1961**.

117.   Stockmayer, W. H. *J. Chem. Phys.* **1950**, *18*, 58-61.

118.   Lebowitz, J. L. *Phys. Rev. A*, **1964**, *133*, 895-899.

119.   Sear, R. P. *J. Chem Phys.* **1999**, *111*, 4800-4806.

120.   Muschol, M.; Rosenberger, F. *J. Chem. Phys.***1997**, *107*, 1953-1957.

121.   Edmond, E. ; Ogston, A. G. Biochem. J. 1968, 109, 569-576.

122.   Warren, P. B.; Ilett, S. M.; Poon, W. C. K. Phys. Rev. E 1995, 5205-5213.

123.   Lekkerkerker, H. N. W. *Coll. and Surf.* **1990**, *51*, 419-426.

124.   Louis, A. A.; Bolhuis, P. G.; Meijer, E. J. Hansen, J. P. *J. Chem. Phys.* **2002**, *116*, 10547-10556.

125.   Taniguchi, T.; Kawakatsu, T.; Kawasaki, K. Slow dynamics in condensed matter. AIP Conference Proceedings **1992**, *256*, 503-504.

126.   Chapman, W. G.; Jackson, G.; Gubbins, K. E. *Mol. Phys.* **1988**, *65*, 1057-1079.

127.   Wertheim, M. S. *J. Stat. Phys.***1984**, *35*, 19-34.

128.   Wertheim, M. S. *J. Stat. Phys.***1984**, *35*, 35-47.

129.   Wertheim, M. S. *J. Stat. Phys.***1986**, *42*, 459-476.

130.   Wertheim, M. S. *J. Stat. Phys.***1986**, *42*, 477-492.

131.   Wertheim, M. S. *J. Chem. Phys.***1986**, *85*, 2929-2936.

132.   Banaszak, M.; Chiew, Y. C.; Radosz, M. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **1993**, *48*, 3760−3765.

133. Pande, J.; Lomakin, A.; Fine, B.; Ogun, O.; Sokolinski, I.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1067-1071.

134. Asherie, N.; Pande, J.; Lomakin, A.; Ogun, O.; Hanson, S. R. A.; Smith, J. B.; Benedek, G. B. *Biophys. Chem.* **1998**, *75*, 213−227.

135. Sela, M.; Lifson, S. *Biochimica et Biophysica Acta* **1959**, *36*, 471-478.

136. Thornton, J. M. *J. Mol. Bio.* **1981**, *151*, 261-287.

137. Thannhauser, T. W.; Konishi, Y.; Scheraga, H. A. *Anal. Biochem.* **1984**, *138*, 181-188.

138. Futami, J.; Tada, H.; Seno, M.; Ishikami, S.; Yamada, H. *J. Biochem.* **2000**, *128*, 245-250.

139. Albertsson, P. Å. *Partition of Cell Particles and Macromolecules*; Wiley: New York, **1986**.

140. Hermanson, G. T. *Bioconjugate Techniques*, Academic Press, Inc., **1995**.

141. Wong, S. S. *Chemistry of Protein Conjugation and Cross-linking*, CRC Press, Inc., **1991**.

142. Waugh, S. M.; Dibella, E. E.; Pilch, P. F. *Biochemistry* **1989***, 28*, 3448-3455.

143. Schneider, C.; Newman, R. A.; Sutherland, D. R.; Asser, U.; Greaves, M. F. *J. Biol. Chem.* **1982**, *257*, 10766−10769.

144. Chen, L.L.; Rosa, J.J.; Turner, S.; Pepinsky, R.B. *J. Biol. Chem.* **1991**, *266*, 18237−18243.

145. Taniuchi, M.; Clark, H. B.; Johnson, E. M. Jr. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 4094-4098.

146.   Migneault, I.; Dartiguenave, C.; Bertrand, M. J.; Waldron, K. C. *BioTechniques* **2004**, *37*, 790-802.

147.   Quiocho, F.A.; Richards, F.M. *Proc. Natl. Acad. Sci.* **1964**, *52*, 833–839.

148.   Yonath, A.; Sielecki, J.M.; Moult, J.; Podjarny, A.; Traub, W. *Biochemistry* **1977**, *16*, 1413–1417.

149.   Fitzpatrick, P.A.; Steinmetz, A.C.U.; Ringe, D.; Klibanov, A.M. *Proc. Natl. Acad. Sci.* **1993**, *90*, 8653–8657.

