Towards further understanding of protein crystallization : phase diagram, protein interactions, nucleation kinetics and growth kinetics

Liu, Yingxin

2010

Liu, Y. X. (2010). Towards further understanding of protein crystallization : phase diagram, protein interactions, nucleation kinetics and growth kinetics. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/38586>

<https://doi.org/10.32657/10356/38586>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**TOWARDS FURTHER UNDERSTANDING OF PROTEIN
CRYSTALLIZATION: PHASE DIAGRAM, PROTEIN
INTERACTIONS, NUCLEATION KINETICS AND GROWTH
KINETICS**

YINGXIN LIU

SCHOOL OF CHEMICAL AND BIOMEDICAL ENGINEERING

2010

TOWARDS FURTHER UNDERSTANDING OF PROTEIN CRYSTALLIZATION: PHASE
DIAGRAM, PROTEIN INTERACTIONS, NUCLEATION KINETICS AND GROWTH KINETICS

YINGXIN LIU 2010

TOWARDS FURTHER UNDERSTANDING OF PROTEIN CRYSTALLIZATION: PHASE DIAGRAM, PROTEIN INTERACTIONS, NUCLEATION KINETICS AND GROWTH KINETICS

YINGXIN LIU

School of Chemical and Biomedical Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2010

ACKNOWLEDGEMENTS

Upon the completion of this thesis, I would like to express my sincere gratitude to my advisor, Professor Ching Chi Bun, for introducing me to this interesting field of protein crystallization and his guidance and support throughout the years. His helpful advice and endless patience have kept me motivated and persistent during my study in NTU and allowed me to carry out this work to the best of my ability.

Many thanks must go to Dr. Wang Xiujuan for her great help in so many ways. She not only provided scientific discussions and critical comments on my research work but also helped me a lot on operating the instruments involved in this work. As a senior, she also shared her own experience on course work and career development with me.

The helpful advice and stimulating discussions on the nucleation mechanism of protein crystals from Dr. Lu Jie is greatly appreciated as well.

I would also like to gratefully acknowledge Nanyang Technological University for providing me the opportunity to take the PhD program with financial support. The supports and research facilities provided by the Division of Chemical and Biomolecular Engineering, NTU are greatly appreciated. It is also my pleasure to express my gratitude to all the staffs and students in Professor Ching's group for their friendship, helps and encouragement.

The last but not the least, I would like to express my thanks to my parents without whose supports I could not have arrived at this point. They always gave me their love and dedication and allowed me great freedom to pursue my interests.

TABLE OF CONTENTS

Acknowledgement	i
Table of Contents	iii
List of Figures	vii
List of Tables	xvii
Nomenclature	xviii
Summary	xxviii
1 Introduction.....	1
2 Literature Review.....	9
2.1 Structure Determination of Protein Molecules	9
2.2 Selected Examples of Compounds Discovered using a Structure-based Drug Design Approach	12
2.3 Applications of Protein Crystallization.....	18
2.4 Protein Crystallization Methods	18
2.5 Phase Behavior of Protein Molecules	21
2.5.1 Solubility Studies	24
2.5.2 Liquid-liquid Phase Separation.....	26
2.6 Second Virial Coefficient and Protein Interactions	30
2.6.1 Static Light Scattering (SLS)	32
2.6.2 Small Angle X-ray Scattering (SAXS)	35
2.6.3 Small Angle Neutron Scattering (SANS)	36
2.6.4 Self-interaction Chromatography (SIC).....	37
2.6.5 Size Exclusion Chromatography (SEC).....	39
2.6.6 Nanoparticle-based Assay.....	40
2.7 Aggregation and Pre-nucleation Behaviors of Protein Molecules.....	40
2.8 Nucleation Studies	44
2.8.1 Classical Nucleation Theory	44
2.8.2 Nucleation Kinetics Studies.....	46

2.8.3 Heterogeneous Nucleation	48
2.8.4 Effects of Seeding on the Nucleation of Protein Crystals	55
2.9 Protein Crystal Growth	56
3 Effects of pH, Salt Type, Salt Concentration and Temperature on the Interactions and Solubility of Hen Egg White Lysozyme	59
3.1 Introduction.....	59
3.2 Experimental Section	61
3.2.1 Materials	61
3.2.2 Solution Preparation	62
3.2.3 Solubility Measurement	62
3.2.4 Second Virial Coefficient Measurement.....	63
3.2.5 Observation of Lysozyme Crystals by Optical Microscopy	64
3.2.6 Raman Microscopy Measurement	64
3.3 Results and Discussion	65
3.3.1 Solubility.....	65
3.3.2 Second Virial Coefficient	68
3.3.3 Correlation of Solubility and Second Virial Coefficient	73
3.3.4 Crystallization Results	79
3.4 Conclusions.....	82
4 Cloud-point Temperature and Liquid-liquid Phase Separation in Supersaturated Lysozyme and Bovine Serum Albumin Solutions.....	84
4.1 Introduction.....	84
4.2 Experimental Section	87
4.2.1 Materials	87
4.2.2 Solution Preparations	87
4.2.3 Sample Preparation	88
4.2.4 Cloud-point Temperature and Second Virial Coefficient Measurement ...	88
4.3 Results and Discussion	90

4.3.1 Cloud-point Temperatures	90
4.3.2 Second Virial Coefficient	94
4.3.3 Correlation of Cloud-point Temperature and Second Virial Coefficient ..	95
4.3.4 Predictions of the Liquid-liquid Co-existence Curve.	98
4.3.5 Linking Models to Experimental Data.....	101
4.4 Conclusions.....	104
5 The Effects of Solution Parameters on the Nucleation and Growth of Lysozyme Crystals	106
5.1 Introduction.....	106
5.2 Experimental Section	109
5.2.1 Materials	109
5.2.2 Solution Preparation	109
5.2.3 Cloud-point Temperature Measurement	109
5.2.4 Solubility and Second Virial Coefficient Measurement	111
5.2.5 Observation of the Crystallization Processes by Optical Microscopy.....	112
5.2.6 Nucleation Rates Measurement	112
5.2.7 Crystal Growth Rate Measurement.....	114
5.2.8 Raman Microscopy Measurement	115
5.3 Theoretical Models	116
5.3.1 Analysis of Measured Nucleation Rates by Classical Nucleation Theory (CNT).....	116
5.3.2 Analysis of the Local Concentration Profile in the Region in between the Growing Crystals and Protein-rich Droplets	117
5.4 Results and Discussion	120
5.4.1 Cloud-point Temperatures for Lysozyme Solutions.....	120
5.4.2 Protein Crystallization without the Liquid-liquid Phase Separation	120
5.4.3 Liquid-liquid Phase Separation and the Subsequent Crystallization Process	129
5.5 Conclusions.....	137

6 Influence of the Roughness, Topography and Physicochemical Properties of Chemically Modified Surfaces on the Heterogeneous Nucleation of Protein Crystals	139
6.1 Introduction.....	139
6.2 Theory	141
6.3 Experimental Section	146
6.3.1 Materials	146
6.3.2 Preparation of Glass Slide Surfaces.....	146
6.3.3 Contact Angle Measurement.....	148
6.3.4 Atomic Force Microscopy (AFM) Measurement	148
6.3.5 Induction Time Measurement	149
6.4 Results.....	150
6.4.1 Contact Angle	150
6.4.2 Surface Topography.....	153
6.4.3 Induction Time.....	156
6.5 Discussion	158
6.6 Conclusions.....	166
7 Conclusions and Future Directions.....	168
7.1 Conclusions.....	168
7.2 Suggestions for Future Work	172
8 References	174
9 List of Publications	199

LIST OF FIGURES

Figure	Title	Page
Figure 2.1	Three types of vapor diffusion setups, hanging drop, sitting drop and sandwich drop method.	20
Figure 2.2	Schematic temperature-density phase diagrams for hard spheres with Yukawa attraction after the work reported by Muschol and Rosenberger [37]: (a) standard phase diagram with a stable gas-liquid co-existence region below the critical point (CP), (b) phase diagram when the attractive range is smaller than one-seventh of the diameter of the colloidal particle.	22
Figure 2.3	Schematic phase diagram for globular proteins: (a) the solubility increases with increasing temperature, (b) the solubility decreases with increasing temperature.	23
Figure 3.1	Typical Debye plots of lysozyme static light scattering data at the scattering angle of 90° in sodium chloride solutions of 0.1 M sodium acetate buffer at pH 4.5, 25 °C.	64
Figure 3.2	Solubility dependence of hen egg white lysozyme on the concentration of the added sodium chloride in 0.1 M sodium acetate buffer, at pH 4.2, T=10, 15, 20 and 25 °C. The lines are added for guide to an eye.	66
Figure 3.3	Solubility dependence of hen egg white lysozyme on the	67

concentration of the added sodium chloride in 0.1 M sodium acetate buffer at $T=25\text{ }^{\circ}\text{C}$, pH 4.2, 4.5 and 5.2. The lines are added for guide to an eye.

Figure 3.4 Solubility dependence of hen egg white lysozyme on the added 68
sodium chloride or sodium nitrate concentration in 0.1 M sodium acetate buffer at $T=25\text{ }^{\circ}\text{C}$, pH 4.5. The lines are added for guide to an eye.

Figure 3.5 Comparison of the second virial coefficient for lysozyme in 69
sodium chloride solution at pH 4.5, $25\text{ }^{\circ}\text{C}$ determined in this work with the data reported previously in the literature [41, 43, 45, 182].

Figure 3.6 Comparison of second virial coefficient determined from static 70
light scattering for lysozyme in various sodium chloride solutions in 0.1 M sodium acetate buffer at pH 4.2, $T=10, 15, 20$ and $25\text{ }^{\circ}\text{C}$. The lines are added for guide to an eye.

Figure 3.7 Comparison of second virial coefficient determined from static 71
light scattering for lysozyme in various sodium chloride solutions in 0.1 M sodium acetate buffer at $T=25^{\circ}\text{C}$, pH 4.2, 4.5 and 5.2. The lines are added for guide to an eye.

Figure 3.8 Comparison of second virial coefficient determined from static 72
light scattering for lysozyme in various salt solutions in 0.1 M sodium acetate buffer at pH 4.5, $T=25\text{ }^{\circ}\text{C}$. The lines are added for

guide to an eye.

- Figure 3.9 Solubility versus second virial coefficient for lysozyme dissolved 74
in various salt solutions of 0.1 M sodium acetate buffer at pH and
temperature.
- Figure 3.10 Dependence of $\Delta\mu_p$ on the salt concentration in 0.1 M sodium 77
acetate buffer at various pH and temperature. The lines are added
for guide to an eye.
- Figure 3.11 Crystallization results of lysozyme in sodium chloride and sodium 80
nitrate solutions in 0.1 M sodium acetate buffer at pH 4.5, 25 °C.
- Figure 3.12 Optical microscopy images of (a) tetragonal and (b) monoclinic 81
lysozyme crystals.
- Figure 3.13 Raman spectra of tetragonal and monoclinic lysozyme crystals. 82
- Figure 4.1 Illustration of a typical result of cloud-point temperature 90
determination for lysozyme solution containing 128.4 mg/mL
lysozyme, 3% w/v sodium chloride in 0.1 M sodium acetate buffer
at pH 4.2: (a) 45 °C, the solution was clear, (b) 12.2 °C, the
solution became cloudy.
- Figure 4.2 Cloud-point temperature determination for lysozyme as a function 91
of lysozyme concentration at four NaCl concentrations. The lines
are added for guide to an eye.

Figure 4.3 Comparison of the cloud-point temperature data measured in this 92
work with the literature data. (a) Comparison of the dependence of
the cloud-point temperature ($C(\text{lysozyme})=87.4 \text{ mg/mL}$, $\text{pH}=4.2$,
 $[\text{NaOAc}]=0.1 \text{ M}$) on the ionic strength with the data reported by
Grigsby *et al.* [29] ($C(\text{lysozyme})=87 \text{ mg/mL}$, $\text{pH}=4.0$, 20 mM Tris
Buffer). (b) Comparison of the dependence of the cloud-point
temperature ($C(\text{NaCl})=4\% \text{ w/v}$, $\text{pH}=4.2$, $[\text{NaOAc}]=0.1 \text{ M}$) on
lysozyme concentration with the work reported by Petsev *et al.*
[190] ($C(\text{NaCl})=4\% \text{ w/v}$, $\text{pH}=4.5$, $[\text{NaOAc}]=0.05 \text{ M}$). (c)
Comparison of the dependence of cloud-point temperature
($C(\text{NaCl})=3\% \text{ w/v}$ and $C(\text{NaCl})=5\% \text{ w/v}$, $\text{pH}=4.2$, $[\text{NaOAc}]=0.1$
 M) on lysozyme concentration with the work reported by Muschol
et al. [37] ($C(\text{NaCl})=3\% \text{ w/v}$ and $C(\text{NaCl})=5\% \text{ w/v}$, $\text{pH}=4.5$,
 $[\text{NaOAc}]=0.1 \text{ M}$).

Figure 4.4 Cloud-point temperature determination for BSA as a function of 94
BSA concentration at three PEG 1500 concentrations. The lines are
added for guide to an eye.

Figure 4.5 Comparison of second virial coefficients of BSA determined from 95
static light scattering in PEG 1500 solutions of 0.1 M sodium
acetate buffer at $\text{pH } 5.0$, $T=15, 20$ and $25 \text{ }^\circ\text{C}$.

Figure 4.6 Dimensionless cloud-point temperature T_c versus dimensionless 96
second virial coefficient B'_{22} of lysozyme in different NaCl

solutions. The lysozyme concentration was 87.4 mg/mL in all cases. The buffer was 0.1 M NaAc buffer at pH 4.2.

- Figure 4.7 Dimensionless cloud-point temperature T_r versus dimensionless 97
second virial coefficient B'_{22} of bovine serum albumin in different
PEG 1500 solutions. The BSA concentration was 117.8 mg/ml in
all cases. The buffer was 0.1 M NaAc buffer at pH 5.0.
- Figure 4.8 Comparison of the liquid-liquid co-existence curves calculated 100
using different methods with the cloud-point temperature data
reported in literature.
- Figure 4.9 Dependence of ε/k for lysozyme calculated from the measured 103
second virial coefficients on temperature at different sodium
chloride solutions. The buffer is 0.1 M NaAc buffer at pH 4.2.
- Figure 4.10 Dependence of ε/k for bovine serum albumin calculated from the 104
measured second virial coefficients on temperature at different
PEG 1500 concentrations. The buffer is 0.1 M NaAc buffer at pH
5.0.
- Figure 5.1 Illustration of a typical result of cloud-point temperature 111
determination for lysozyme solution (lysozyme, 128.4 mg/mL,
NaCl 3% w/v) in 0.1 M sodium acetate buffer, pH 4.2: (a) 45 °C,
the solution was uniform, (b) 12.2 °C, the protein-rich droplets
separated from the solution.

- Figure 5.2 A typical example of the measurement of nucleation rate. The 113
number of crystals in one droplet of 0.5 μL was plotted against the
time from initial measurement. The slope of the straight line was
calculated to be the nucleation rate J .
- Figure 5.3 Illustration for the measurement of the growth rates for (110) and 114
(101) faces of tetragonal lysozyme crystals.
- Figure 5.4 Schematic illustration of a crystal nucleating and growing from a 119
liquid-liquid phase separation solution.
- Figure 5.5 Local concentration profile calculated at different initial overall 119
concentrations assuming the initial moment of the system to be in a
quasi-steady state.
- Figure 5.6 Cloud-point temperature as a function of lysozyme concentration 121
at four sodium chloride solutions in 0.1 M sodium acetate buffer,
at pH 4.2, 10 $^{\circ}\text{C}$.
- Figure 5.7 Images of lysozyme crystals growing from solution consisting of 121
23 mg/mL lysozyme, 3% w/v NaCl, in 0.1 M sodium acetate
buffer at pH 4.2, 10 $^{\circ}\text{C}$. The pictures were taken at (a) 15 min, (b)
25 min, (c) 35 min and (d) 45 min after initial observation.
- Figure 5.8 Dependence of nucleation rate J on the lysozyme concentration at 123
four sodium chloride concentrations in 0.1 M sodium acetate
buffer at pH 4.2, 10 $^{\circ}\text{C}$. Dash lines are fitted with the classical

nucleation theory.

- Figure 5.9 Dependence of $\gamma^3 / (k^3 T^3 (\ln S)^2)$ on the sodium chloride 125
concentration at fixed lysozyme concentrations.
- Figure 5.10 Crystallization window of lysozyme crystals obtained in various 127
crystallization conditions in 0.1 M sodium acetate buffer at pH 4.2,
10 °C.
- Figure 5.11 Growth rates of tetragonal lysozyme crystals in 0.1 M sodium 127
acetate buffer, at pH 4.2, 10 °C in various crystallization
conditions. The lines are drawn for guide to an eye.
- Figure 5.12 Images of lysozyme crystals growing from a solution consisting of 128
55.7 mg/mL lysozyme, 4% w/v NaCl in 0.1 M sodium acetate
buffer at pH 4.2, 10 °C. These pictures were taken at (c) 10 min,
(d) 20 min, (e) 30 min and (f) 40 min after initial observation. (a)
The solution was uniform right after mixing at 45 °C. (b) Protein-
rich droplets separated from the solution at 10 °C.
- Figure 5.13 FT-Raman spectra of (a) 50.6 mg/mL, (b) 88.4 mg/mL, (c) 126.3 132
mg/mL and (d) 177.1 mg/mL lysozyme solutions.
- Figure 5.14 Schematic illustration of the liquid-liquid co-existence curve for 134
lysozyme-NaCl solution.
- Figure 6.1 Schematic plots for homogeneous and heterogeneous nucleation 143

rates with different coefficients (ϕ').

- Figure 6.2 Geometry of a sphere cap shaped nucleating solution on the rough 143
surface.
- Figure 6.3 Contact angles for water and crystallization solution (lysozyme, 20 151
mg/mL; NaCl, 2% w/v) on various surfaces with 1 min's
stabilizing time.
- Figure 6.4 The time dependence of the contact angles for water and 151
crystallization solution (lysozyme, 20 mg/mL; NaCl, 2% w/v) on
the surfaces of P4VP and PMMA.
- Figure 6.5 Illustration of the time dependence of the contact angles for 152
crystallization solution (lysozyme, 20 mg/mL; NaCl, 2% w/v) on
the surface of P4VP coated glass slide: (a) 52.65° (0 min), (b)
 32.51° (1 min), (c) 17.22° (12 min).
- Figure 6.6 Three-dimensional AFM images of the investigated surfaces of (a) 154
bare glass slide, (b) poly(2-hydroxyethyl methacrylate) coated
glass slide, (c) poly-L-glutamic acid coated glass slide, (d)
poly(methyl methacrylate) coated glass slide, (e) ATPES-silanized
glass slide, and (f) poly(4-vinyl pyridine) coated glass slide. The
scanner size is $1 \times 1 \mu\text{m}$.
- Figure 6.7 Three-dimensional AFM images of the investigated surfaces of (a) 155
bare glass slide, (b) poly(2-hydroxyethyl methacrylate) coated

glass slide, (c) poly-L-glutamic acid coated glass slide, (d) poly(methyl methacrylate) coated glass slide, (e) ATPES-silanized glass slide, and (f) poly(4-vinyl pyridine) coated glass slide. The scanner size is $5 \times 5 \mu\text{m}$.

Figure 6.8 The induction times of protein crystallization (20 mg/mL lysozyme; 2% w/v NaCl; 4 °C) on various surfaces. Error bars denote standard deviation ($n_s=5$). 157

Figure 6.9 Lysozyme crystals grown on the surfaces of (a) bare glass slide and (b) poly(4-vinyl pyridine) coated glass slide. The crystallization conditions of the two images were the same, which were conducted at 4 °C and with the concentration of lysozyme 20 mg/mL and the concentration of NaCl 2% w/v. 158

Figure 6.10 $\Delta G_{het}^* / \Delta G_{hom}^*$ ratio as a function of the contact angle on different roughness. 160

Figure 6.11 Schematic illustration of the processes of protein aggregation and crystallization on ideally flat surface. (a) Lysozyme molecules are dispersed in solution, (b) The adsorption of lysozyme molecules: the adsorption of lysozyme molecules may be reversible; lysozyme molecules may arrive at the surface randomly and migrate over the surface to form the absorption layer, (c) Lysozyme molecules form multi-layers layer by layer, (d) Some of the molecules form small aggregates, (e) The formation of quasi-critical clusters. (f) The 163

formation of crystals from the quasi-critical clusters.

Figure 6.12 Figure 6.12. Schematic illustration of protein crystallization and 164
the formation of large cluster on rough surface. (a) Lysozyme
molecules are dispersed in solution. (b) The adsorption of
lysozyme molecules from the solution to the surface; the presence
of hillocks on the surface may block the lateral migration of the
molecules. (c) The blocked molecules are packed into compact
aggregates. (d) The molecules that are packed with bond angles
corresponding to protein crystals form quasi-critical clusters; the
molecules that are randomly packed form fractal clusters. (e)
Quasi-critical clusters nucleate and grow into crystals, while the
fractal clusters grow into larger clusters.

Figure 6.13 Figure 6.13. Schematic illustration of the swelling of P4VP in 166
contact with the crystallization solution (lysozyme, 20 mg/mL;
NaCl, 2% w/v).

LIST OF TABLES

Table	Title	Page
Table 2.1	PDB statistics summary of released entries as of May 2009 [92].	11
Table 2.2	Selected examples of the compounds discovered using structure-based drug design approach.	15
Table 2.3	Heterogeneous nucleation surfaces studies summarized in this chapter.	49
Table 3.1	Second virial coefficient (B_{22}), solubility (c_e) and $\Delta\mu_p$ for lysozyme under various conditions.	77
Table 5.1	Second virial coefficient (B_{22}), solubility (c_e) and protein interaction ($\frac{\varepsilon}{kT_{283.15}}$) in various sodium chloride solutions at pH 4.2, T=10°C	125
Table 5.2	Tyrosine doublet ratio for lysozyme samples.	133
Table 6.1	Root mean square (RMS) and the average roughness (R_a) of chemically modified glass slides, as derived from AFM analysis.	156

NOMENCLATURE

B_{22}	Second virial coefficient
Π	Osmotic pressure
c_p	Protein concentration
R	Gas constant
T	Temperature
M_w	Molecular weight
N_A	Avogadro's number
$W(r)$	Sum of pairwise potential of protein molecules
r	Center-to-center distance between two protein molecules
$W_{hs}(r)$	Hard sphere potential
$W_{elec}(r)$	Electric double-layer-repulsion potential
$W_{disp}(r)$	Dispersion potential
$W_{spec}(r)$	Specific potential
σ	Hard-sphere diameter of the protein molecules
z	Valence of the protein molecule in the solution condition

e	Elementary charge
ϵ_o	Dielectric permittivity of free space
ϵ_r	Relative dielectric permittivity of water
κ	Inverse Debye length
I	Ionic strength of the solution
k	Boltzmann constant
H	Effective Hamaker constant for protein interactions
ϵ_{spec}	Depth of the square well
δ	Width of the square well
K	System-specific constant
n_0	Solvent refractive index
λ	Wavelength
n_p	Refractive index of protein solution
R_θ	Excess Rayleigh scattering ratio
B_3	Third virial coefficient
c	Solution concentration

s	Scattering vector
θ	Scattering angle
$I(c, s)$	Intensity of scattering light
$S(c, s)$	Strcuture factor
ρ	Molecular protein concentration
q	Scattering vector
$P(q)$	Form factor
$S(q)$	Structure factor
B'_{22}	Dimensionless second virial coefficient
k'	Retention factor
V_r	Retention volume of a solute to be eluted from the column
V_o	Retention volume of the solute that is of free interaction with the packing particles of the column to be eluted
B_{22}^{HS}	Hard sphere second virial coefficient
ϕ	Phase ratio
A_s	Total available surface area

ρ_s	Number of immobilized molecules per unit area
K_D	Local solute distribution coefficient
K_0	Partition coefficient of solute molecules between chromatographic phases in the limit of infinite dilution
$\langle C_i \rangle$	Average concentration of the mobile phase in the pulse
t_r	Solute retention time
t_o	Retention time of the completely excluded molecules
t_T	Retention time of the completely included molecules
V_T	Retention volume of the completely included molecules
$g(\tau)$	Intensity autocorrelation function
t	Time
τ	Delayed time
$I(t)$	Dynamic light scattering intensity at time t
$I(t + \tau)$	Dynamic light scattering intensity of the signal delayed by a time τ
Γ	Delay constant
A	Background

B	Instrument constant
D	Translational diffusion coefficient
R_H	Hydrodynamic radius
k_B	Boltzmann's constant
η	Solvent viscosity
ΔG	Total free energy barrier
ΔG_s	Free energy change of the transformation per unit area
a	Radius of the spherical nucleus
ΔG_v	Free energy change of the transformation per unit volume
γ	Interfacial tension
a_c	Radius of the critical nucleus
S	Supersaturation ratio
ΔG_{crit}	Critical free energy barrier
v	Volume of the molecule
c_e	Solubility
J	Nucleation rate

h	Planck's constant
ΔG_a	Energy barrier of the diffusion from the bulk solution to the cluster
N_1	Number of molecules
$C(NaCl)$	Sodium chloride concentration
$C(salt)$	Salt concentration
ρ'	Number density of the protein molecule
m	The number of water molecules that can be placed in the volume of one protein molecule
A'	Fitting parameter
z'	Fitting parameter
λ'	Range of protein interaction
p	Anisotropy
$\mu_p(xtal)$	The chemical potential of protein in crystalline phase
$\mu_p(soln)$	The chemical potential of protein in solution
$\mu_p^o(soln)$	The standard chemical potential of protein in solution
y_p	The activity coefficient

$\Delta\mu_p$	$\mu_p^o(so\ln) - \mu_p(xtal)$
T_{cloud}	Cloud-point temperature
T_r	Dimensinless cloud temperature
B_{22}'	Dimensionless second virial coefficient
$c(r)$	Direct correlation function
$c_{ref}(r)$	Direct correlation function for the reference system
C	Fourier transform of $c(r)$
C_{ref}	Fourier transform of $c_{ref}(r)$
$u_1(r)$	Perturbation part of the total energy
$u(r)$	Total energy
$U_1(r)$	Fourier transform of $u_1(r)$
P	System pressure
P_{ref}	Pressure of the reference system
μ	Chemical potential of the system
μ'	Chemical potential of the reference system
η'	Protein volume fraction of the liquid phase

ε	Protein interaction energy
η'_c	Critical protein volume fraction of the liquid phase
α	Phase α
β	Phase β
κ'	Fitting parameter
T_c	Critical Temperature
N	Number of crystals
M	Longest width of (110) face of the tetragonal lysozyme crystal
L_A	The distance between two (110) faces of the the tetragonal crystal
G_{110}	Growth rate of (110) face of tetragonal lysozyme crystal
G_{101}	Growth rate of (101) face of tetragonal lysozyme crystal
b	Constant
C'	Protein concentration in the 'buffer zone'
r'	Diffusion radius
C_n	Protein concentration in the nucleus

C_{tr}	Crossover concentration
C_L	Protein concentration of the protein-lean phase
r_{LL}	Liquid-liquid co-existence boundary
A^*	Collision factor
t_{ind}	Induction time
V	Volume of crystallization solution
ϕ'	Ratio of the Gibbs free energy of heterogeneous nucleation to homogeneous nucleation
ΔG_{hom}	Free energy barrier for homogeneous nucleation
ΔG_{het}	Free energy barrier for heterogeneous nucleation
ΔG_{hom}^*	Critical free energy barrier for homogeneous nucleation
ΔG_{het}^*	Critical free energy barrier for heterogeneous nucleation
B'	Constant
B'_{het}	Constant B' for heterogeneous nucleation
B'_{hom}	Constant B' for homogeneous nucleation
θ'	Contact angle between nucleus and surface

R'	Main radius of the spherical cap
r''	Radius of the cone on the surface
h'	Height of the cone on the surface
n_s	Number of experiments
n'	Number of the cones on the surface
A_{SL}	Solid-liquid interfacial area
A_L	Surface area of the liquid spherical cap
Ω	Molar volume
α'	Ratio of r'' to R'
β'	Ratio of h' to R'
γ_L	Nucleus-liquid interfacial energies
γ_i	Nucleus-substrate interfacial energies
γ_s	Liquid-substrate interfacial energies
θ_Y	Young contact angle
Γ_w	Roughness ratio

SUMMARY

Protein crystallization has been well recognized to be of great importance in several applications. Crystallization of protein crystals is the necessary first step in these applications. Given the lack of fundamental understanding of protein crystallization mechanism, protein crystals are usually obtained by the trial-and-error method. Screening for optimal crystallization conditions always requires a large amount of time and protein. Thus, knowledge of phase behavior, crystal nucleation and growth kinetics and the interactions that govern them is crucial for the understanding of protein crystallization mechanism and the design of optimal crystallization conditions.

In this dissertation, a full-picture investigation of protein crystallization has been presented. The phase diagram was obtained by measuring solubility and cloud-point temperature of lysozyme in solutions of various conditions. Protein interactions were characterized by measuring the second virial coefficient. It was found that there was a systematic variation between the solubility and the second virial coefficient; the solubility decreased with the increasing of protein attractive interactions, and the liquid-liquid phase separation was also driven by the net attractive interactions between protein molecules. The experimental data were also compared with the theoretical predictions. A good agreement was exhibited between the experimental results and theoretical predictions. In addition, a quantitative description of the liquid-liquid co-existence curve was achieved by modifying random phase approximation.

Nucleation kinetics of tetragonal lysozyme crystals was measured at various solution conditions. The nucleation rates were found to increase with the increase of the protein

and sodium chloride concentration when crystallization occurred in the solid-liquid co-existence region, and the broken dependence of nucleation rate on protein concentration around the liquid-liquid co-existence boundary was identified. The suppression of nucleation in the protein-rich phase was investigated in terms of protein interactions as determined by Raman spectroscopy and the self-assembly of protein molecules. The concentration distribution in the region between the growing nucleus and the protein-rich phase was derived to explain the increase of nucleation rate with the increase of the overall protein concentration in the liquid-liquid co-existence region. The different shapes and morphologies of lysozyme crystals were attributed to different growth rates of the (110) and (101) face of tetragonal lysozyme crystals and the protein interactions.

In addition, the influences of roughness, topography and physicochemical properties of chemically modified surfaces on the heterogeneous nucleation of protein crystals were investigated. An optical microscope with a heating/cooling stage was employed to measure the induction time of heterogeneous nucleation. The surface topography and roughness were characterized by the atomic force microscopy. Contact angles for crystallization solution on the investigated surfaces were measured by a contact angle meter. A theoretical model was proposed to relate the free energy barrier for nucleation to surface roughness and contact angle between the nucleating solution and the surface investigated. From the theoretical analysis, the energetic barrier to heterogeneous nucleation was found to increase at higher contact angles, and to decrease at higher roughness. The theoretical analysis was quite consistent with the experimental results.

1 Introduction

Protein crystallization has been well recognized to be of great importance in several applications, such as the determination of the three-dimensional structure of protein molecules, drug delivery, and the development of novel catalytic materials [1]. Detailed knowledge of the three-dimensional structures of protein molecules and protein-ligand complexes is essential to the understanding of genome structure-function relationship and the structure-based drug design. According to the statistics of protein data bank (PDB), X-ray crystallography is currently predominant in structure determination of protein molecules. However, a crystal of suitable size and diffraction quality must first be obtained in order to determine the structures of protein molecules by X-ray crystallography. On the other hand, the purification of some proteins can be achieved by protein crystallization. Cross-linked protein crystals that are of greatly enhanced stability and performance can be applied to facilitate sustained drug delivery and extend the shelf-life of protein pharmaceuticals [2-4]. The improved stability of the enzyme crystals also allows the commercialization of the enzymes as chemical and environmental catalysts [5]. In addition, protein crystals can be applied as chiral chromatographic media due to their micro-porous structures. The investigations of protein crystallization mechanisms and the applications of protein crystals usually involve the research work from physics, chemistry, biochemistry, molecular biology, and chemical and material engineering. Thus, protein crystallization is an important interdisciplinary area.

Due to the lack of fundamental understanding of the mechanism of protein crystallization and the complexity of protein molecules, it is difficult to design protein crystallization conditions rationally. Usually, protein crystals are obtained by trial-and-

error method. The screening for optimal protein crystallization conditions is achieved by systematically varying solution variables, such as precipitant type and concentration, protein concentration, pH, and temperature. Thus, a large amount of experiments are required. It is usually costly and time-consuming. The preparation of protein crystals that are of suitable quality is the main bottleneck to the applications of protein crystals in different disciplines. To overcome this bottleneck is the major driving force for protein crystallization studies.

Previous studies on protein crystallization have involved nucleation, crystal growth, and thermodynamic phase behavior of protein over a wide range of solution parameters. In order to grow high quality protein crystals, it is necessary to know protein solution phase behaviors and the interactions that govern them. Currently, experimental measurement is the main source of the determination of protein solubility. A lot of measurements have been carried out on the solubility of protein molecules. For example, the solubility data of hen egg white lysozyme have been measured as a function of salt type, salt concentration, pH, and temperature [6-12]. Mikol and Giegé investigated the effects of the concentration of ammonium sulfate, solution pH, and temperature on the solubility of concanavalin A systematically [13]. Additionally, the solubility of ovalbumin [14], concanavalin A [15], equine serum albumin [16], and glucose isomerase [17] were reported.

The effects of liquid-liquid phase separation on protein crystallization have been investigated in the past two decades [18-28]. Different mechanisms have been proposed to explain the influences of the liquid-liquid phase separation on protein crystallization. For example, the presence of a metastable liquid-liquid critical point was proposed to change

the nucleation pathway dramatically and enhance the nucleation by density fluctuation [25, 26]. Therefore, the study on the location of the liquid-liquid co-existence boundary and the strategies to shift it provides a systematic approach to control protein crystallization. The location of liquid-liquid co-existence curve can be determined by measuring the cloud-point temperatures using the dynamic light scattering (DLS) [29], thermooptical analysis technique [30], micro-differential scanning calorimeter [31], and small angle X-ray scattering (SAXS) [32-34]. By employing these techniques, the effects of pH, electrolyte identity, and electrolyte concentrations on the cloud-point temperature of lysozyme in electrolyte solutions were investigated systematically [29-31, 35-37]. More recently, the cloud-point temperatures of bovine serum albumin (BSA) [38], urate oxidase [32] and bovine pancreatic trypsin inhibitor (BPTI) [33, 34] in polyethylene glycol solutions were determined as well. Additionally, models based on van der Waals equation of state, statistical-mechanical perturbation theory, Gibbs free energy model, and Monte Carlo simulations were applied to predict the liquid-liquid co-existence curve.

Phase behaviors of protein solution are known to be driven by protein interactions. Understanding the correlation between protein interactions and protein phase behavior is essential to control the nucleation and growth of protein crystals. As a thermodynamic parameter that described protein interactions, second virial coefficient (B_{22}) was first proposed by George and Wilson [39] as a predictor for protein crystallization. They measured the second virial coefficient values for a variety of proteins in crystallization conditions. It was found that the values of B_{22} for solutions in optimal crystallization conditions were within a narrow range of -1×10^{-4} to -8×10^{-4} mol mL g⁻². This observation

suggested that B_{22} might be an important parameter to predict protein crystallization conditions. Since then a lot of efforts have been made to investigate the influences of solution parameters on protein interactions and correlate protein interactions with protein phase behaviors. The effects of pH, electrolyte concentration, and poly (ethylene glycol) (PEG) on the B_{22} of lysozyme and chymotrypsinogen [40, 41] and the effects of glycerol on the B_{22} of bovine pancreatic trypsin inhibitor (BPTI) [42] were investigated by SLS. Bonneté *et al.* [43, 44] have applied SAXS to measure the B_{22} values of urate oxidase as a function of solution pH, temperature, and the addition of PEG. Besides the traditional methods SLS and SAXS, self-interaction chromatography [45, 46], size exclusive chromatography [47] and a nanoparticle-based assay [48] were also developed and applied to measure the second virial coefficient. The measured second virial coefficient was also found to be closely related to solubility and cloud-point temperature [27, 49, 50].

Protein crystallization process is considered to be consisting of two processes; nucleation and crystal growth. In this respect, an understanding of the influences of solution variables on protein nucleation and growth kinetics is important in the control of the size and quality of the crystals obtained. Some previous studies have focused on the effects of solution variables, such as solution temperature, pH, *etc.*, on the nucleation rates of protein crystals [19, 20, 51-54]. Considerable researches have also been carried out to study the effects of the liquid-liquid phase separation on the nucleation of protein crystals [18-26, 28, 35, 36, 52, 55]. Since heterogeneous nucleation can decrease the free energy barrier for nucleation and grow protein crystals at low supersaturation, various materials [56-72] have been tested as heterogeneous nucleation surfaces for protein crystallization.

As another way to induce heterogeneous nucleation, various seeding materials were applied to promote protein crystallization [73-76]. Some groups also investigated the effects of electric and magnetic fields on the nucleation of protein crystals [77, 78]. A systematic investigation on protein crystal growth not only provides guidelines for growing better crystals but also provides insights into the mechanisms dominating the perfection of the quality of protein crystals. The investigation on the growth mechanism was usually carried out by measuring the growth rate of a protein crystal using optical microscopy. The influences of temperature, pH and salt concentrations on the growth rates of tetragonal lysozyme crystals were investigated and correlated with the quality of protein crystal [79-81]. In addition, electron microscopy [82], atomic force microscopy (AFM) [83] and advanced optical microscopy [84] were applied to study the defect generation mechanism of protein crystals.

As mentioned above, different aspects of protein crystallization have been investigated, but most of these studies just focused on one or two sides of the whole crystallization process. When a full picture of protein crystallization is sought, the experimental conditions and methods were normally inconsistent with each other. Furthermore, the effects of many parameters are not well understood and highly nonlinear in their effects on other parameters, hence there is still no general theory to describe protein crystallization. A further understanding such as the mechanism of protein interaction on protein nucleation and crystal growth is necessary for tuning protein crystallization in a controlled manner.