150.   Lusty, C.J. *J. Appl. Cryst.* **1999**, *32*, 106–112.

151.   Monsan, P.; Puzo, G.; Mazarguil, H. Biochimie **1975**, *57*, 1281-1292.

152.   Richards, F. M.; Knowles, J. R. *J. Mol. Biol.* **1968**, *37*, 231-233.

153.   Hardy, P. M.; Nicholls, A. C.; Rydon, H. N. *J. Chem. Soc. Perkin Trans. I*, **1976**, 958-962.

154.   Hardy, P. M.; Hughes, G. J.; Rydon, H. N. *J. Chem. Soc. Perkin Trans. I*, **1979**, 2282-2288.

155.   Lubig, R.; Kusch, P.; Roper, K.; Zahn, H. *Monatsh. Chem.* **1981**, *112*, 1313-1323.

156.   Strynadka, N. C. J.; James, M. N. G. *J. Mol. Biol.* **1991**, *220*, 401-424.

157.   Jolles, P. *Angew. Chem. Int. Ed.* **1969**, *4*, N227-239.

158.   Paduano, L.; Annunziata, O.; Pearlstein, A. J.; Miller, D. G.; Albright, J. G. *J. Cryst. Growth* **2001**, *232*, 273-284.

159.   Annunziata, O.; Paduano, L.; Pearlstein, A. J.; Miller, D. G.; Albright, J. G. Use *J. Am. Chem. Soc.* **2000**, *122*, 5916-5928.

160.   Wang, J.; Dauter, M.; Alkire, R.; Joachimiak, A.; Dauter, Z. *Acta. Crystallogr. Sect. D* **2007**, 63, 1254-1268.

161. Towbin, H.; Staehelin, T.; Gordon, J. *Proc. Natl. Acad. Sci. U S A.* **1979**, 76, 4350-4354.

162. Gold. *Organic Compounds in Biological Systems, 2nd ed.* John Wiley & Sons, Inc., **1990**.

163. Curry, S. *Vox Sang.* **2003***, 83 Suppl 1*, 315–319.

164. Kopp, J.; Schwede, T. *Nucleic Acids Res*. **2004**, *32*, D230-D234.

165. Kopp, J.; Schwede, T. *Nucleic Acids Res*. **2006**, *34*, D315-D318.

166. Solomon, E. I.; Sundaram, U. M.; Machonkin, T. E. *Chem. Rev.* **1996**, *96*, 2563-2606.

167. Bonner, W. D., Jr. *Annu. Rev. Plant. Physiol.* **1957**, *8*, 427-452.

168. Yaropolov, A. I.; Skorobogat'ko, O. V.; Vartanov, S. S.; Varfolomeyev, S. D. *Appl. Biochem. Biotech.* **1994**, *49*, 256-280.

169. Xu, F. *Industrial Biotech.* **2005**, *1*, 38-50.

170. Gouto, S. R.; Herrera, J. L. T. Biotech. Advances **2006**, *24*, 500-513.

171. Barton, S. C.; Kim, H.; Binyamin, G.; Zhang, Y.; Heller, A. *J. Phys. Chem. B*, **2001**, *105*, 11917–11921.

172. Kim, Y.; Cho, N.; Eom, T.; Shin, W. *Bull. Korean Chem. Soc.* **2002**, *23*, 985-989.

173. Piontek, K.; Antorini, M.; Choinowski, T. *J. Biol. Chem.* **2002**, *277*, 37663-37669.

174. Dean J. F. D.; Eriksson K. E. L. *Holzforschung* **1994**, *48*, 21–33.

175. Baldrian, P. *FEMS Microbiol Rev.* **2006**, *30*, 215-242.

176. Sakurai, T.; Kataoka, K. *The Chemical Record* **2007**, *30,* 220 – 229.

177. Madzk, C.; Mimmi, M. C.; Caminade, E.; Brault, A.; Baumberger, S.; Briozzo, P.; Mougin, C.; Jolivalt, C. *Prot. Eng., Design & Selection* **2006**, *19*, 77–84.

178. Koudelka, G. B.; Hansen, F. B.; Ettinger, M. J. *J. Biol. Chem.* **1985**, *260*, 15561-15565.

179. d'Acunzo, F.; Barreca, A. M.; Galli, C. *J. Mol. Catal. B: Enzymatic* **2004**, *31*, 25–30.

180. Zille, A.; Munteanu, F.; Gübitz, G. M.; Cavaco-Paulo, A. *J. Mol. Catal. B: Enzymatic* **2005**, *33*, 23-28.

181. Steele, J. C. H.; Tanford, C.; Reynolds, J. A. *Methods Enzymol.* **1978**, *48*, 11-23.

182. Foster, J. F.; Sterman, M. D. *J. Am. Chem. Soc.* **1956**, *78*, 3656.

183. Sophianopoulos, A. J.; Rhodes, C. K.; Holcomb, D. N.; van Holde, K. E. *J. Biol. Chem.* **1962**, *237*, 1107−1112.

184. Liu, C.; Asherie, N.; Lomakin, A.; Pande, J.; Ogun, O.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 377.

185. Lomakin, A.; Teplow, D. B.; Benedek, G. B. *Quasielastic Light Scattering for Protein Assembly Studies,* Amyloid Proteins: Methods and Protocols. Methods in Molecular Biology, vol. 299, Edited by: E. M. Sigurdsson, Humana Press Inc.: Totowa, NJ, **2005**.

186. Berne, B. J.; Pecora, R. *Dynamic Light Scattering with Applications to Chemistry, Biology, and Physics*. Dover Publications, inc.: Mineola, NY, **2000**.

187. Annunziata, O.; Buzatu, D.; Albright, J. G. *Langmuir* **2005**, *21*, 12085-12089.

188. Born, M.; Wolf, E. *Principles of Optics*. Cambridge University Press. **1997**.

189. Reiss, H.; Frisch, H. L.; Lebowitz, J. L. *J. Chem. Phys.* **1959**, *31*, 369-380.

190. Reiss, H.; Frisch, H. L.; Helfand, E.; Lebowitz, J. L. *J. Chem. Phys.* **1960**, *32*, 119-124.

191. Lebowitz, J. L.; Helfand, E.; Praestgaard, E. *J. Chem. Phys.* **1965**, *43*, 774-779.

192. Lekkerkerker, H. N. W.; Stroobants, A. *Physica A* **1993**, *195*, 387-397.

193. Attard, P. *J. Chem. Phys.* **1993**, *98*, 2225-2231.

194. Sastry, S.; Truskett, T. M.; Debenedetti, P. G.; Torquato, S.; Stillinger, F. H. *Mol. Phys.* **1998**, *95*, 289-297.

195. Krekelberg, W. P.; Ganesan, V.; Truskett, T. M. *J. Chem. Phys.* **2006**, *124*, 214502.

196. Amokrane, S.; Regnaut, C. *Phys. Rev. E* **1996**, *53*, 1990-1993.

197. Kalos, M. H.; Whitlock, P. A. *Monte Carlo Methods, Vol I: Basics*, John Wiley & Sons: New York, **1986**.

198. Barrat, J.; Hansen, J. *Basic Concepts for Simple and Complex Liquids*, University Press: Cambridge, **2003**.

199. Vergara, A.; Capuano, F.; Paduano, L.; Sartorio, R. *Macromolecules* **2006**, *39*, 4500.

200. Gast, A. P.; Hall, C. K.; Russell, W. B. *J. Colloid Interface Sci.* **1983**, *96*, 251-267.

201. Baker, J. A.; Henderson, D. *J. Chem. Phys.* **1967**, *47*, 2856-2861.

202. Baker, J. A.; Henderson, D. *J. Chem. Phys.* **1967**, *47*, 4714-4721.

203. Leonard, P. J.; Baker, J. A.; Henderson, D. *Trans. Faraday Soc.* , **1970**, *66*, 2439 – 2452.

204.   Carnahan, N. F.; Starling, K. E. *J. Chem. Phys.* **1969**, *51*, 635-636.

205.   Yethiraj, A.; Carol, K. H. *J. Chem. Phys.* **1991**, *95*, 8494-8506.

206.   Reichl, L. E. *A Modern Course in Statistical Physics*; University of Texas Press: Austin, TX, 1980.

207.   Kulkarni, A. M.; Chatterjee, A. P.; Schweizer, K. S.; Zukoski, C. F. *J. Chem. Phys.* **2000**, *113*, 9863-9873.

208.   Mansoori, G. A.; Carnahan, N. F.; Starling, K. E.; Leland, T. W. *ibid.* **1971**, *54*, 1523.

209.   Andrews, F. C.; Ellerby, H. M. *J. Chem. Phys.* **1981**, *75*, 3542-3552.

210.   Speedy, R. J. *J. Chem. Phys.* **1982**, *77*, 2702-2704.

211.   Oie, T.; Topol, I. A.; Burt, S. K. *J. Phys. Chem.*, **1994**, 98, 1121–1128.

212.   Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, **1953**.