The main objective of this work is to present a full-picture investigation of protein crystallization, namely a detailed study on the phase behavior and crystallization mechanism of protein molecules including nucleation kinetics, crystal growth kinetics, crystal morphologies, solubility, second virial coefficient, and cloud-point temperatures using consistent experimental conditions and methodologies. In addition, since heterogeneous nucleation can decrease the free energy barrier for nucleation and grow protein crystals at low supersaturation, some theoretical and experimental efforts on the influence of the roughness, topography and physicochemical properties of chemically modified surfaces on the heterogeneous nucleation of protein crystals were investigated as well.

In this work, lysozyme from hen egg white was selected as a model protein. Different from some proteins that are sensitive to the practical conditions in which they are handled, *e.g.* pH and temperature, lysozyme is relatively more stable. The relatively cheap price and the availability of the thermophysical properties are important factors for us to choose lysozyme as a model protein. On the basis of previous research, more findings and deeper understanding on the mechanism of protein crystal nucleation, growth and morphology were revealed using more reliable and efficient experimental methodologies. Therefore, lysozyme is a useful model protein for the full-picture investigation of the protein crystallization mechanism.

The work presented in this thesis was organized in the sequence as follows.

In Chapter 2, the background and existent knowledge of protein crystallization are summarized. This should elucidate the context and importance of the present work.

In Chapter 3, the effects of pH, temperature, salt type and salt concentration on the solubility and strength of protein interactions were investigated systematically. The solubility was measured by crystallization of supersaturated solutions to reach the equilibrium. The strength of protein interactions was characterized through measuring second virial coefficient by SLS. In parallel, crystallization trials in corresponding conditions were performed. This chapter also included the search of the correlation between protein interactions and protein solubility.

In Chapter 4, the liquid-liquid phase separation of lysozyme and bovine serum albumin was characterized by measuring the liquid-liquid co-existence curves for lysozyme solutions and bovine serum albumin solutions under different practical conditions. In order to correlate the cloud-point temperature with second virial coefficient, the second virial coefficients for lysozyme and bovine serum albumin in various precipitant concentrations and temperatures were measured. The quantitative ability of the theoretical models to predict the experimentally measured liquid-liquid co-existence curves for lysozyme and γ II-crystallin in various solution conditions and the dependence of the energy of protein interaction on temperature were also tested.

Chapter 5 presents a detailed study on the phase behavior and mechanism of lysozyme crystallization. The nucleation and crystal growth rates, crystal morphologies, solubility, second virial coefficient and cloud-point temperatures under different solution conditions

were experimentally measured and theoretically analyzed. The observations and analysis in this chapter were expected to provide a further understanding and guideline for protein crystallization.

In Chapter 6, the influence of the factors, such as surface roughness and topography, on the heterogeneous nucleation of hen egg white lysozyme was examined by carrying out heterogeneous nucleation on a series of chemically modified surfaces. The induction time of heterogeneous nucleation, the surface topography, the surface roughness and the contact angle between the crystallization solution and the foreign surfaces were measured and analyzed. A theoretical model was also proposed to correlate the free energy barrier for heterogeneous nucleation with surface roughness and the contact angle between nucleating solution and the surface. This part of work was expected to provide a further understanding of the heterogeneous nucleation mechanism of protein crystals and an inspiration in the design of the novel heterogeneous nucleant.

In Chapter 7, the conclusions were drawn and some suggestions were provided for the future work.

2 Literature Review

The purpose of this chapter is to provide a detailed introduction on the background of protein crystallization. It includes the applications of protein crystallization, crystallization techniques and the relevant research work in the published literature regarding the fundamental understanding of the mechanism of protein crystallization.

2.1 Structure Determination of Protein Molecules

It is known that proteins play a key role in almost every biological process. Their specific biological functions are determined by the three-dimensional arrangement of the amino acids at the active sites. High-resolution three-dimensional structures of protein molecules are pivotal for the understanding of the biological functions. By May 2009, there have been a total of 57,424 entries available in Protein Data Bank (PDB). As summarized in Table 2.1, X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and electron microscopy (EM) are currently three predominant experimental methods in the determination of the three-dimensional structures of protein molecules. These three methods have their own advantages and limitations and can complement each other.

Around 15% of the structures that are available in PDB were determined using NMR spectroscopy. It serves as a complementary way to X-ray crystallography. As reported in PDB, the structures of some proteins that do not provide crystals of X-ray diffraction quality were determined using liquid-state NMR spectroscopy [85]. Liquid-state NMR spectroscopy allows the determination of the atomic resolution of protein molecules in solution. It comprises the steps including protein sample preparation, data collection,

assignment of chemical shifts of the nuclear spins, assignment of NMR-derived experimental information, and structure calculations. It requires a high solubility in water to obtain an interpretable signal, which may be a severe obstacle for liquid-state NMR structure determination. Although the introduction of transverse relaxation-optimized spectroscopy (TROSY), residual dipolar coupling (RDC), and labeling strategies has allowed liquid-state NMR to determine the three-dimensional structure of proteins with a molecular weight up to ~100 kDa [86], there is still a size limit for proteins to be studied.

In contrast to NMR spectroscopy, the X-ray crystallography does not limit to the size and complexity of the protein molecules. The X-ray crystallography is currently predominant in the structure determination of protein molecules. According to the statistics of the PDB entries as in Table 2.1, around 90% of the structures were determined using X-ray crystallography. The structure is determined by irradiating a diffraction quality crystal and recording the spatial distribution and intensities of the diffracted X-ray beams as long as the diffraction quality crystal is available. It comprises of several steps including protein purification and crystallization, post-crystallization treatment, data collection and phasing [87]. The recent advances in every stage of protein X-ray crystallography, especially the applications of genetic engineering, robotics, synchrotron data collection and high powered computing have accelerated protein crystallography for structural genomics [88]. However, the key to the success of the X-ray crystallography is to obtain the single crystals that are of diffraction quality. Due to the lack of the fundamental understanding of the mechanisms of protein crystallization, the growth of the diffraction quality crystals are the most difficult and time-consuming step in the protein structure determination.

In recent years, emerging methods in cryo-electron microscopy allow the structure determinations of the molecules and molecular assemblies that are too large or too heterogeneous to be investigated using traditional X-ray crystallography [89, 90]. Another recent advance for the structure determination of protein molecules is the solid-state magic-angle-spinning NMR spectroscopy. Although protein structure determined by the solid-state NMR has not been reported, the solid-state NMR spectroscopy provides a possibility to determine the structures of the insoluble proteins at an atomic level [91]. The advances in both instrumentation and software provide a promising possibility to improve the low throughput of the structure determination of protein structures.

Table 2.1. PDB statistics summary of released entries as of May 2009 [92].

Experimental method	X-ray	NMR	EM	Other	Total
Proteins	45,962	6,834	155	110	53,061
Nucleic acid	1,141	850	16	4	2,011
Protein/nucleic acid complexes	2,112	144	59	4	2,319
Other	17	7	0	9	33
Total	49,232	1,835	230	127	57,424

2.2 Selected Examples of Compounds Discovered using a Structure-based Drug Design Approach

A traditional effort for drug discovery usually consists of cycles of synthesis and testing of candidate compounds on cultured cells or animals in a trial-and-error way. Although the introduction of high-throughput screening assay has enabled researchers to develop and screen a large amount of candidate compounds, the traditional drug design approach has some inherent limitations. Different from the traditional drug design, the structure-based drug design begins with the knowledge of the three-dimensional structure information and the specific chemical responses of the target molecules. Using structural information to guide small-molecule design and generate new chemicals is now a mainstream in the drug discovery process [93]. In the past two decades, the structure-based drug design has been applied in the discovery of the drugs for the treatments of various diseases, such as human immunodeficiency virus (HIV) infection, cancer, Alzheimer's disease, *etc.* The increase of the impact of the structure-based drug design in drug discovery is illustrated in Table 2.2 by summarizing the selected examples of the compounds that were designed based on the three-dimensional structure information of the target molecules since 1996.

In the past few years, the inhibition of the HIV protease has been considered as one therapeutic approach for AIDS. The discovery of the novel inhibitors of HIV protease is considered as a classic example of the applications of the structure-based drug design. Based on the three-dimensional structure of the inhibitor/HIV-1 protease complex as revealed by the X-ray crystallography, 5,6-dihydro-4-hydroxy-2-pyrones were discovered as potential agents for the treatment of HIV infection [94]. Guided by an interactive structure-based drug design process, one of *p*-cyanophenyl sulfonamide derivatives, as

illustrated in Table 2.2, was designed and found to be a promising inhibitor of HIV protease with excellent enzyme inhibitory activity and potent antiviral activity [95]. In addition, tipranavir disodium (PNU-140690E) was developed as a potent, orally bioavailable, nonpeptidic HIV protease inhibitor with the help of the crystallographic data on the inhibitor-protease complex [96]. In a recent report, a compound with broad-spectrum activity against several key mutants of HIV protease was developed by analyzing the binding mode of the clinical candidate and existing protease inhibitor to HIV protease [93, 97].

Cancer has long been recognized as a disease of aberrant cellular proliferation [98]. The targets of the therapy of cancer are the mutated, deleted or over-expressed proliferation regulation proteins in cancer cell lines. The cyclin-dependent kinases (CDKs) are important in controlling entry into and transition through each phase of the cell cycle. Therefore, the discovery for the inhibitors to CDKs which control the cell cycle provides a potential approach for the therapy of cancer. As potential cancer therapy pharmaceutical ingredients, imidazo[1,2-a]pyridines [98], furanopyrimidine and pyrrolopyrimidine inhibitors were developed as inhibitors [99] to CDKs, using the knowledge of their binding modes to the target proteins, the emerging structure-activity trends and protein-ligand affinity as obtained through the X-ray crystallography of the inhibitor/target complex. The X-ray crystallographic structures of proteasome and inhibitor/proteasome complex were solved and used to assist in understanding interaction between the non-covalent proteasome inhibitors and the proteasome. A new class of potent cellular active inhibitors was also discovered to act non-covalently to the proteasome and show high specificity for the chymotrypsin-like activity of the enzyme and high antiproliferative

activity [100]. The discovery of the fact that farnesylation is required for ras transforming activity has led to an intensive research for farnesyltransferase (FTase) inhibitors (FTIs) as anticancer agents. Li *et al.* [101] have reported the design, synthesis and testing of a series of novel inhibitors of FTase using the three-dimensional structure of tipifarnib. As proteins that are involved in the regulation of mitosis, aurora kinases have been found to be attractive targets for the design of the anticancer drugs. In a recent report, Coumar and co-workers [102] developed a new compound as the inhibitor for aurora kinase by modification of the existing aurora kinase A inhibitor. The new compound has shown ~450-fold improved aurora kinase A inhibition potency compared to the existing compound.

Alzheimer's disease (AD) is a neurodegenerative disease of the brain that is characterized by the progressive formation of the insoluble amyloid plaques and fibrillary tangles. Inhibition of β -amyloid precursor protein (APP) proteolysis, which is caused by β -secretase (BACE-1), has been considered as a central concept in the therapy of AD [103]. On the basis of the co-crystal structure of inhibitor in the BACE-1, Stachel *et al.* [103] successfully identified a β -secretase inhibitor that was of low molecular weight and displayed ~500 fold selectivity of β -secretase versus cathepsin D and renin. Novel P1-P3 linked macrocyclic BACE-1 inhibitors were designed and synthesized based on the crystallographic data on the complex of open-chain inhibitor and BACE-1. The novel compounds were found to be more active without the increasing of their molecular weight [104]. Other than these compounds, a lot of novel compounds have been discovered to be the candidates of active pharmaceutical ingredients for the treatment of arthritis [105],

obesity [106], inflammation [107-111], cardiovascular disease [112, 113], influenza [113], and liver cirrhosis [114] as listed in Table 2.2.

Table 2.2. Selected examples of the compounds discovered using structure-based drug design approach.

Disease	Target Protein	Compound Structure	Reference
HIV	HIV protease		[94]
HIV	HIV protease		[95]
HIV	HIV protease		[96]
HIV	HIV protease		[97]
Cancer	CDKs		[98]
Cancer	CDKs		[99]
Cancer	Proteasome		[100]

Table 2.2 (Continued)

Disease	Target Protein	Compound Structure	Reference
Cancer	FTase		[101]
Alzheimer's disease	BACE-1		[103]
Alzheimer's disease	BACE-1		[104]
Arthritis	Caspase-1		[105]
Obesity	PPAR γ		[106]
Inflammation	Adenosine deaminase		[107]
Inflammation	Rho kinase		[108]
Inflammation	Adenosine deaminase		[109]

Table 2.2 (Continued)

Disease	Target Protein	Compound Structure	Reference
Inflammation	Adenosine deaminase		[110]
Inflammation	Glycogen synthase kinase 3		[111]
Cardiovascular disease	Tissue factor VIIa		[112]
Influenza	Neuraminidase		[113]
Liver Cirrhosis	Hepatitis C NS3-4A protease		[114]

2.3 Applications of Protein Crystallization

As mentioned previously, the X-ray crystallography is currently the most powerful method for the determination of the three-dimensional structures of protein molecules. The detailed understanding of the three-dimensional structures of protein molecules is vital for the structure-based drug design and the understanding of genome structure-function correlation [52]. Thus, protein crystallization is always considered as the necessary first step in protein structure determination. Protein crystals were also crystalline materials that were of various chemical and physical properties, which enabled them to be potentially applicable in drug delivery and the development of novel catalytic materials. The controllable dissolution rate of protein crystals enabled them to be utilized to achieve sustained protein drug release rate. The most pioneering example was the formulation in which insulin was precipitated and crystallized with Zn(II) and protamine. The sustained release was achieved in this way to treat diabetes mellitus successfully [2, 3]. In addition, protein crystals could be strengthened both chemically and mechanically by cross-linking to work as the catalysts in chemical and environmental industry. The microporous properties of protein crystals can also be taken advantage to make protein crystals to be potential materials as chiral chromatographic media [5].

2.4 Protein Crystallization Methods

Vapor diffusion technique is the most frequently used method to crystallize protein. In this technique, a small droplet containing protein, buffer, and precipitant is placed on a siliconized glass cover slip. The concentration difference between the crystallization droplet and the reservoir drives the system towards equilibrium by diffusion. As illustrated in Figure 2.1, there are three types of vapor diffusion setups; hanging drop, sitting drop

and sandwich drop methods. In addition to vapor diffusion methods, dialysis, batch and micro-batch techniques are applied. If a protein is less soluble close to low ionic strength, it is possible to crystallize it simply by dialysis. The dialysis experiment can be set up using dialysis bag, collodion thimbles, or Zeppezauer cells [115]. The protein solution is placed in the dialysis device and dialyzed against the buffer. In batch and micro-batch technique, the precipitant and the target protein solution are simply mixed. In this way, supersaturation can be achieved directly rather than by diffusion. The goal of all these different techniques is to achieve a state of supersaturation by adjusting the solutions conditions to make crystallization possible.

Due to the lack of fundamental understanding of the mechanism of protein crystallization and the complexity of protein molecules, protein crystals are usually obtained by the trial-and-error method. Generally, the screening for protein crystallization conditions can be achieved by systematically varying solution variables, such as precipitant type and concentration, pH, temperature, *etc* [116]. Thus, a large amount of experiments are required to determine the optimal crystallization conditions. It is usually costly and time-consuming. Under the pressure that the structures of protein molecules, especially the protein targets for drug design, should be obtained in a reasonable time frame and the limited amount of proteins, crystallization robots handling crystallization reagents in nanoliter scale have been developed to speed up the screening process and reduced the consumption of proteins. C-200 robot offered by Gilson-Cyberlab (Middleton, WI, USA) is a conventional crystallization robot that requires 1 to 2 μL solution per experiment. A lot of efforts have been made to reduce the amount of crystallization solution required per experiment. For example, a crystallization robot and digital image

capture system (Agincourt) that was able to carry out 60,000 micro-vapor diffusion crystallization experiments per day was developed by Genomic Institute of the Novartis Research Foundation (San Diego, CA, USA) and Syrrx (San Diego, CA, USA) [117]. Each crystallization experiment required a droplet of 80 nL crystallization solution. However, the equipment cost of these systems is so high that many individual laboratories cannot afford. In order to overcome this problem, microfluidics has been applied to combine and confine crystallization reagents in volumes as small as or even smaller than those dispensed by robotic systems [118]. The precise control of fluid flow provided by plug-based crystallization has been shown to be a promising method to screen protein crystallization conditions.

Figure 2.1. Three types of vapor diffusion setups; hanging drop, sitting drop and sandwich drop method.

2.5 Phase Behavior of Protein Molecules

The phase behavior of colloidal particles sensitively depends on the range of attractive interaction. Generally the colloidal particles are assumed to be hard spheres. The schematic temperature-density phase diagrams for hard spheres with attractive Yukawa potentials are illustrated in Figure 2.2. In Figure 2.2 (a), gas is the stable low-density phase at high temperature. The phase separates into two co-existence phases; gas and liquid as the temperature is low. At high densities, the gas-liquid and gas-solid co-existence curves intersect at the triple point. The further increase in the density leads to the complete solidification. When the attractive interaction range is smaller than one-seventh of the diameter of the hard sphere the phase diagram undergoes a significant change. As shown in Figure 2.2 (b), the triple point disappears and the gas-liquid co-existence region moves underneath the gas-solid co-existence curve, which is similar to the phase diagrams of globular proteins as observed in previous studies [119]. Based on the previous experimental data on the phase diagrams of protein solutions, a generic phase diagram of protein molecules is depicted in Figure 2.3. The phase diagram in Figure 2.3 (a) applies to the proteins, whose solubility increases with increasing temperature, *e.g.* lysozyme [35, 37] and γ -crystallin [120], while the phase diagram in Figure 2.3 (b) applies to the proteins, whose solubility decreases with increasing temperature, for example HbS [121]. It is found that there are three regions in phase diagram, liquid-solid co-existence region, liquid-liquid co-existence region, and gelation region. The presence of the liquid-liquid co-existence and gelation regions is considered to be intrinsic features for globular proteins. As suggested by the previous studies, the locations of these phase boundaries vary with solution parameters, such as pH, temperature, protein concentration, *etc.* It appears that the detailed understanding of the phase diagram would be advantageous in determining the

optimal crystallization conditions to obtain the crystals that are suitable for the X-ray diffraction.

Figure 2.2. Schematic temperature-density phase diagrams for hard spheres with Yukawa attraction after the work reported by Muschol and Rosenberger [37]: (a) standard phase diagram with a stable gas-liquid co-existence region below the critical point (CP), (b) phase diagram when the attractive range is smaller than one-seventh of the diameter of the colloidal particle.

Figure 2.3. Schematic phase diagram for globular proteins: (a) the solubility increases with increasing temperature, (b) the solubility decreases with increasing temperature.

2.5.1 Solubility Studies

Protein solubility is a critical parameter with the advent of interest in protein crystallization process. Generally, protein solubility is either obtained by dissolution of protein crystals into the protein-free solution or crystallization of super-saturated solution to reach the equilibrium concentration. In the first technique, a known mass of protein crystals is placed in a protein-free solution under the conditions of interest and stirred to ensure the thorough mixing of proteins. Then a clear sample is removed to analyze the concentration of the protein solutions. When the concentration reaches a constant value the system is assumed to be saturated. The final concentration is noted as the solubility [122]. It is also possible to measure the solubility starting from a super-saturated solution. The equilibrium of the solution is achieved through the growth of crystals. However, it is more difficult to establish equilibrium in this way compared to the first technique. The crystal growth may cease due to the presence of the impurities on the surface of crystals.

The traditional methods as described above are usually costly and time-consuming. In order to achieve rapid and efficient solubility determination, a lot of alternative methods have been developed. Cacioppo *et al.* [6, 7] developed a micro-column technique for the rapid determination of protein solubility data. In this technique, the protein microcrystals in the column beds were equilibrated at the conditions of interest. The measurement rate was maximized and the consumption of protein was reduced. Sazaki *et al.* [123], Gray *et al.* [16] and Suzuki *et al.* [17] have reported the application of the Michaelson interferometer to the measurement of the solubility of lysozyme, equine serum albumin and glucose isomerase. The Michaelson interferometer correlates the protein solubility with the concentration gradient around a crystal that is growing, dissolving or in

equilibrium with the environment. It is rapid, allowing one solubility data point to be determined in less than 2 hours. Scintillation [124], Fourier transformed infrared spectroscopy (FTIR) [125] and isothermal micro-calorimetry [126] were also developed to measure the solubility of protein. These techniques require some specialized and expensive equipment. A spin filter method was also developed to measure the solubility of lysozyme and carboxypeptidase G₂. In this method the crystal slurry was equilibrated with protein solution and then centrifuged through a filter. This technique was rapid and does not require specialized or expensive equipment [127].

It is known that the solubility of protein molecules depends on several factors, such as pH, temperature, solvent, precipitant type, precipitant concentration, and methods of solubility measurement, *etc.* Measurements of the dependence of the protein solubility on these factors can be valuable for determining the phase behavior of a particular system and selecting suitable operation parameters. Research on the solubility of proteins in water and aqueous solutions can be dated back to more than a century ago [128]. The solubility data of hen egg white lysozyme crystals have been measured as a function of salt concentration along with other parameters, such as pH, temperature and salt type [6-12, 17]. The dependence of the solubility of concanavalin A on the concentrations of ammonium sulfate, solution pH, and temperature were measured systematically by Mikol and Giegé [13] using a microquantitation assay. The resulting phase diagram has permitted the analysis of the solubility of concanavalin A and the design of the efficient ways to control the crystallization of concanavalin A. In addition, the solubility of ovalbumin in ammonium sulfate solution [14], canavalin A [15], equine serum albumin [16], and glucose isomerase [17] were reported. As mentioned before, reliable solubility information is necessary for

the study of nucleation kinetics. The extensive solubility data of hen egg white lysozyme make it a suitable candidate in the study of the nucleation and crystal growth mechanisms.

2.5.2 Liquid-liquid Phase Separation

A detailed understanding of the phase transformations of aqueous protein solutions, especially the liquid-liquid phase separation (LLPS), is of considerable importance for protein crystallization and some human diseases, such as cataract [129]. It has been both experimentally suggested [18-24, 52] and theoretically predicted [25, 26, 28, 55] that LLPS affects protein crystallization. LLPS is also the cause of a wide class of human diseases including sickle cell disease [121, 130, 131] and cataract [120, 132]. The study of the location of the liquid-liquid co-existence boundary and the strategies to shift it provides a systematic approach to control protein crystallization and is also crucial in the search of the disease treatment.

LLPS of lysozyme in salt solutions was first reported by Ishimoto and Tanaka in 1977 [35]. Subsequently, the effects of pH, salt identity, and concentrations on the cloud-point temperature for lysozyme in electrolyte solutions were investigated systematically to provide an insight into protein interaction and its correlation with crystallization. The cloud-point temperature was measured as a function of the ionic strength of the sodium phosphate buffer, its pH, and the identity and concentration of the salts by Taratuta *et al.* [36]. It was observed that the cloud-point temperature did not change significantly with the total ionic strength of the sodium phosphate buffer at fixed pH. The cloud-point temperature increased monotonically with salt concentration. The cloud-point temperature was found to increase with pH as well. The obtained cloud-point temperature data were

found to be characterized very well by modeling the attractive interaction using thermodynamic Gibbs free energy. The phase boundaries for metastable liquid-liquid phase separation in hen egg white lysozyme solutions with 3%, 5% and 7% w/v at pH 4.5 were determined using cloud-point temperature measurement by Muschol and Rosenberger [37]. It was found that the critical temperatures for the liquid-liquid co-existence curves were increased approximately linearly with salt concentration and the critical concentration did not vary much with the salt concentration. The phase diagram of lysozyme solutions was also proposed by combining these liquid-liquid co-existence curves with the solubility curves and applied to correlate with protein crystallization. In addition, the effects of electrolyte type, concentration, solution pH, and glycerol on the cloud-point temperature of lysozyme were investigated using dynamic light scattering [37], thermooptical analysis technique [30] and micro-differential scanning calorimeter [31]. It was of great interest that the dependence of the cloud-point temperature on the salt concentrations was differed with the types of the cations and anions. For the salts with monovalent cations, the cloud-point temperature was increased with the salt concentration. For the salts with divalent cations, the cloud-point temperature was increased with the salt concentration with a maximum, which was attributed to the ion binding to the protein surface and the subsequent water structuring. At fixed salt concentrations, the cloud-point temperature was found to decrease with rising anion kosmotropic character. Moreover, the addition of glycerol was observed to decrease the cloud-point temperature of lysozyme solutions. In order to further characterize the protein-protein interactions using cloud-point temperature determination, the random phase approximation (RPA) together with the square-well potential was applied to investigate the specific effects of different ions. The

interaction energy could be fitted as a function of ionic strength, which provided an approach to predict the cloud-point temperature of lysozyme in a particular salt solution.

In recent years, such kind of studies have been extended to other proteins, such as bovine serum albumin (BSA) [38], urate oxidase [32], and bovine pancreatic trypsin inhibitor (BPTI) [33, 34]. Wang *et al.* [38] reported the effects of the concentration of polyethylene glycol (PEG) 1450 on the liquid-liquid phase separation of BSA by measuring the temperature dependence of the cloud-point temperature on the PEG 1450 concentration and BSA/PEG 1450 partitioning. A thermodynamic perturbation theory was also applied to relate the effects of PEG 1450 on the liquid-liquid phase separation of BSA to the effects of PEG 1450 to protein-protein interaction and the partition of BSA/PEG 1450 to protein-PEG interactions. Based on these results, a depletion-interaction model was found to be able to account for both protein-PEG interaction and the effects of PEG on the protein-protein interaction sufficiently. The phase diagram of urate oxidase in the presence of PEG 8000 was characterized by measuring the solubility and the dilute part of the liquid-liquid phase separation [32]. On the basis of the protein-protein interaction as determined using the SAXS and the crystal growth process of urate oxidase in PEG 8000 solutions observed by optical microscope, it was concluded that the liquid-liquid phase separation preceeded and slowed down the crystallization. The phase diagram of BPTI/350 mM KSCN at pH 4.9 was determined by measuring the solubility curve and liquid-liquid co-existence curve using optical microscope. Solutions above the liquid-liquid co-existence curve and the dilute and dense phases of the solutions below the liquid-liquid co-existence curve were characterized using SAXS. The experimental observations of optical microscopy and SAXS were coupled with the data obtained from numerical simulation. It

was found that solutions above the liquid-liquid co-existence curve were composed of both decamers and monomers. It was the attractive interactions between decamers that drive the liquid-liquid phase separation and liquid-solid separation [33, 34]. Additionally, the liquid-liquid co-existence curves for γ -crystallin, hemoglobin and their natural variants [120, 121, 130-136] were measured to provide a sound reference in the search of the treatment of cataract and sickle cell anemia.

The similarity between the phase diagrams of colloidal solutions and globular protein solutions enables the modeling of the phase diagram of protein solution using the theories developed for colloidal solutions. The first model was based on the equation of state proposed by van der Waals [137]. This equation of state was further developed to be more general by characterizing the model system with their molecular interactions and kinetic energy. Another general approach was to apply a statistical-mechanical perturbation theory. In this approach, the protein solutions were assumed to be composed of hard spheres. The interactions between these hard spheres were repulsive interactions. The attractive interactions were the perturbation to the repulsive interactions of the hard sphere system [38, 138-140]. The Gibbs free energy model [36, 55, 141], which described the water-mediated interactions between protein molecules in phenomenological level was also used to describe the properties of the observed liquid-liquid phase separation. In addition, many Monte Carlo simulations have been made to predict the liquid-liquid co-existence curve [142, 143]. All the described modeling methods are based on the assumption that the interactions between protein molecules are isotropic. However, these isotropic models only described the phase diagram of protein solutions qualitatively. The curves generated by modeling were not as wide as that determined by experimental measurement.

2.6 Second Virial Coefficient and Protein Interactions

The challenge of crystallizing proteins has led to a significant amount of research in understanding the correlation between protein interactions and the solution conditions that are conducive to crystallization [144]. As a thermodynamic parameter that described the protein-protein interactions, the second virial coefficient (B_{22}) was firstly proposed by George and Wilson [39] as a predictor for protein crystallization experiment. B_{22} represents a Boltzmann-weighted average measure of protein-protein interactions; positive value corresponds to repulsive interactions and negative value corresponds to attractive interactions [144]. In the work reported by George and Wilson, the values of B_{22} have been measured for a variety of proteins in the crystallization conditions. It was found that the values of the B_{22} for the solutions in crystallization conditions were within a narrow range of $-1 \times 10^{-4} \sim -8 \times 10^{-4} \text{ mol mL g}^{-2}$. The solutions, whose B_{22} values were out of this range would either result in no crystallization or disordered aggregates. This observation suggested a rational approach to predict whether proteins that have not been crystallized before could assemble into crystals suitable for the X-ray diffraction. Since then a lot of efforts have been made to improve the efficiency of the measurement of the second virial coefficient, investigate the influences of solution parameters, such as ionic strength, solution pH, and temperature, on protein interactions, and correlate the second virial coefficient with the conditions conducive to protein crystallization.

Usually B_{22} for protein solution is defined by the osmotic virial expansion [145] as

$$\Pi = RTc_p \left(\frac{1}{M_w} + B_{22}c_p + \cdots \right) \quad (2.1)$$

in which Π is the osmotic pressure, c_p is the protein concentration, R is the gas constant, T is the temperature and M_w is protein molecular weight. The sum of the pairwise potential of protein molecules can be related to the second virial coefficient by [146]

$$B_{22} = -\frac{1}{2} \frac{N_A}{M_w^2} \int_0^\infty 4\pi r^2 (\exp[-W(r)/kT] - 1) dr \quad (2.2)$$

where N_A is the Avogadro's number, $W(r)$ is the sum of pairwise potentials of protein molecules and r is the center-to-center distance between two protein molecules. The sum of pairwise potentials is given by

$$W(r) = W_{hs}(r) + W_{elec}(r) + W_{disp}(r) + W_{spec}(r) \quad (2.3)$$

where $W_{hs}(r)$ is the hard-sphere potential, $W_{elec}(r)$ is the electric double-layer-repulsion potential, $W_{disp}(r)$ is the dispersion potential and $W_{spec}(r)$ is the specific potential that describes the hydrophobic interactions. The hard-sphere potential is the contribution for the size of the particles only and given by

$$W_{hs}(r) = \begin{cases} \infty, & \text{for } r \leq \sigma \\ 0, & \text{for } r > \sigma \end{cases} \quad (2.4)$$

with σ the effective hard-sphere diameter of the protein molecules.

According to Debye-Huckel theory, the electric double-layer repulsive can be calculated by

$$W_{elec}(r) = \frac{z^2 e^2 \exp[-\kappa(r - \sigma)]}{4\pi\epsilon_o\epsilon_r(1 + \kappa\sigma/2)^2} \quad (2.5)$$

where z is the valence of the protein molecule in the solution condition, e is the elementary charge, ϵ_o is the dielectric permittivity of free space, ϵ_r is the relative dielectric permittivity of water, κ is the inverse Debye length given by

$$\kappa^2 = \frac{2e^2 N_A I}{kT \epsilon_o \epsilon_r} \quad (2.6)$$

in which I is the ionic strength of the solution and k is the Boltzmann constant. The attractive dispersion potential can be calculated as

$$W_{disp}(r) = -\frac{H}{36} \left(\frac{\sigma}{r} \right)^6 \quad (2.7)$$

where H is the effective Hamaker constant for protein interactions. The specific interaction can be calculated by applying the square-well potential to characterize the specific interactions as

$$W_{spec}(r) = \begin{cases} -\epsilon_{spec}, & \text{for } \sigma \leq r \leq \sigma + \delta \\ 0, & \text{for } r < \sigma + \delta \end{cases} \quad (2.8)$$

In this equation, ϵ_{spec} is the depth of the square well and δ is the width of the square well.

The values of B_{22} were able to be measured by SLS, membrane osmometry, SAXS, SANS, self-interaction chromatography, size exclusion chromatography, and self-interaction nanoparticle spectroscopy. The principles of the measurements of the second virial coefficient using these techniques and their applications will be summarized as next sections.

2.6.1 Static Light Scattering (SLS)

Static light scattering (SLS) is a traditional way to measure the second virial coefficient (B_{22}). It measures the time-averaged intensity of the scattered light and relates it to the osmotic compressibility and the B_{22} of the interacting particles. For a system containing water, polymer and salt, the light scattering equation is given by [147]

$$\frac{Kc_p}{R_\theta} = \frac{1}{M_w} + 2\frac{N_A B_{22}}{M_w^2} c_p + 3B_3 c_p^2 + \dots \quad (2.9)$$

where B_3 is the third virial coefficient. If the scattering molecules are smaller than 1/20 of the wavelength of the incident laser light beam, the light scattering does not have the angular dependence. Therefore, the measurement is usually carried out at the scattering angle of 90° . The second virial coefficient can be determined by truncating after the first-order term [147, 148] as

$$\frac{Kc_p}{R_\theta} = \frac{1}{M_w} + 2\frac{N_A B_{22}}{M_w^2} c_p \quad (2.10)$$

Here, K is the system-specific constant

$$K = \frac{4\pi^2 n_0^2}{N_A \lambda^4} \left(\frac{dn_p}{dc_p} \right)^2 \quad (2.11)$$

with the solvent refractive index n_0 , the wavelength λ , and the refractive index increment of the protein (dn_p/dc_p) . R_θ is the excess Rayleigh scattering ratio of the polymer solutions over the salt solution [146]. In a typical SLS measurement, the excess scattering ratio Kc_p/R_θ of protein solutions was plotted against the protein concentration to obtain the molecular weight and the second virial coefficient.

In the past few years, the effects of solution parameters on protein interactions in protein solutions have been investigated by measuring the second virial coefficient using SLS. Using SLS, Rosenbaum *et al.* [41] measured lysozyme second virial coefficients at pH 4.6 and pH 7.4 as a function of ionic strength and temperature and described the protein interactions using the adhesive hard sphere potential. The result suggested that

simple models for protein interactions could be useful for investigating protein crystallization. The effects of pH and electrolyte concentration on protein interactions in lysozyme and chymotrypsinogen solutions were investigated by static light scattering [40]. At low electrolyte concentration, second virial coefficient depended on pH and change from positive to negative as the pH increases. All coefficients at high salt concentration were slightly negative and depend weakly on pH. The addition of non-adsorbing polymers to protein solutions was found to induce protein attractive interaction due to the unbalanced osmotic pressure arising from the exclusion of polymer molecules from the region between proteins [149]. The interactions between protein molecules in the presence of poly (ethylene glycol) (PEG) were probed through the measurement of the second virial coefficient using static light scattering. The obtained experimental data were compared with the predictions from Asakura-Oosawa (AO) model and polymer reference interaction site model (PRISM). The comparison suggested that these two models failed to predict the strength of the protein interactions both quantitatively and qualitatively. A better understanding could be achieved through incorporating in a simple mean-field manner the changes in polymer solution screening length as the spinodal for the polymer-solvent demixing transition was approached [149]. As a polyol that stabilized the protein structure, the influence of glycerol on the interactions between bovine pancreatic trypsin inhibitor (BPTI) molecules was investigated. It was suggested that the addition of the glycerol would increase the repulsive interactions between BPTI molecules [42]. In addition, the effect of secondary structure on the potential mean force for poly-L-lysine [146], the influence of alcohols on the lysozyme interactions [150] and nonequivalence of second virial coefficients of lysozyme and ovalbumin [151] were investigated using static light scattering.

2.6.2 Small Angle X-ray Scattering (SAXS)

With solutions of mono-disperse spherical particles, the total intensity $I(c, s)$ scattered at a scattering angle of 2θ is expressed as a function of scattering vector s , $s = 2 \sin \theta / \lambda$, by

$$I(c, s) = I(0, s) \cdot S(c, s) \quad (2.12)$$

in which $I(0, s)$ is the form factor and $S(c, s)$ is the structure factor that characterizes the particle interactions. This equation works for quasi-spherical and/or poly-disperse particles as well. Thus the interactions between the spherical or quasi-spherical protein molecules can be analyzed using SAXS as well. The X-ray structure factor at $s = 0$ can be related to the osmotic pressure by

$$S(c, 0) = (RT / M_w)(\partial \Pi / \partial c)^{-1} \quad (2.13)$$

With weak interactions and at low protein concentrations, the second virial coefficient can thus be measured by [43]

$$\frac{1}{S(c, 0)} = 1 + \frac{2M_w B_{22}}{c} \quad (2.14)$$

The interactions of urate oxidase interactions have been investigated by measuring the second virial coefficients using small angle X-ray scattering (SAXS) [43, 44] as a function of solution pH, temperature, precipitant, and the addition of poly(ethylene glycol) (PEG). The crystallization experiments in the corresponding conditions were performed in parallel to correlate the second virial coefficient with the crystallization slot. The B_{22} values were found to be in the range of $-0.4 \times 10^{-4} \sim 0.1 \times 10^{-4} \text{ mol mL g}^{-2}$. By comparing this range with the crystallization slot proposed by George and Wilson [39], these second virial

coefficient values were at the low end of the crystallization slot. The author also proposed the use of the dimensionless second virial coefficient to predict the successful crystallization conditions.

2.6.3 Small Angle Neutron Scattering (SANS)

Small angle neutron scattering (SANS) intensity distribution from protein solutions is given by [40]

$$I(q) = \rho P(q) S(q) \quad (2.15)$$

where ρ is the molecular protein concentration, q is the scattering vector that is a function of the scattering angle θ and wavelength λ . $P(q)$ is the form factor that is an orientation-averaged function of the size and shape of protein molecules. $S(q)$ is the structure factor that is related with the potential of mean force between protein molecules. Both the form factor and structure factor vary with the scattering vector q . The form factor could be approximated to be 1 when q approaches to 0. In this limit, the structure factor

$$S(0) \cong (1 + 2B'_{22}\rho)^{-1} \quad (2.16)$$

in which B'_{22} is the dimensionless second virial coefficient. The second virial coefficient can be obtained from the structure factor using an extrapolation of the SANS data as $q \rightarrow 0$ via a polynomial least squares fit approximation. The influence of 2-methyl-2,4-pentanediol on the *Hm* MalDH interactions was also probed by measuring the second virial coefficients using small angle neutron scattering.

The work described above was based on the scattering of the protein molecules in the solution. However, these techniques are usually costly and protein-consuming. Therefore a lot of efforts have been made to invent the new techniques, such as self-interaction chromatography, size exclusion chromatography and self-interaction nanoparticle spectroscopy to measure the second virial coefficient.