213.   Kanesaka, I.; Snyder, R. G.; Strauss, H. L. *J. Chem. Phys.* **1986**, *84*, 395-397.

214.   Kashchiev, D. *Nucleation: Basic Theory with Applications*; Butterworth−Heinemann: Oxford, **2000**.

215.   Tanford, C.; Wagner, M. L. *J. Am. Chem. Soc.* **1954**, *76*, 3331−3336.

216.   Baxter, R. J. *J. Chem. Phys.* **1968**, *49*, 2770−2774.

217.   Roy, J. J.; Abraham, T. E. *J. Mol. Catal. B: Enzymatic* **2006**, *38*, 31-36.

218.   Cabana, H.; Jones, J. P.; Agathos, S. N. J. Biotech. **2007**, 132, 23-31.

219.   Shah, S. S.; Sharma, A.; Gupta, M. N. *Anal. Biochem.* **2006**, *351*, 207-203.

VITA

Ying Wang was born in Anqin city, Anhui, China at 10:10 am on October 4th, 1980. He then moved to Chongqing, China in 1981. He graduated from Yucai high school and attended Tsinghua University, Beijing, China in September, 1999. He received his BS degree in chemistry with a thesis on "Determination of Herbal Medicines using NIR and Clustering Analysis" on July, 2003. After one year's work as research associate in the analysis center at Tsinghua University, he went to Texas Christian University to pursue a Ph.D. degree in chemistry. In 2004, He joined Prof. Onofrio Annunziata's group as a graduate research associate investigating the phase behaviors of protein aqueous solution. During the graduate study, he was active in both teaching and research. He has served as teaching assistant in the physical laboratory class for four semesters. He also actively participated in the research poster expositions at TCU. He will earn his Ph.D. degree in chemistry with an emphasis in macromolecular and biophysical chemistry in December, 2008. After graduation, he is planning to become a famous scientist or a billionaire.

ABSTRACT

PROTEIN CONDENSATION IN CROWDED AQUEOUS SOLUTIONS

by Ying Wang, Ph.D., 2008
Department of Chemistry
Texas Christian University

Dissertation Advisor: Onofrio Annunziata, Assistant Professor of Chemistry

The formation of protein condensed phases (e.g., aggregates, gels and crystals) from aqueous solutions is important in materials science, biology, medicine and biotechnology. However this process is not well-understood and physicochemical studies on protein condensation are fundamental for all the above areas. Protein condensation induced in the presence macromolecular additives is known as macromolecular crowding. Although this process is not well-understood, it is believed that the main action of these additives is to enhance protein condensation by depletion interactions related to the size of macromolecules.

Polyethylene glycol (PEG), a hydrophilic nonionic polymer, is one of the most widely used macromolecular-crowding agents. Upon the addition of PEG with an average molecular weight of 1450 g/mol (PEG1450) to bovine serum albumin (BSA) aqueous solutions, protein condensation into oil-like spherical droplets is observed by lowering

temperature. This phenomenon is related to liquid−liquid phase separation (LLPS). To investigate the crowding effect of PEG on protein condensation, two thermodynamically independent properties of the LLPS phase boundary were experimentally characterized: (1) the effect of PEG1450 concentration on the LLPS temperature, (2) BSA/PEG1450 partitioning in the two liquid coexisting phases. A thermodynamic perturbation theory, which includes one parameter describing depletion interactions, was used to examine our experimental results. It was found that both independent properties of the phase boundary could be quantitatively described by the same value of the depletion-interaction parameter. However, the depletion-interaction parameter determined from our LLPS experiments was found to be significantly different from that theoretically obtained from the molecular size of BSA and PEG1450 by treating PEG as an ideal chain. The non-ideal behavior of the PEG chain was hypothesized as the main reason for the observed discrepancy. Thus, a more realistic model for the PEG chain was introduced. Monte-Carlo simulations were then performed to calculate novel values of the depletion-interaction parameter from the size of BSA and PEG1450. It is then shown that our new model for PEG chains significantly improves the agreement between experimental results and theory.

The role of protein oligomerization on protein condensation was also investigated. It is shown that LLPS of protein aqueous solutions can be isothermally induced by protein oligomerization. For BSA and lysozyme as protein models, protein oligomerization in the presence of glutaraldehyde as a chemical cross-linker can evolve either towards the formation of cross-linked protein condensed droplets or macroscopic amorphous aggregation, depending on the LLPS properties of the protein monomer.

Finally, laccase condensation from aqueous solutions was investigated. Laccase is an enzyme used in the petroleum, paper and food industry. The thermodynamic behavior of laccase condensation and the morphology of laccase condensed phases was observed to strongly depend on the nature of the additive (PEG or $(NH_4)_2SO_4$). Interestingly, laccase condensation could be also induced by using liquid-liquid interfaces. At low enzyme concentrations, it is shown that laccase condensation can be driven by BSA-laccase cross-linking and BSA condensation. This has produced catalytically-active aggregates. The reported preliminary results on laccase provide the basis for more systematic experimental and theoretical investigations on this system.