2.6.4 Self-interaction Chromatography (SIC)

A novel technique of measuring the second virial coefficient (B_{22}) was proposed by Tessier *et al.* [45] in order to achieve a rapid measurement of B_{22} . In this method, protein molecules were immobilized on the chromatographic particles and the retention of the same protein was measured in isocratic elution. The relative retention of the protein was related with B_{22} . The chromatographic retention was characterized experimentally in terms of retention factor

$$k' = \frac{V_r - V_o}{V_o} \quad (2.17)$$

in which V_r is the retention volume of a solute to be eluted from the column, and V_o is the retention volume of the solute that is of free interaction with the packing particles of the column to be eluted. The second virial coefficient can be calculated from the retention factor by

$$B_{22} = B_{22}^{HS} - \frac{k'}{\rho_s \phi} \quad (2.18)$$

where ϕ is the phase ratio and defined as

$$\phi = \frac{A_s}{V_o} \quad (2.19)$$

with A_s the total available surface area and ρ_s the number of immobilized molecules per unit area. B_{22}^{HS} is the hard sphere contribution and defined as

$$B_{22}^{HS} = \frac{2\pi\sigma^3}{3} \quad (2.20)$$

where σ is the diameter of protein molecule.

The second virial coefficient values for lysozyme and chymotrypsinogen over a wide range of pH and ionic strengths were measured by self-interaction chromatography [45]. The results obtained were comparable to those that were determined using traditional scattering methods. It was suggested that self-interaction chromatography would be an efficient alternative that was time and protein saving to the traditional methods. The application of the self-interaction chromatography was explored to measure the second virial coefficient values of ovalbumin, ribonuclease A, bovine serum albumin, α -lactalbumin, myoglobin, cytochrome c and catalase [46]. It was shown that, at low salt concentrations protein-protein interactions were either attractive or repulsive, possibly due to the anisotropy of the protein charge distribution. The influences of different salts on the interactions of protein molecules followed the Hofmeister series, which can be interpreted as a water-mediated effect between protein and salt molecules. The effects of pH on protein interactions were investigated for ovalbumin, catalase, β -lactoglobulin A and B, ribonuclease A and lysozyme [47]. For ovalbumin and catalase, the protein interactions were revealed to become increasingly attractive with decreasing pH.

2.6.5 Size Exclusion Chromatography (SEC)

The second virial coefficient can also be determined from the retention time measurement in size exclusion chromatography. The solute retention time was known to be dependent on the solute's size sensitively in size exclusion chromatography. And the thermodynamic non-ideality was also found to result in the concentration dependent retention times, which can be applied to quantify the second virial coefficient. According to Bloustine *et al.* [152], the second virial coefficient (B_{22}), the local solute distribution coefficient (K_D), the partition coefficient of solute molecules between chromatographic phases in the limit of infinite dilution (K_0), and the average concentration of the mobile phase in the pulse $\langle C_i \rangle$ can be correlated as

$$\ln\left(\frac{K_D}{K_0}\right) = 2B_{22} \frac{N_A}{M_w} \langle C_i \rangle (1 - K_D) \quad (2.21)$$

The local distribution coefficient K_D is given by

$$K_D \equiv \frac{C_p}{C_i} = \frac{t_r - t_o}{t_T - t_o} = \frac{V_r - V_o}{V_T - V_o} \quad (2.22)$$

where t_r and V_r are the solute retention time and volume, t_o and V_o are the retention time and volume of completely excluded molecules, and t_T and V_T are the retention time and volume of the completely included molecules. $\ln K_D$ is plotted as a function of $\langle C_i \rangle (1 - K_D)$. The second virial coefficient can be obtained from the slope of the plot.

Size exclusion chromatography was also applied to determine the second virial coefficient from retention time measurements for lysozyme and bovine serum albumin solutions. The good agreement between the values obtained from size exclusion

chromatography and light scattering suggested that the size exclusion chromatography could be an efficient alternative as well [152].

2.6.6 Nanoparticle-based Assay

In a more recent report, Tessier *et al.* [48] have invented a nanoparticle-based assay to guide the screening of the crystallization conditions for the proteins that have not been crystallized before. In this method, the protein molecules were conjugated with the gold nanoparticles. The protein/gold conjugated particles were then added to the solutions of interest for crystallization. It was found that the optical properties of the particles are sensitive to the interparticle separation distance. By testing the assay with bovine serum albumin and ovalbumin, the solutions whose second virial coefficients were in the crystallization slot as revealed by George and Wilson [39] led to the maximum change of the color of the suspension particles. This method has presented a remarkable efficiency. Given this, this method may be a potential alternative to the traditional methods in the screening of the crystallization conditions for protein molecules.

2.7 Aggregation and Pre-nucleation Behaviors of Protein Molecules

The crystallization of protein molecules is known to begin with the aggregation of protein molecules [153]. In order to obtain a detailed understanding of protein crystallization process, it is necessary to investigate the pre-nucleation behavior of protein molecules. A popular method of studying the pre-nucleation behavior of protein molecules is dynamic light scattering (DLS). In this method, the fluctuations in the average intensity of scattered light arising from different scattering sources are measured. These fluctuations are mainly the results of the Brownian motion of the protein molecules. The time

dependence of these fluctuations can be related to the translational diffusion coefficient of the protein molecules [154]. In order to quantify the time dependence of these fluctuations, the intensity autocorrelation function is measured and the correlation time of the intensity fluctuations is obtained from the time decay. For a mono-disperse solution of protein molecules, the intensity autocorrelation function is defined for a signal $I(t)$ as

$$g(\tau) = \lim_{T' \rightarrow \infty} \frac{1}{2T'} \int_{-T'}^{T'} I(t)I(t+\tau)dt \quad (2.23)$$

where t is the time and $I(t+\tau)$ is the signal delayed by a time τ . For the mono-disperse solution, the autocorrelation function can be expressed as

$$G(\tau) = A[1 + Be^{-2\Gamma\tau}] \quad (2.24)$$

in which Γ is the decay constant, A is the background, and B is a constant depending on the instrument parameters. The decay constant can be correlated with the scattering vector as

$$\Gamma = Dq^2 \quad (2.25)$$

with the translational diffusion coefficient D and the scattering vector q given by

$$q = 4\pi \frac{n_0}{\lambda} \sin \frac{\theta}{2} \quad (2.26)$$

where n_0 is the refractive index of the solvent, λ is the wavelength of the extinction light source, and θ is the scattering angle. The hydrodynamic radius of the particle R_H can be related to the diffusion coefficient through Stokes-Einstein equation using

$$D = \frac{k_B T}{6\pi\eta R_H} \quad (2.27)$$

where k_B is the Boltzmann's constant, T is the absolute temperature, and η is the viscosity of the solvent. Using DLS, the diffusion coefficient and the hydrodynamic radius

of the protein molecules are probed to provide an insight into the pre-nucleation process of protein crystallization.

Using DLS, the dynamics of lysozyme precrystallization cluster formation was studied by Georgalis *et al.* [155-157]. It was found that individual lysozyme molecules aggregated rapidly and formed large precrystallization fractal clusters in the initial stages of crystallization. The growth of these clusters took place either in the reaction limited aggregation regime (RLCA), diffusion limited aggregation regime (DLCA) or the crossover regime between the DLCA and RLCA (XOVER). The crystals that were of X-ray diffraction quality grew from the DLCA regime. Based on these observations, it was proposed that a systematic investigation on the influences of the parameters, such as the growth regime and the limiting size of the clusters, on the crystallization of protein molecules would be beneficial in the control of the protein crystallization in its early stages. These studies were extended by Georgalis and her colleagues to determine the optimal protein crystallization conditions by studying the influences of the electrolyte type and concentration and protein concentration [158, 159]. Both static and dynamic light scattering were applied to investigate the crystallization behavior of lysozyme in NaCl and $(\text{NH}_4)_2\text{SO}_4$ solutions of various lysozyme and electrolyte concentrations. DLS measurement revealed that mass fractals initially undergo diffusion-limited-like aggregation coexisted with individual lysozyme molecules and oligomers. SLS experiments also revealed the progressive restructuring or growth of compact structures at later stages of the aggregation. On this basis, a tentative scheme was proposed for lysozyme crystallization to explain the association of individual lysozyme molecules to form oligomer, nuclei, fractal, gel or crystals. The variation of the friction factor was

measured by DLS during the isothermal growth of lysozyme crystals as well [160]. It was observed to increase to a maximum as the crystal form and then decrease to a constant value. In addition, by conducting the kinetic study on the early stages of the crystallization process of orthorhombic and needle-like lysozyme crystals, Tanaka *et al.* [161] suggested that the early stages of the crystallization of both crystals could be explained by a diffusion-limited aggregation model and the random aggregates were formed in the early stage of the crystallization regardless of the morphology of the crystal.

Other than lysozyme, DLS has been applied to study the precrystallization behavior of other proteins. Kadima *et al.* [162] have investigated the aggregation processes of canavalin via DLS. The sizes of aggregates formed in solutions of various pH, salt concentrations, and protein concentrations were deduced from the fluctuation intensity correlation functions measured for the solutions of the protein. Results indicated that the barrier to crystallization of canavalin was the formation of the trimer. These researchers also extended this study to 2Zn-insulin hexamer and compared with canavalin. Different from canavlin, the 2Zn-insulin hexmer did not aggregate into higher oligomers, which suggested that the composition of a protein, its organization, and the characteristics of the internal bonding interactions were important factors that influenced the aggregation of the protein molecules. The differences in the molecular association behaviors of 2Zn-insulin hexamer and canavalin could be attributed to the differences in the internal bonding. It was the metal that played an important role in the assembly of insulin molecules and covalent bonds that led to the aggregation of canavalin [153]. In addition, the influences of several factors, such as the protein concentration, the salt type and concentration, pH, temperature, and the addition of polyols on the precrystallization behaviors of aminoacyl-tRNA

synthetase [163], ribonuclease [154], and β -galactosidase [164] were diagnosed with DLS. The previous work on the application of DLS to access the prenucleation and crystallization behavior of protein molecules work suggests that DLS is a promising technique in the understanding of the aggregation, cluster formation, and nucleation mechanism of protein molecules.

2.8 Nucleation Studies

2.8.1 Classical Nucleation Theory

Nucleation has always been recognized as the necessary first step in protein crystallization process. It plays an important role in the number, size, and quality of the crystals obtained. The condition of supersaturation alone is not sufficient to cause the nucleation. Nucleation may occur spontaneously or may be induced by some external stimulus, such as agitation, mechanical shock, friction, *ect.* In classical nucleation theory, the nucleus is assumed to be of spherical shape. The free energy barrier for homogenous nucleation is considered to consist of the volume excess free energy and the surface excess free energy. The volume excess free energy is due to the aggregation and the change of phase. The surface excess free energy is due to the creation of a new surface. The total free energy is given by [165]

$$\Delta G = \Delta G_s + \Delta G_v = 4\pi a^2 \gamma + \frac{4}{3}\pi a^3 \Delta G_v \quad (2.28)$$

where ΔG_v is the free energy change of the transformation per unit volume, a is the radius of spherical nucleus and γ is the interfacial tension. The maximum of the free energy barrier corresponding to the formation of the critical nucleus can be obtained by setting

$$\frac{d\Delta G}{da} = 8\pi a \gamma + 4\pi a^2 \Delta G_v = 0 \quad (2.29)$$

Thus the radius of the critical nucleus is

$$a_c = -\frac{2\gamma}{\Delta G_v} \quad (2.30)$$

The critical free energy barrier is

$$\Delta G_{crit} = \frac{4\pi a_c^2 \gamma}{3} \quad (2.31)$$

The supersaturation ratio S can be related with the size of the critical nucleus a_c in

$$\ln\left(\frac{c}{c_e}\right) = \ln S = \frac{2\gamma}{kT a_c} \quad (2.32)$$

in which v is the volume of the molecule, k is the Boltzmann's constant and T is the absolute temperature. Therefore, the critical free energy barrier can be related to the supersaturation ratio by substituting the equation above into equation (2.30)

$$\Delta G_{crit} = \frac{16\pi\gamma^3 v^2}{3 (kT \ln S)^2} \quad (2.33)$$

The homogeneous nucleation can be expressed by

$$J = \left(\frac{kT}{h}\right) N_1 \exp\left(\frac{-\Delta G_a}{kT}\right) (4\pi a_c^2) \exp\left(\frac{-16\pi\gamma^3 v^2}{3k^3 T^3 (\ln S)^2}\right) \quad (2.34)$$

with the Planck's constant h , the energy barrier of the diffusion from the bulk solution to the cluster ΔG_a and the number of the molecules N_1 . This equation applies to both homogeneous and heterogeneous nucleation. According to classical nucleation theory, the parameters, *e.g.* solution pH, temperature, the presence of foreign surfaces, *etc.*, affect the nucleation kinetics by influencing the supersaturation and interfacial tension. A detailed understanding of the nucleation kinetics is advantageous in the pursuit of the growth of the

protein crystals that are of suitable quality for the X-ray diffraction. Many investigations regarding protein crystal nucleation kinetics have been reported in the past few decades. In the next few paragraphs, these studies will be summarized. The first part is mainly on the influences of the solutions parameters on the homogeneous nucleation kinetics by affecting the supersaturation. The effects of the foreign nucleants on the heterogeneous nucleation kinetics through influencing the interfacial tension will be summarized in the second part. The third part will cover the influence of seeding on protein crystal nucleation.

2.8.2 Nucleation Kinetics Studies

According to classical nucleation theory, many factors, such as solution temperature, pH, *etc.*, affect nucleation rates. In this respect, an understanding of the influences of these factors on the nucleation rate is important for the control of the size and quality of the crystals to be obtained. Some previous studies have focused on the effects of solution parameters on the nucleation rates of protein crystals. The effects of temperature, solution pH, and supersaturation on the nucleation of tetragonal lysozyme crystals were investigated by Judge *et al.* [51] by counting the number of crystals formed in batch crystallization plates after a prolonged period. It was found that the initial supersaturation and temperature did affect the number of the crystals obtained. The number of the crystals formed in batch crystallization plates was found to increase with the increase of supersaturation. The solution pH greatly affected protein nucleation due to the charge changes on the protein's surface. At constant supersaturation, the number of crystals formed decreased with the increase of solution pH. Crystal size also increased substantially with the increase of pH. Supersaturation was also found to have a greater

effect on the crystal size at a higher pH and a smaller effect on the crystal size at a lower pH.

A temperature jump technique was developed by Galkin and Vekilov [19, 20, 52] to study nucleation kinetics of hen egg white lysozyme at 12.6 °C. In this technique, the temperature was first lowered to the nucleation temperature and then raised to a higher temperature at which nucleation rate was practically zero but crystals that already formed could grow to detectable size. The nucleation rates at the nucleation temperature could be obtained through dividing the number of crystals formed by the nucleation time. The nucleation rates of hen egg white lysozyme at different supersaturations as determined using this technique indicated a broken dependence on supersaturation, which was beyond the prediction of classical nucleation theory. The broken dependence was attributed to the existence of the liquid-liquid phase separation. The nucleation rates passed a maximum in the vicinity of the liquid-liquid phase boundary and decreased when the conditions were deep into the liquid-liquid co-existence region. The influence of glycerol and polyethylene glycol on changing the location of liquid-liquid co-existence boundary and nucleation rates of lysozyme was also investigated in this method [21]. It was suggested that the systematic investigations on the relationship between the liquid-liquid phase separation and nucleation kinetics would be advantageous in the control of protein crystals.

The nucleation rate data were also obtained using a laser particle counter [53, 54]. The instrument used the principles of near angle light scattering to detect particles in the solution. As the revolving laser beam passed through the particles in the sensing zone, light from the beam would be scattered. The scattered light was collected and caused

electrical pulses. The number and amplitude of the pulse could be used to measure the number and size of the particles. In this method, the initial nucleation rates of hen egg white lysozyme at different NaCl concentrations, 4 °C and pH 4.5 with 0.1 M sodium acetate buffer were measured. The nucleation rates varied with supersaturation as predicted by classical nucleation theory. However, there was a broken dependence of the nucleation rates on the supersaturation due to the occurrence of liquid-liquid phase separation, and the liquid-liquid phase separation also affected the morphology of the crystals obtained. Comparing the obtained nucleation kinetics data with the classical nucleation theory, a conclusion was drawn that there might be two mechanisms; heterogeneous nucleation at low protein concentration and homogeneous nucleation at high protein concentrations. In addition, the surface energy was revealed to vary little with ionic strength.

2.8.3 Heterogeneous Nucleation

One of the obstacles in the structure determination of protein molecules through the X-ray crystallography is either getting no crystals or getting the crystals that are not of the suitable quality for the X-ray diffraction. Protein crystallization is usually achieved by inducing spontaneous nucleation at appropriate supersaturations. However, excessive supersaturation can also increase the rate of crystal growth, which would result in the crystals that are of inadequate quality for the X-ray diffraction. Therefore, protein crystallization is preferred to be induced in metastable conditions to obtain the crystals that are of larger size and better X-ray diffraction quality. Since heterogeneous nucleation can decrease the free energy for nucleation and grow protein crystals at low supersaturation,

usually in metastable conditions, many experimental and theoretical efforts have been made regarding this issue.

Table 2.3. Heterogeneous nucleation surfaces studies summarized in this chapter.

Protein	Surfaces	Ref
Canavalin	Mineral surfaces	[56]
Concanavalin B		[59]
Catalase		[60]
Lysozyme		
Lysozyme	Mineral substrates	[57]
Conalbumin		
Lysozyme	Mineral substrates	[58]
Lysozyme	Poly-L-lysine modified glass slides	[61]
		[62]
Lysozyme	Bare glass slide; Silanized surfaces; Surface covered by -CH ₃ ; Glass slide coated by Cr or Ag; Surface treated with poly-L-lysine	[166]
Lysozyme	Behenic acid film	[167]
		[168]
Streptavidin	Biotinylated lipids	[65]

Table 2.3 (*Continued*)

Protein	Surfaces	Ref
Lysozyme	Silanized mica;	[67]
Trichosanthin	Silanized glass cover slips	
K115		
Hoat		
K169		
Lysozyme	Silanized low-protein-binding polystyrene wells	[68]
Concanavalin A	Polymeric films	[69]
Lysozyme		
Lysozyme	Porous glass substrate	[169]
Thaumatococcus		
Apoferitin		
Lysozyme	Porous silicon	[70]
Trypsin		
Catalase		
Thaumatococcus		
Phycobiliprotein		
Concanavalin A		
Lysozyme	Poly(vinylidene fluoride) membranes	[170]
Lysozyme	Chemically modified glass slides	[171]

Table 2.3 (*Continued*)

Protein	Surfaces	Ref
Lysozyme Concanavalin A Thaumatococcus	Chemically modified mica	[66]
Lysozyme Thaumatococcus Trypsin c-phycocyanin α -actin-binding protein myosin-binding protein-C Lobster α -crustacyanin	Gel-glass	[71]

As summarized in Table 2.3, various materials have been tested as heterogeneous nucleation surfaces for protein crystallization in the past two decades. The first attempt to nucleate protein molecules on heterogeneous surfaces was reported by McPherson and Shlichta in 1988 [56]. In their work, 50 mineral samples were tested as potential heterogeneous or epitaxial nucleants for canavalin, concanavalin B from jack bean, beef liver catalase and hen egg-white lysozyme. For each protein, it was found that there was a set of mineral substrates that promoted the nucleation. The nucleation of protein crystals was affected by the modification of crystal habit, promotion of the unit cell properties [17] or involving of a direct lattice match. After this initiative, a lot of efforts were also made to further investigate the key factor of the exogenous growth effects of the mineral surfaces [57-60]. The effects of exogenous mineral substrates were found to be more pronounced in

the case of lattice matching than that of the minerals exhibiting lattice mismatching. As one of the candidates of epitaxial nucleants, poly-L-lysine [61, 62] coated surfaces selectively controlled the orientation of lysozyme crystals to either (110) or (101) form depending on the solution conditions. The adhesion between the resulted crystals and the surfaces could even be correlated with the bond energy in crystal lattice to investigate the protein contact [166]. In the work reported by Kubo *et al.* [167, 168] that the coverage and surface morphology of behenic acid film influenced the orientation of the lysozyme crystals strongly but not the nucleation rate. Biotinylated lipids were also tested as the epitaxial nucleant for streptavidin. It was observed that biotinylated lipids nucleated the growth of the 2D crystals of streptavidin. Seeded by these 2D crystals, the 3D crystals suitable for the X-ray diffraction were obtained in the conditions at which the crystals did not grow from the solution in the absence of 2D crystals.

Other than these epitaxial nucleants, some surfaces were also used as heterogeneous nucleants whose effectiveness was due to the non-specific attractive and local interactions between protein and the surfaces. For example, freshly cleaved mica surfaces were chemically modified using mixtures of *n*-propyltriethoxysilane and 3-aminopropyltriethoxysilane (APTES) in different percentages ranging from 0% to 100% [66] and 3-aminopropyltriethoxysilane (APTES) [67]. The surfaces treated in the first method worked as the heterogeneous nucleants for lysozyme, concanavalin A and thaumatin. A shorter induction time and smaller starting protein concentrations were observed with an increase of the density of the ionisable groups on the mica surface. And the experimental results have shown several advantages of the APTES-modified mica surfaces over the conventional silanized glass cover slips on the heterogeneous nucleation

of the proteins including lysozyme, trichosanthin, K115, Hoat and K169 with a shorter induction time, a larger crystal size and a smaller number of crystals. The use of silanized low-protein-binding polystyrene wells also led to a transition from homogeneous nucleation to heterogeneous nucleation, which was not observed for the untreated polystyrene wells [68]. In addition, it was suggested that polymeric films, such as sulfonated polystyrene, cross-linked gelatin films with adsorbed poly-L-lysine or entrapped poly-L-aspartate and silk fibroin with entrapped poly-L-lysine or poly-L-aspartate [69], would be useful for inducing the crystallization of protein crystals at low starting protein concentration. Their effects were attributed to the promotion of the molecular collisions and the clustering with the symmetry for the formation of the crystal nuclei caused by the non-specific attractive and local interactions between the protein and the film surface.

Moreover, the presence of a surface microstructure conducive to nucleation was also considered as an approach in the pursuit of the universal heterogeneous nucleant for protein crystals. Porous silicon [70] was successfully applied to induce the nucleation of lysozyme, trypsin, catalase, thiamin, phycobiliprotein, and concanavalin A in conditions where the solutions remain metastable in the absence of the nucleant. Its effectiveness was due to the size distribution of the pores. Only the pores of the sizes most suited to protein molecular diameter and the shape of the initial aggregates that it formed could induced efficient nucleation. Gel-glass grain [71] and porous glass substrate [169] was also found to be efficient in inducing the heterogeneous nucleation. In a more recent report, Curcio *et al.* [170] has investigated the influence of the structural properties of poly(vinylidene fluoride) (PVDF) membranes on the heterogeneous nucleation of lysozyme crystals. They

related the free energy barrier for nucleation to the contact angles between the protein solution and the surface and the porosity of the surface. From a theoretical point of view, the free energy barrier increased at higher contact angles and decreased at higher porosity. The experimental results were consistent with the predictions. Inspired by this previous work, Liu *et al.* [171] investigated the influence of the surface roughness, topography, and physicochemical properties of five chemically modified surfaces on the heterogeneous nucleation of lysozyme crystals. In their work, the glass slides were treated with poly-L-glutamic acid (PLG), poly(2-hydroxyethyl methacrylate) (P2HEMA), poly(methyl methacrylate) (PMMA), poly(4-vinyl pyridine) (P4VP) and (3-aminopropyl)triethoxysilane (APTES). Based on the nucleation kinetics, contact angle between crystallization solution and the surface, and surface roughness, the free energy barrier was correlated with the contact angle and surface roughness. The free energy barrier was found to increase at higher contact angles and decrease at higher surface roughness. The surface with specific topography was also expected to increase the possibility of the formation of a critical nucleus.

On the basis of these studies, it was concluded that different surfaces may affect heterogeneous nucleation through various mechanisms; for example, modification of the supersaturation profile near the surface due to the concentration polarization, adsorption of the solute onto the surface, specific interactions with the solute resulting from the surface chemistry characteristics, the presence of a surface microstructure conducive to nucleation, or introduction of spatial characteristics related to the crystalline lattice [58, 170, 171]. So far, the exact mechanisms of heterogeneous nucleation on surfaces are still not clear, and none of the studied surfaces has been shown to work as a universal nucleant. The advent

of the manufacturing techniques of nanostructures may suggest a promising future in the endeavor in the pursuit of finding the universal nucleant.

The addition of foreign seeding particles is another way to induce the heterogeneous nucleation. According to classical nucleation theory, the addition of small ‘foreign seeding’ particles to a supersaturated solution can greatly increase the rate, at which crystals nucleate [172]. The presence of a foreign seeding particle that is larger than the critical nucleus can also increase the possibility to induce the heterogeneous nucleation in metastable conditions. In order to enhance the nucleation in the crystallization solutions, where protein crystals cannot grow from, some effort has been investigated on the effects of foreign seeding particles on the nucleation. For example, D’Arcy *et al.* [75, 76] have introduced the horse hair into the solutions in metastable conditions successfully to induce the crystallization of the test proteins including glucose isomerase, lysozyme, trypsin and Fab-D.

2.8.4 Effects of Seeding on the Nucleation of Protein Crystals

The introduction of seeds into the solution is known to be able to induce higher local concentrations so that the free energy barrier for nucleation can be lower and the nucleation can be induced [76]. Thus, seeding is an efficient method to influence the nucleation and growth of protein crystal. Some researchers have tried to apply seeds to control the nucleation and improve the quality of protein crystals. Bancel *et al.* [73] reported a novel approach, in which a focused laser beam was used to select and transfer single microscopic seed crystals in a growing solution. They successfully seeded and grew crystals of tomato bushy stunt virus. Different from conventional method of seeding drops,

a dynamic seeding method was reported by Zhu *et al.* [74]. This method simplified not only the control of the number of micro-seeds but also the manipulation of macro-seeds. It was applied in the crystallization of novel protein CutCm, which proved that the method was useful for optimizing protein crystallization. A repeated seeding technique [173] was developed to grow large single crystals of globular proteins. In addition, a small and carefully washed crystal was used to seed a protein solution. After the growth stopped, the crystal was removed and inserted into a fresh protein solution, which enabled the further growth of the crystal. This technique applied to several proteins, suggesting it might be a promising way to improve the quality of crystals once the small protein crystals were obtained. As a technique, which permits the decoupling of crystal nucleation and growth, seeding is a potentially powerful tool for the growth of the crystals that are of suitable quality for the X-ray crystallography, but it does require much more efforts on the manipulation of the seeds.

2.9 Protein Crystal Growth

It is well recognized that a detailed understanding of the fundamental aspects of protein crystal growth should not only provide guidelines for growing better crystal, but also provide insights into the mechanisms dominating the perfection of protein crystals. Since 1950s, both theoretical and experimental efforts have been made to understand the process of protein crystal growth.

Tetragonal lysozyme crystal is the most investigated protein crystal for growth studies [174]. The growth rates were usually obtained by measuring the changes in the dimensions of a particular face of protein crystal using optical microscopy. In this way, the influences

of temperature, pH, and salt concentrations [79-81] on the (110) and (101) face growth rates of tetragonal lysozyme crystals were investigated and correlated with the quality of crystal growth. It was found that the growth rate and crystal habit were strongly influenced by protein concentration. A slow growth rate produced more ordered crystals. Different crystal growth models were employed to analyze and interpret the mechanism of protein crystal growth. However, there is still no generally applicable model for the range of conditions studied. The most likely interpretation was that a defect-mediated mechanism dominated at low supersaturation and two-dimensional nucleation dominated at high supersaturation. Through modeling the growth rates of tetragonal lysozyme crystals, Li *et al.* proposed the hypothesis that the growth units for protein crystals might be larger than tetramer [175]. This hypothesis was further supported by the periodic bond analysis carried out by Nadarajah *et al.* [174]. A complete periodic bond chain analysis of tetragonal lysozyme crystals was carried out and coupled with an approach incorporating the molecular orientations of the crystal structure. The analysis revealed the growth unit for the (101) surface of tetragonal lysozyme was tetramer while that for the (110) surface the growth unit was octamer.

In addition to crystal size and surface morphology, better X-ray diffraction quality crystals are usually considered to be of fewer defects. Thus, the systematic investigation of defect generation mechanism would be helpful in growing crystals of good diffraction quality. Electron microscopy was applied to examine the surfaces of lysozyme crystals and deduce the mechanisms of crystal growth by Durbin *et al.* [82]. Although electron microscopy provides a better understanding of the fundamental processes of crystallization of macromolecules, it is difficult to observe the continuous crystal growth in solution.

Atomic force microscopy (AFM) was applied to observe the growth of protein crystals in liquid [83]. But, as with electron microscopy, AFM did not provide the information on the lattice structure of a crystal. In order to conduct a nondestructive *in situ* observation technique by which dislocations inside a crystal can be observed during growth, advanced optical microscopy [84] was employed to visualize defects in a tetragonal lysozyme crystal *in situ*. The dislocations were found to appear in bundles and were probably generated at the periphery of the relatively large inclusions inside the crystal.

3 Effects of pH, Salt Type, Salt Concentration and Temperature on the Interactions and Solubility of Hen Egg White Lysozyme

3.1 Introduction

It is known that proteins play a key role in almost every biological process. Their specific biological functions are determined by the three-dimensional arrangement of the amino acids at the active sites. Knowledge of the detailed three-dimensional structures of protein molecules is essential for the structure-based drug design and the understanding of the biological functions of protein molecules. By May 2009, around 90% of protein structures were determined using X-ray crystallography. However, the key to the success of X-ray crystallography is to obtain single crystals that are of diffraction quality. Since the mechanism of protein crystallization is not well understood, crystals are typically obtained using trial-and-error strategies. Thus, knowledge of protein solution phase behavior, in particular solubility, and the protein interactions that govern it is crucial for predicting and designing crystallization conditions to grow high quality crystals.

It is known that the solubility of protein molecules depends on a lot of parameters, such as pH, temperature, solvent, precipitant type, precipitant concentration, *etc* [42]. Measurements of the dependence of the protein solubility on these factors can be valuable for determining the phase behavior of a particular system and selecting suitable operation parameters for protein crystallization experiment. Traditionally, protein solubility could be obtained either by dissolution of protein crystals into the protein-free solution or crystallization of supersaturated solution to reach the equilibrium [122]. However, the traditional methods as described above are always costly and time-consuming. In order to

achieve rapid and efficient solubility determination, micro-column technique [6, 7], Michaelson interferometer [16, 17, 123], scintillation [124], Fourier transformed infrared spectroscopy [125] and isothermal micro-calorimetry [176] were also applied to measure the solubility of protein. Using these methods, the solubility data of hen egg white lysozyme crystals have been measured as a function of salt concentrations along with other parameters, such as pH and temperature [6-12, 17]. In addition, the solubility data of concanavalin A [13], ovalbumin [14], canavalin A [15], equine serum albumin [16] and glucose isomerase [17] were determined. As mentioned before, reliable solubility information serves as a basis for finding optimal conditions for protein crystal growth.

The analysis of crystal contacts indicated that protein crystallization depended on the strength of protein interactions [47]. Hence, understanding the principles by which protein interactions can be tuned is vital to control the nucleation and growth of protein crystals. The strength of protein interactions can be probed by measuring the second virial coefficient (B_{22}) [177]. In an early work reported by George and Wilson [39], the second virial coefficients for solutions that were of the conditions to generate high quality crystals were shown to fall in a narrow range of -1×10^{-4} to -8×10^{-4} mol mL g⁻². The solutions whose B_{22} values were larger than this range would result in no crystallization. On the other hand, the strong attractive protein interactions in solutions whose B_{22} values were more negative than -8×10^{-4} mol mL g⁻² led to the formation of amorphous precipitate. The relation between second virial coefficient and the crystallization of proteins has also been verified for various proteins [40, 145, 178, 179]. Moreover, some efforts have been made to correlate solubility with second virial coefficient [27, 41, 49, 180, 181].

The main objective of this work was to increase our understanding of the effects of pH, temperature, salt type, and salt concentration on the solubility and protein interactions of hen egg white lysozyme. In this work, the solubility and second virial coefficient values of lysozyme were measured in different practical conditions of pH, temperature, and the addition of salts. Crystallization trials in the corresponding conditions were carried out in parallel. The obtained results were applied as the basis for discussing the relation between the second virial coefficient and the phase behavior of lysozyme solution. Consistent with previous studies, both solubility and second virial coefficients for lysozyme solutions were dependent on salt type, salt concentration, pH, and temperature. The data suggested that there was clearly a systematic variation between the solubility and B_{22} over the B_{22} range studied. The solubility decreased as B_{22} became larger and more negative. In addition, salts with different anions were also found to be of different abilities to precipitate protein. The incorporation of different anions in lysozyme crystals also resulted in the formation of crystals that were of different morphologies. These results will contribute to a better understanding of the thermodynamic behavior of lysozyme molecules. Meanwhile, the experimental results will be useful in the development of the optimal operations to precipitate a target protein and obtain high-quality crystals.

3.2 Experimental Section

3.2.1 Materials

Hen egg white lysozyme (L7651, 3 × re-crystallized, lyophilized), sodium chloride ($\geq 99.5\%$), sodium nitrate ($\geq 99.5\%$), magnesium chloride (anhydrous, $\geq 98.0\%$), sodium acetate ($\geq 99.5\%$), acetic acid (HPLC grade), sodium hydroxide ($\geq 99.5\%$), toluene (HPLC grade), and paraffin oil (for IR spectroscopy) were purchased from Sigma-Aldrich,

Singapore. All reagents were used without any further treatment. The buffers were prepared using ultrafiltered, ultrapure water from a Milli-Q[®] ultra-pure water purification system to result in a 0.1 M sodium acetate solutions of pH 4.2, 4.5 and 5.2.

3.2.2 Solution Preparation

Lysozyme solutions of different concentrations were prepared by dissolving a specific amount of lysozyme powder in the sodium acetate buffer. Sodium chloride, sodium nitrate and magnesium chloride were dissolved in the same buffer to prepare the precipitant solutions of desired concentrations. The pH of both solutions measured with a pH meter (Mettler Toledo SevenMulti) was adjusted by sodium hydroxide solution and acetic acid. The solutions were passed through 0.22 μm sterile filters to remove foreign nucleation sources before use. A Shimadzu UV 2450 spectrophotometer was applied to determine the concentration of the lysozyme solution by measuring the absorbance at 280 nm using an extinction coefficient of 2.64 [52, 54, 182].

3.2.3 Solubility Measurement

Solubility was measured for lysozyme by determining the concentration of protein in the liquid in equilibrium with macroscopically observable crystals. The solutions with excess lysozyme crystals were incubated at desired temperature for several weeks until the solution equilibrium was established. The supernatant in equilibrium was then filtered through a 0.1 μm Anotop sterile filter (inorganic membrane). The concentration of lysozyme in the filtered supernatant was determined using the Shimadzu UV spectrometer and noted as the solubility. These measurements were repeated for at least three times to verify the reproducibility of the data.

3.2.4 Second Virial Coefficient Measurement

In this work, SLS experiment was performed using a ZetaPALS analyzer (Brookhaven Instrument Company, USA) to measure the second virial coefficient values of lysozyme in solutions of various precipitant concentrations, pH, and temperature. A 35 mW solid state laser at 660 nm was used as the incident light source. All the experiments were carried out at the scattering angle of 90°. For each experiment, the apparatus was first calibrated with HPLC grade toluene. Both toluene and the solvent were filtered using a 0.02 µm Anotop syringe filter (inorganic membrane). And the lysozyme-precipitant mixtures were filtered using a 0.1µm Anotop syringe filter before injecting to the equipment. In a typical static light scattering measurement, the intensity of light scattered by a protein solution in excess of the background scattered by the solvent was measured. The variable $\frac{Kc_p}{R_\theta}$ obtained from the measurement was plotted against protein concentration. Typical Debye plots of lysozyme scattering data at the scattering angle of 90° in various sodium chloride concentration of 0.1 M sodium acetate buffer at pH 4.5, 25 °C were illustrated in Figure 3.1. Since the diameter of lysozyme molecule is smaller than 1/20 of the wavelength of the incident light, Raleigh scattering equation (equation 2.10) can be applied to analyze the experimental data [42]. Both the molecular weight and second virial coefficient can be obtained through analyzing the experimental data by applying equation (2.10). Since the molecular weight of lysozyme has been characterized well, the second virial coefficient can be obtained from the slope of the Debye plot. The measurements were repeated for at least three times to ensure the reproducibility of the data.

Figure 3.1. Typical Debye plots of lysozyme static light scattering data at the scattering angle of 90° in sodium chloride solutions of 0.1 M sodium acetate buffer at pH 4.5, 25 °C.

3.2.5 Observation of Lysozyme Crystals by Optical Microscopy

The crystals obtained were observed by an Olympus BX51 polarized-light microscope with a CCD camera.

3.2.6 Raman Microscopy Measurement

A Renishaw inVia Reflex Raman spectrometer system was used in this study, equipped with a high powered Renishaw HeNe laser, at least 17 mW at 633 nm, allowing measurement of the Raman spectrum from 100 cm^{-1} to 3200 cm^{-1} , spectral resolution of 2 cm^{-1} , variable laser spots size from 1 to 300 μm . The obtained crystal samples were placed on a silicon wafer and then placed on the stage of Raman microscope for Raman spectra collection. The samples were irradiated by the Diode laser operating at 300 mW. For each

sample, 100 scans were loaded. The scan range was from 500 cm^{-1} to 2000 cm^{-1} with the resolution of 2 cm^{-1} . All the spectra were acquired and processed using the Renishaw WiRE 2.0 software.

3.3 Results and Discussion

3.3.1 Solubility

In this work, the solubility of lysozyme was examined as a function of temperature, pH, precipitant type, and precipitant concentration. The solubility data reported in this work are also listed in Table 3.1. Figure 3.2 presents the solubility dependence of hen egg white lysozyme on sodium chloride concentration and temperature at pH 4.2. The solubility was found to increase with increasing temperature and decrease with increasing sodium chloride concentration. It was also of interest to notice that the change in solubility as a function of temperature depended on the sodium chloride concentration. It was largest in 0.34 M sodium chloride solution and decreased in solutions of higher sodium chloride concentrations. These results agreed with the work reported by Cacioppo *et al.* [7]. As illustrated in Figure 3.3, the effects of pH on the solubility of lysozyme varied with sodium chloride concentration. The solubility of lysozyme in 0.34 M sodium chloride solution decreased with the increase of pH. However, in solutions of higher sodium chloride concentrations, such as 0.85 M and 1.20 M sodium chloride solutions, the solubility increased with increasing pH. In addition, sodium nitrate could also be used as a precipitant. The dependence of solubility data of hen egg white lysozyme on the precipitant type were shown in Figure 3.4. For both sodium chloride and sodium nitrate, the solubility decreased with the increasing of salt concentration. The solubility data of

hen egg white lysozyme in sodium nitrate solutions were smaller than that of hen egg white lysozyme in sodium chloride solutions.

Figure 3.2. Solubility dependence of hen egg white lysozyme on the concentration of the added sodium chloride in 0.1 M sodium acetate buffer, at pH 4.2, T=10, 15, 20 and 25 °C.

The lines are added for guide to an eye.

Figure 3.3. Solubility dependence of hen egg white lysozyme on the concentration of the added sodium chloride in 0.1 M sodium acetate buffer at $T=25\text{ }^{\circ}\text{C}$, pH 4.2, 4.5 and 5.2. The lines are added for guide to an eye.

Figure 3.4. Solubility dependence of hen egg white lysozyme on the added sodium chloride or sodium nitrate concentration in 0.1 M sodium acetate buffer at $T=25\text{ }^{\circ}\text{C}$, pH 4.5. The lines are added for guide to an eye.

3.3.2 Second Virial Coefficient

Values of second virial coefficient (B_{22}) have been measured by SLS experiments in precrystallization conditions for lysozyme in sodium chloride, sodium nitrate and magnesium chloride solutions for the pH range 4.2~5.2, temperature range 10~25 $^{\circ}\text{C}$, sodium chloride concentrations 0.34~1.50 M, sodium nitrate concentrations 0.2~1.0 M and magnesium chloride concentrations 0.2~1.5 M. Second virial coefficient of lysozyme was measured at sodium chloride concentrations ranging from 0.1 M to 1.5 M at pH 4.5, 25 $^{\circ}\text{C}$. The second virial coefficient data measured in this work are also listed in Table 3.1. As shown in Figure 3.5, B_{22} monotonically decreased with the increasing of sodium chloride concentration. At low sodium chloride concentration, electrostatic repulsion

dominated protein interaction and B_{22} was positive. Upon increasing ionic strength, electrostatic repulsion was screened and B_{22} became negative, indicating protein-protein interaction was attractive. The data were also compared with the data obtained from literature at pH 4.5 measured by various methods [41, 43, 45, 183] in Figure 3.5. Our data were found to agree well with the literature data.

Figure 3.5. Comparison of the second virial coefficient for lysozyme in sodium chloride solution at pH 4.5, 25 °C determined in this work with the data reported previously in the literature [41, 43, 45, 183].

Figure 3.6 illustrates B_{22} for lysozyme in sodium chloride solutions at pH 4.2 as a function of temperature. At 25 °C, B_{22} decreased with the increasing of the sodium chloride concentration. As the temperature was decreased, the B_{22} trends were similar to that at 25 °C, but the B_{22} values were more negative at lower temperatures, which

indicated that the strength of the attractive interactions between protein molecules increased as the decrease of temperature. The effects of pH on the B_{22} of lysozyme in sodium chloride solutions at 25 °C were shown in Figure 3.7. At a fixed sodium chloride concentration, B_{22} was found to decrease with the increasing of the solution pH. It was suggested by Dumetz *et al.* [47] that the role of pH on the protein-protein interaction could be attributed to the protonation state of charged amino acids and the α -carboxyl and α -amino groups at the surface of proteins. The pI value for lysozyme is 11.0. The rising pH from 4.2 to 5.2 lowered the net positive charge on the surface of protein, which resulted in the increase of the attractive interactions between protein molecules.

Figure 3.6. Comparison of second virial coefficient determined from static light scattering for lysozyme in various sodium chloride solutions in 0.1 M sodium acetate buffer at pH 4.2, T=10, 15, 20 and 25 °C. The lines are added for guide to an eye.

Figure 3.7. Comparison of second virial coefficient determined from static light scattering for lysozyme in various sodium chloride solutions in 0.1 M sodium acetate buffer at $T=25^{\circ}\text{C}$, pH 4.2, 4.5 and 5.2. The lines are added for guide to an eye.

It was worth noting that the effects of salt concentration on the second virial coefficient depended strongly on the specific nature of the ions. The B_{22} for lysozyme in sodium nitrate and magnesium chloride solutions were also measured and plotted in Figure 3.8. Similar to the effect of sodium chloride, there was a systematic progression of B_{22} towards more negative values as the concentration of sodium nitrate was increased. But the B_{22} values for lysozyme in sodium nitrate solutions were more negative than that in sodium chloride solutions, which was due to the different abilities of NO_3^- and Cl^- to precipitate proteins [183, 184]. On the other hand, when the concentrations of magnesium chloride and the sodium chloride were of the same, the second virial coefficient for lysozyme in magnesium chloride solution was more negative than that in sodium chloride solution.

Figure 3.8. Comparison of second virial coefficient determined from static light scattering for lysozyme in various salt solutions in 0.1 M sodium acetate buffer at pH 4.5, T=25 °C. The lines are added for guide to an eye.

One of the possible reason was that the ability of Mg^{2+} to precipitate protein was larger than that of Na^+ . The other reason was that the amount of Cl^- in magnesium chloride solutions was larger than that in sodium chloride solutions at the same salt concentration. Thus, the screening effect of magnesium chloride solution was stronger than that of sodium chloride solution. In addition, for magnesium chloride, a minimum second virial coefficient was observed at the the magnesium chloride concentration of 0.90 M. The result that the second virial coefficient increased with the increasing of magnesium chloride concentration when the magnesium chloride concentration was larger than 0.90 M could be attributed to the increase in the repulsive interactions due to the binding of the highly kosmotropic ion Mg^{2+} to the surface of lysozyme molecule. The subsequent

structuring of water around the cation could produce a repulsive barrier [29, 45]. Similar effects of magnesium cation on the second virial coefficient of lysozyme were also reported by Tessier *et al.* [45] for lysozyme in sodium bromide solutions.

3.3.3 Correlation of Solubility and Second Virial Coefficient

In order to investigate the correlation between solubility (c_e) and second virial coefficient (B_{22}), solubility values were plotted versus B_{22} in lysozyme solutions of various salt concentrations, pH and temperatures in Figure 3.9. There was clearly a systematic variation between the solubility and B_{22} over the B_{22} range studied. The data also indicated that the solubility decreased as B_{22} became larger and more negative. This is consistent with the prediction that the increase in protein attractive interactions would reduce protein solubility [49]. A theoretical explanation of this relation was given by Haas *et al.* [27]. In their work, solubility was related to second virial coefficient as

$$B_{22} = (4/M_w \rho') \left[1 - A' \left\{ (\phi_s / m)^{-(2/z')} - 1 \right\} \right] \quad (3.1)$$

$$\phi_s = \frac{c_e}{1000 \rho'} \quad (3.2)$$

where ρ' was the number density of the protein, m was the number of water molecules that can be placed in the volume of one protein molecule, c_e was the solubility in mg/mL and A' and z' were fitting parameters. A' was assumed to be dependent on the anisotropy p and the range of protein-protein interaction λ' as $A = p(\lambda'^3 - 1)$. Small A' represented strong anisotropy and short range of protein-protein interaction. While z' was related to the morphology of the crystals obtained.

Figure 3.9. Solubility versus second virial coefficient for lysozyme dissolved in various salt solutions of 0.1 M sodium acetate buffer at various pH and temperature.

The solubility was plotted versus second virial coefficient for lysozyme in sodium chloride solutions at pH 4.2, T=10, 15, 20 and 25 °C in Figure 3.9. All the data were found to fall approximately on a single curve. At a fix pH value, we can also observe that lysozyme solubility was almost the same for solutions in which lysozyme molecules experienced the same strength of protein interaction. The experimental data were also compared with the results obtained according to the Haas-Drenth-Wilson Model described by equation (3.1). The best fit was achieved with z' and A' equal to 6 and 0.049, respectively.

Figure 3.9 also presents a comparison of solubility and the second virial coefficient for lysozyme in various sodium chloride solutions at 25 °C, pH 4.2, 4.5 and 5.2. It was shown

that the data obtained at the same pH fall almost on the same curve. The comparison of the experimental data with the predictions by Haas-Drenth-Wilson model was illustrated in Figure 3.9. Clearly, the model can describe the experimental data both quantitatively and qualitatively with z' of the value of 6. The other parameter A' was 0.049 for the data obtained at pH 4.2. And it was increased to 0.053 and 0.091 to achieve the best fit for the data obtained at pH 4.5 and 5.2.

Sodium nitrate can be used as crystallizing precipitant for lysozyme. Figure 3.9 also presents the results of solubility versus second virial coefficient as a function of sodium nitrate and sodium chloride concentration at pH 4.5, 25 °C. Although the solubility range was limited for lysozyme in sodium nitrate solution, the correlation of second virial coefficient and solubility was still evident. The Haas-Drenth-Wilson theoretical model was also applied to fit the experimental data. It was found that the best fit was achieved with A' and z' of 0.0007 and 3 for lysozyme in sodium nitrate solution.

The theoretical correlation between protein solubility and second virial coefficient can also be related from the view point of the thermodynamic equilibrium between protein in a solid phase and the same protein in solution [49]. According to the work reported by Guo *et al.* [49], when a thermodynamic equilibrium was established between protein crystals and protein solution, the chemical potential of protein in crystalline phase $\mu_p(xtal)$ was equal to that of protein in solution $\mu_p(soln)$. Thus,

$$\mu_p(xtal) = \mu_p(soln) \quad (3.3)$$

At conditions of saturation,

$$\mu_p(xtal) = \mu_p^o(so\ln) + RT \ln y_p c_e \quad (3.4)$$

where $\mu_p^o(so\ln)$ is the standard chemical potential of protein in solution and y_p is the activity coefficient. Thus,

$$y_p c_e = \exp(-\Delta\mu_p / RT) \quad (3.5)$$

$$\Delta\mu_p = \mu_p^o(so\ln) - \mu_p(xtal) \quad (3.6)$$

By taking the nonideality of the solution into consideration,

$$\mu_p(so\ln) = \mu_p^o(so\ln) + RT \ln c_e + 2RTM_w B_{22}c_e \quad (3.7)$$

So the second virial coefficient and solubility was correlated by

$$B_{22} = \frac{-\Delta\mu_p}{RT} \frac{1}{2M_w c_e} - \frac{\ln c_e}{2M_w c_e} \quad (3.8)$$

where R is the gas constant, T is the temperature, M_w is the molecular weight of protein molecule, and c_e is the solubility. Thus the driving force for crystallization, $\Delta\mu_p$ of lysozyme molecules in various solution conditions can be calculated from equation (3.8) and listed in Table 3.1. In order to demonstrate the effects of pH, temperature, precipitant type and precipitant concentration on $\Delta\mu_p$, the calculated $\Delta\mu_p$ were plotted versus precipitant concentrations at different temperature and pH in Figure 3.10. As illustrated in Figure 3.10, $\Delta\mu_p$ was found to decrease with increasing temperature and increase with increasing sodium chloride concentration at constant pH. The trend of the dependence of $\Delta\mu_p$ on salt concentration was the same at different pH values. But the effects of pH on $\Delta\mu_p$ differed at different sodium chloride concentration. From Figure 3.10, it was also observed that $\Delta\mu_p$ in 0.34 M sodium chloride solution increased at higher pH but decreased with increasing pH in 0.85 M and 1.20 M sodium chloride solutions. In addition,

the $\Delta\mu_p$ for lysozyme in sodium nitrate solution was larger than that in sodium chloride solution over the range of salt concentrations studied, as summarized in Figure 3.10.

Figure 3.10. Dependence of $\Delta\mu_p$ on the salt concentration in 0.1 M sodium acetate buffer at various pH and temperature. The lines are added for guide to an eye.

Table 3.1. Second virial coefficient (B_{22}), solubility (c_e) and $\Delta\mu_p$ for lysozyme under various conditions.

Precipitant	pH	T(°C)	$B_{22} \times 10^4$ (mol.mL/g ²)	c_e (mg/mL)	$\Delta\mu_p$ (kcal/mol)
0.34 M NaCl	4.2	25	-1.61	33.52	2.10
0.51 M NaCl	4.2	25	-2.74	12.88	2.64
0.68 M NaCl	4.2	25	-3.65	6.25	3.05
0.85 M NaCl	4.2	25	-4.46	3.60	3.36
1.20 M NaCl	4.2	25	-6.3	1.89	3.73

Table 3.1 (*Continued*)

Precipitant	pH	T(°C)	$B_{22} \times 10^4$ (mol.mL/g ²)	c_e (mg/mL)	$\Delta\mu_p$ (kcal/mol)
0.34 M NaCl	4.2	20	-2.21	16.10	2.46
0.51 M NaCl	4.2	20	-3.81	5.49	3.07
0.68 M NaCl	4.2	20	-4.63	2.84	3.44
0.85 M NaCl	4.2	20	-5.72	2.08	3.62
1.20 M NaCl	4.2	20	-7.45	1.14	3.96
0.34 M NaCl	4.2	15	-3.46	14.14	2.52
0.51 M NaCl	4.2	15	-4.69	3.41	3.28
0.68 M NaCl	4.2	15	-5.37	2.08	3.55
0.85 M NaCl	4.2	15	-6.47	1.53	3.73
1.20 M NaCl	4.2	15	-8.20	0.76	4.12
0.34 M NaCl	4.2	10	-3.89	9.47	2.68
0.51 M NaCl	4.2	10	-5.94	3.22	3.26
0.68 M NaCl	4.2	10	-6.82	1.89	3.55
0.85 M NaCl	4.2	10	-7.69	1.33	3.74
1.20 M NaCl	4.2	10	-9.82	0.57	4.21
0.34 M NaCl	4.5	25	-1.74	32.39	2.13
0.51 M NaCl	4.5	25	-3.16	11.36	2.71
0.68 M NaCl	4.5	25	-4.01	6.06	3.07
0.85 M NaCl	4.5	25	-4.85	4.17	3.28
1.20 M NaCl	4.5	25	-6.54	2.08	3.68
0.34 M NaCl	5.2	25	-4.62	25.76	2.24

Table 3.1 (*Continued*)

Precipitant	pH	T(°C)	$B_{22} \times 10^4$ (mol.mL/g ²)	c_e (mg/mL)	$\Delta\mu_p$ (kcal/mol)
0.51 M NaCl	5.2	25	-6.34	12.50	2.73
0.68 M NaCl	5.2	25	-7.13	7.95	2.96
0.85 M NaCl	5.2	25	-8.69	5.49	3.16
1.20 M NaCl	5.2	25	-9.90	3.22	3.45
0.2 M NaNO ₃	4.5	25	-5.51	2.46	2.50
0.5 M NaNO ₃	4.5	25	-9.17	1.52	3.87
0.75M NaNO ₃	4.5	25	-12.4	1.14	4.04
1.0 M NaNO ₃	4.5	25	-14.9	0.95	4.15

3.3.4 Crystallization Results

It was reported by George and Wilson [39] that the values of the B_{22} for the solutions in crystallization conditions were within a narrow range of $-1 \times 10^{-4} \sim -8 \times 10^{-4}$ mol mL g⁻². The solutions whose B_{22} values were out of this range would either result in no crystallization or disordered aggregates. In order to test the application of the concept of “crystallization slot” to the crystallization of lysozyme in sodium chloride and sodium nitrate solution at pH 4.5 and 25 °C, crystallization trials were performed and were summarized in Figure 3.11. At low salt concentration, electrostatic repulsion dominated protein interactions and B_{22} was positive. No crystals were found to grow from the solution. Upon the increasing of the salt concentration, electrostatic repulsion was screened and B_{22} became negative, indicating protein-protein interaction was attractive. Good quality symmetrical crystals were obtained, as shown in the region named “crystallization slot” in Figure 3.11.

In this slot, the second virial coefficients were in the range of $-1 \times 10^{-4} \sim -8 \times 10^{-4} \text{ mol mL g}^{-2}$. Upon the increasing of the salt concentration, the attractive interactions were strengthened. Therefore, the second virial coefficient was more negative than $-8 \times 10^{-4} \text{ mol mL g}^{-2}$. Thus the protein attractive interaction was too strong for the lysozyme molecules to adjust their orientations to form the crystals that are of suitable quality for the X-ray diffraction [53]. These crystals of bad X-ray diffraction quality were typically of the sea-urchin morphology as shown in the third slot of Figure 3.11. The formation of the sea-urchin morphology crystal could be attributed to the strong attractive protein-protein interaction. Thus, the “crystallization slot” proposed by George and Wilson applied to the experiments in this work.

Figure 3.11. Crystallization results of lysozyme in sodium chloride and sodium nitrate solutions in 0.1 M sodium acetate buffer at pH 4.5, 25 °C.

It addition, it was interesting to note that the crystals obtained in the presence of sodium chloride and sodium nitrate were of different morphologies, as shown in Figure 3.12. Like many proteins, lysozyme can crystallize in various crystal structures depending on the crystallizing agent and temperature [185]. Tetragonal ($P4_32_12$) lysozyme crystals were obtained in sodium chloride solution when temperature was not larger than 25 °C. Monoclinic ($P2_1$) lysozyme crystals were obtained in the sodium nitrate solution at 25 °C. The crystals obtained in the conditions of this work were consistent with the previous research work. The different morphologies may be attributed to the incorporation of anions in the crystals, which could be further supported by the Raman spectra of the tetragonal lysozyme chloride crystals and monoclinic lysozyme nitrate crystals, as compared in Figure 3.13. The incorporation of NO_3^- anion in the crystals was confirmed by the presence of the peak at around 1050 cm^{-1} in the spectrum of monoclinic lysozyme nitrate crystals [186]. But these various crystal structures that were caused by the incorporating of different anions were not polymorphs strictly [185].

Figure 3.12. Optical microscopy images of (a) tetragonal and (b) monoclinic lysozyme crystals.

Figure 3.13. Raman spectra of tetragonal and monoclinic lysozyme crystals.

3.4 Conclusions

In this work, the solubility and second virial coefficient of lysozyme were examined as a function of temperature, pH, precipitant type, and precipitant concentration. The solubility of lysozyme was found to increase with increasing temperature and decrease with increasing sodium chloride or sodium nitrate concentration. The effects of pH on the solubility of lysozyme varied with sodium chloride concentration. In low sodium chloride concentration solution, lysozyme solubility decreased at higher pH while it increased with increasing pH in solutions of high sodium chloride concentrations. The solubility data for lysozyme in sodium nitrate solutions were smaller than that of lysozyme in sodium chloride solutions. The second virial coefficient values for the solutions of corresponding conditions were measured as well. It was noted that there was a strong correlation between the solubility and the second virial coefficient. The solubility decreased as the second

virial coefficient became larger and more negative. The driving force for crystallization $\Delta\mu_p$ decreased with increasing temperature and increase with increasing sodium chloride concentration at constant pH. The hypothesis on the relation between the second virial coefficient and the solubility were verified by fitting the experimental data with the prediction of Haas-Drenth-Wilson model. The ‘crystallization slot’ proposed by George and Wilson [39] was verified in this work as well. Crystals of high quality were obtained when the value of the second virial coefficient was in the ‘crystallization slot’. No crystallization or crystals that were not of good diffraction quality were obtained when the second virial coefficients were out of this range. Moreover, the adsorption of different anions onto the surface of protein molecules was found to result in crystals of different morphologies.

4 Cloud-point Temperature and Liquid-liquid Phase Separation in Supersaturated Lysozyme and Bovine Serum Albumin Solutions

4.1 Introduction

Preparation of single crystals that are of suitable diffraction quality is indispensable for the determination of the three-dimensional structure of protein molecules by the X-ray crystallography, but it is difficult. This difficulty could be partly due to the fact that protein solutions exhibit various kinds of phase transition phenomena, including crystallization and liquid-liquid phase separation [18]. It has been both experimentally suggested [18-24, 52] and theoretically predicted [25, 26, 28, 55] that the liquid-liquid phase separation affected protein crystallization. A detailed understanding of the liquid-liquid phase separation, especially the study of the location of the liquid-liquid co-existence boundary and the strategies to shift it, provides a systematic approach to control protein crystallization.

A liquid-liquid co-existence curve can be determined by measuring the cloud-point temperature. It was first reported for lysozyme-salt solution by Ishimoto and Tanaka in 1977 [35]. Subsequently, the effects of pH, salt identity, and salt concentrations on the cloud-point temperature for lysozyme in electrolyte solutions were investigated systematically to gain an insight into protein interaction and its correlation with crystallization [29-31, 36, 37, 187]. In recent years, such kind of studies has been extended to other proteins, such as bovine serum albumin (BSA) [38], urate oxidase [32] and bovine pancreatic trypsin inhibitor [33, 34]. Additionally, the liquid-liquid co-existence curves for γ -crystallin, hemoglobin and their natural variants [21, 120, 121, 130, 132, 135, 136] were

measured to provide a sound reference in the search of the treatment of cataract and sickle cell anemia.

From a theoretical point of view, the phase behavior of protein molecules depends on the interaction between them. For a sufficiently attractive interaction, the phase diagram of protein solutions exhibits a meta-stable liquid-liquid co-existence curve that subtends the solid-liquid co-existence curve. Since the second virial coefficient is a dilute solution thermodynamic parameter that reflects the extent of protein interaction [188], it is of interest to investigate the correlation between the cloud-point temperature and the second virial coefficient.

The similarity between the phase diagrams of colloidal solutions and globular protein solutions enabled the modeling of the phase diagram of protein solution using the theories developed for colloidal solutions. The first model was based on the equation of state proposed by van der Waals [137]. This equation of state was further developed to be more general by characterizing the model system with their molecular interactions and kinetic energy. Another general approach was to apply a statistical-mechanical perturbation theory, in which the protein solution was assumed to be composed of hard spheres with repulsive interactions only. The attractive interaction was assumed to be the perturbation to the repulsive forces of the hard sphere system [38, 138-140]. The Gibbs free energy model [36, 55] which described the water-mediated interactions between protein molecules in phenomenological level was also used to describe the properties of the observed liquid-liquid phase separation. In addition, many Monte Carlo simulations have been made to predict the liquid-liquid co-existence curve [142, 143]. However, these models only

describe the phase diagram of protein solutions qualitatively. The curves generated by modeling are not as wide as that determined by experimental measurement.

In this work, the liquid-liquid co-existence curves in supersaturated lysozyme solutions with 2%, 3%, 4% and 5% w/v NaCl at pH 4.2 and BSA solutions with 60, 70, 80 mg/mL polyethylene glycol 1500 (PEG 1500) at pH 5.0 were determined by measuring the cloud-point temperature. The second virial coefficients for lysozyme and BSA in solutions of various precipitant concentrations and temperatures were also measured to investigate the correlation between the cloud-point temperature and the second virial coefficient. In addition, random phase approximation (RPA) with square-well potential [29] was applied to correlate the cloud-point temperature data obtained from literature. These literature data were also compared with the results obtained by the Gibbs free energy model [189] and Monte Carlo simulation [142]. It was found that these models could not describe the liquid-liquid co-existence curves of both lysozyme and γ II-crystallin system quantitatively. It is of interest to find that the liquid-liquid co-existence curve could be better described through modifying the RPA by assuming the energy of the interaction between protein molecules to be temperature dependent. The modified RPA was found to describe the liquid-liquid co-existence curve quantitatively. Moreover, the energy of protein interactions was calculated from the experimentally measured second virial coefficient. The temperature dependence of the energy of protein interactions was observed, which could serve as a support for the above assumption.

4.2 Experimental Section

4.2.1 Materials

Hen egg white lysozyme (L7651, 3 × re-crystallized, lyophilized), bovine serum albumin (BSA) (05470, crystallized, lyophilized, $\geq 98.0\%$), sodium chloride ($\geq 99.5\%$), sodium acetate ($\geq 99.5\%$), acetic acid (HPLC grade), sodium hydroxide ($\geq 99.5\%$), toluene (HPLC grade), and paraffin oil (for IR spectroscopy) were purchased from Sigma-Aldrich, Singapore. Polyethylene glycol of molecular weight 1,500 (g/mol) (PEG 1500) (biochemika ultra) was obtained from Chemicals Testing & Calibration Laboratory, Singapore. All reagents were used without any further treatment. The buffers were prepared with ultrafiltered, ultrapure water from a Milli-Q® ultra-pure water purification system.

4.2.2 Solution Preparations

Both protein solution and salt solution were prepared at twice of the final concentration separately. The benefit of this approach was that the salt was introduced into the protein solution upon the mixing for measurement. This could avoid the crystallization of protein solution [187]. Proper amounts of lysozyme and NaCl were dissolved in 0.1 M sodium acetate buffer at pH 4.2 separately to obtain stock solutions for cloud-point temperature measurement. The pH of each solution measured with a pH meter (Mettler Toledo Seven Multi) was readjusted to the desired pH with concentrated hydroxide solution and acetic acid. The solutions were subsequently filtered through 0.22 μm sterile filters to remove the foreign particles before use. BSA and PEG 1500 solutions of different concentrations were prepared in the same way. The only difference was that 0.1 M sodium acetate buffer at pH 5.0 was used as the buffer. The concentrations of the pure protein solutions were

determined by measuring the absorbance at 280 nm using a Shimadzu UV spectrophotometer. The extinction coefficients for lysozyme and BSA are 2.64 and 0.667, respectively [190]. The concentrations of NaCl and PEG 1500 solutions were calculated based on the mass of NaCl and PEG 1500 and the volume of the solutions.

4.2.3 Sample Preparation

The separately prepared lysozyme and sodium chloride solutions were allowed to equilibrate at a temperature of 45 °C prior to mixing. At the same temperature, the stock solutions of lysozyme and NaCl were mixed manually in equal volume in a 1.5 mL Eppendorf tube. It was followed by depositing a droplet of the pre-mixed solution (approximately 0.5 μ L) into a micro quartz cell (15 mm diameter $\times$ 2 mm depth) of a Linkam heating/cooling microscope stage (Linkam THMS 600) and sealed with paraffin oil to stop evaporation. The samples of the BSA-PEG 1500-buffer system were prepared in the same way as described above.

4.2.4 Cloud-point Temperature and Second Virial Coefficient Measurement

An optical microscope (Olympus BX51) with a heating/cooling stage (Linkam THMS600) was used to measure the cloud-point temperature. The stage is designed to observe the thermo-behavior of a sample under the microscope. Temperature can be controlled by liquid nitrogen within 0.1 °C in the range from -192 to 600 °C, which allows a rapid equilibrium. The paraffin oil sealed sample was placed in this stage and observed visually by the optical microscope with an Olympus Analysis CVIII cooled CCD camera. The sample was equilibrated at 45 °C for 30 min. The solution was clear at this temperature, as shown in Figure 4.1 (a). Then the temperature was reduced at a rate of 1

°C/min for lysozyme-NaCl-buffer system and 10 °C/min for BSA-PEG 1500-buffer system until the onset of LLPS. As shown in Figure 4.1 (b), the onset of LLPS was associated with a dramatic clouding of the solution due to the different refraction index of the protein-rich and protein-poor domain. Further lowering of the temperature brought about no detectable changes in the solution droplet. The solutions became clear when the temperature of the sample was increased. The temperature at which the droplet became cloudy firstly was noted as the cloud-point temperature T_{cloud} for the convenience of the comparison. Each measurement was repeated at least three times to assure the reproducibility. The method required a very small amount of protein and short time to determine the cloud-point temperature. The second virial coefficients for lysozyme and BSA in solutions of various precipitant concentrations and temperatures were determined in the method as described in section 3.2.4.

Figure 4.1. Illustration of a typical result of cloud-point temperature determination for lysozyme solution containing 128.4 mg/mL lysozyme, 3% w/v sodium chloride in 0.1 M sodium acetate buffer at pH 4.2: (a) 45 °C, the solution was clear, (b) 12.2 °C, the solution became cloudy.

4.3 Results and Discussion

4.3.1 Cloud-point Temperatures

In this work, the cloud-point temperatures (T_{cloud}) of lysozyme and BSA were measured as a function of precipitant and protein concentrations. Figure 4.2 presents T_{cloud} dependence of lysozyme on lysozyme and sodium chloride concentration at pH 4.2. At 2% w/v sodium chloride solution, T_{cloud} was found to increase monotonically with the concentration of lysozyme. As the sodium chloride concentration was increased, the trend was similar to that at 2% w/v NaCl, but T_{cloud} increased as the concentration of NaCl increased from 2% w/v to 5% w/v. The cloud-point temperature data obtained in this work were compared with the data reported in literature, as shown in Figure 4.3. It was noted

that at the lysozyme concentration of 87.4 mg/mL, our experimental data on T_{cloud} were quite consistent with the data measured using DLS method [29]. In addition, the dependence of T_{cloud} on lysozyme concentration at 3%, 4% and 5% w/v NaCl at pH 4.2 and 0.1 M sodium acetate buffer was compared with the liquid-liquid co-existence curves obtained by Petsev *et al.* [191] and Muschol *et al.* [37] in Figure 4.3. The data obtained in this work were in good agreement with literature values with some deviations. These differences could be considered to be not significant by taking the differences in pH into account. Therefore, the cloud-point temperature measurement method in this work was reasonable.

Figure 4.2. Cloud-point temperature determination for lysozyme as a function of lysozyme concentration at four NaCl concentrations. The lines are added for guide to an eye.

Figure 4.3. Comparison of the cloud-point temperature data measured in this work with the literature data. (a) Comparison of the dependence of the cloud-point temperature ($C(\text{lysozyme})=87.4 \text{ mg/mL}$, $\text{pH}=4.2$, $[\text{NaOAc}]=0.1 \text{ M}$) on the ionic strength with the data reported by Grigsby *et al.* [29] ($C(\text{lysozyme})=87 \text{ mg/mL}$, $\text{pH}=4.0$, $20 \text{ mM Tris Buffer}$). (b) Comparison of the dependence of the cloud-point temperature ($C(\text{NaCl})=4\% \text{ w/v}$, $\text{pH}=4.2$, $[\text{NaOAc}]=0.1 \text{ M}$) on lysozyme concentration with the work reported by Petsev *et al.* [191] ($C(\text{NaCl})=4\% \text{ w/v}$, $\text{pH}=4.5$, $[\text{NaOAc}]=0.05 \text{ M}$). (c) Comparison of the dependence of cloud-point temperature ($C(\text{NaCl})=3\% \text{ w/v}$ and $C(\text{NaCl})=5\% \text{ w/v}$, $\text{pH}=4.2$, $[\text{NaOAc}]=0.1 \text{ M}$) on lysozyme concentration with the work reported by Muschol *et al.* [37] ($C(\text{NaCl})=3\% \text{ w/v}$ and $C(\text{NaCl})=5\% \text{ w/v}$, $\text{pH}=4.5$, $[\text{NaOAc}]=0.1 \text{ M}$).

The cloud-point temperatures for BSA in the presence of PEG 1500 in 0.1 M sodium acetate buffer at pH 5.0 were measured as well. According to Figure 4.4, at a fix PEG 1500 concentration, T_{cloud} of BSA increased at higher BSA concentration. At a fixed BSA concentration, T_{cloud} increased monotonically with the concentration of PEG 1500.

Figure 4.4. Cloud-point temperature determination for BSA as a function of BSA concentration at three PEG 1500 concentrations. The lines are added for guide to an eye.

4.3.2 Second Virial Coefficient

Values of second virial coefficient (B_{22}) for lysozyme were plotted versus sodium chloride concentrations at pH 4.2 and temperatures ranging from 10 to 25 °C in Figure 3.6. At 25 °C, B_{22} decreased with the increasing of the sodium chloride concentration. As the temperature was decreased, the B_{22} trends were similar to that at 25 °C, but the B_{22} values

were more negative at lower temperatures. The effects of temperature on B_{22} of BSA in PEG 1500 solutions were illustrated in Figure 4.5. At fixed temperature, B_{22} was found to decrease with the increasing of the concentration of PEG 1500. At fixed PEG 1500 concentration, B_{22} decreased with temperature.

Figure 4.5. Comparison of second virial coefficients of BSA determined from static light scattering in PEG 1500 solutions of 0.1 M sodium acetate buffer at pH 5.0, T=15, 20 and 25 °C. The lines are added for guide to an eye.

4.3.3 Correlation of Cloud-point Temperature and Second Virial Coefficient

In order to investigate the correlation between the cloud-point temperature (T_{cloud}) and the second virial coefficient (B_{22}), the dimensionless cloud-point temperature (T_r) was plotted versus dimensionless second virial coefficient (B'_{22}) for lysozyme solutions at various sodium chloride concentrations and temperature at pH 4.2 in Figure 4.6 and for

BSA solutions at various PEG 1500 concentrations and temperatures at pH 5.0 in Figure 4.7. According to Berger *et al.* [50] the dimensionless cloud temperature (T_r) and dimensionless second virial coefficient (B'_{22}) could be calculated as

$$T_r = \frac{T_{cloud} - T}{T_{cloud}} \quad (4.1)$$

$$B'_{22} = \frac{B_{22}}{B_{22}^{HS}} \quad (4.2)$$

Figure 4.6. Dimensionless cloud-point temperature T_r versus dimensionless second virial coefficient B'_{22} of lysozyme in different NaCl solutions. The lysozyme concentration was 87.4 mg/ml in all cases. The buffer was 0.1 M NaOAc buffer at pH 4.2.

where T was the temperature at which the second virial coefficient was measured. An example was taken to examine the correlation between T_r and B'_{22} at the lysozyme concentration of 87.4 mg/mL and BSA concentration of 117.8 mg/mL. A systematic variation between T_r and B'_{22} were illustrated in both Figure 4.6 and Figure 4.7. The data also indicated that T_r increased as B'_{22} became larger and more negative. This observation was in a good agreement with the prediction that the liquid-liquid phase separation was driven by the net protein attractive interaction. The cloud-point temperature increased with the increase of the strength of the attractive interaction [29-31].

Figure 4.7. Dimensionless cloud-point temperature T_r versus dimensionless second virial coefficient B'_{22} of bovine serum albumin in different PEG 1500 solutions. The BSA concentration was 117.8 mg/ml in all cases. The buffer was 0.1 M NaOAc buffer at pH 5.0.

4.3.4 Predictions of the Liquid-liquid Co-existence Curve

It is known that the RPA has been applied to study the phase separation in a protein solution induced by addition of neutral polymer and/or simple electrolyte [29, 30, 138, 139, 192, 193]. The RPA was a perturbation theory that approximated the direct correlation function $c(r)$ by [138]

$$c(r) = c_{ref}(r) - \frac{u_1(r)}{kT} \quad (4.3)$$

where k was Boltzmann constant, T was temperature, $c_{ref}(r)$ was the direct correlation function for the reference system, and $u_1(r)$ was the perturbation part of the total energy $u(r)$. Taking the Fourier transforms of both sides of equation (4.3) gave

$$C = C_{ref} - \frac{U_1}{kT} \quad (4.4)$$

in which

$$C_{ref} = 4\pi \int c_{ref}(r) r^2 dr \quad (4.5)$$

$$U_1 = 4\pi \int u_1(r) r^2 dr \quad (4.6)$$

According to Hansen and McDonald [194], the isothermal compressibility could be used to relate the pressure of the system P to C . Hence, the equation for pressure P was [29]

$$\frac{P}{\rho' kT} = \frac{P_{ref}}{\rho' kT} + \frac{\rho' U_1(r)}{2kT} \quad (4.7)$$

where P_{ref} was the pressure of the reference system, ρ' was the protein number density, and $U_1(r)$ was the perturbation energy per unit density. By applying the Gibbs-Duhem equation, the chemical potential of the system μ was [29, 138]

$$\frac{\mu - \mu'}{kT} = \ln \rho' + \frac{\mu_{ref}}{kT} + \frac{\rho' U_1(r)}{kT} \quad (4.8)$$

In the RPA, pressure P and chemical potential μ were used to describe the liquid-liquid phase equilibrium. μ' was the reference chemical potential that was the same in both liquid phases. The compositions of the two liquid phases α and β could be calculated by solving the two equation as bellows.

$$\mu^\alpha = \mu^\beta \quad (4.9)$$

$$P^\alpha = P^\beta \quad (4.10)$$

The reference system was assumed to be hard-sphere fluid. According to Carnahan-Starling equation [195]

$$\frac{P_{ref}}{\rho' kT} = \frac{1 + \eta' + \eta'^2 - \eta'^3}{(1 - \eta')^3} \quad (4.11)$$

$$\frac{\mu_{ref}}{\rho' kT} = \frac{\eta'(8 - 9\eta' + 3\eta'^2)}{(1 - \eta')^3} \quad (4.12)$$

with η' being the protein volume fraction of the liquid phase.

The perturbation energy per unit density was given by

$$U_1 = 4\pi \int W(r) r^2 dr \quad (4.13)$$

In this work, a square-well potential of mean force $W(r)$ was applied to describe the interactions between two protein molecules as

$$W(r) = \begin{cases} \infty, & r < \sigma \\ -\varepsilon, & \sigma < r < \lambda'\sigma \\ 0, & \lambda'\sigma < r \end{cases} \quad (4.14)$$

where r was the center-to-center distance between two protein molecules, σ was the protein diameter, λ' was the attractive range, and ε was the depth of the square-well which described the interactions between protein molecules. According to ten Wolde and

Frenkel [25], a meta-stable liquid-liquid co-existence region existed when λ' was not larger than 1.25. Thus, the range of λ' should fall in the range of

$$1 < \lambda' \leq 1.25 \quad (4.15)$$

In this work, we chose the typically values of 1.15 and 1.25 as examples to calculate the reduced liquid-liquid co-existence curve (T/T_c vs η'/η'_c).

Figure 4.8. Comparison of the liquid-liquid co-existence curves calculated using different methods with the cloud-point temperature data reported in literature.

The reduced liquid-liquid co-existence curves (T/T_c vs η'/η'_c) generated by RPA, the Gibbs free energy model proposed by Haas and Drenth [189] and the Gibbs ensemble Monte Carlo method [142] with different values of λ' were plotted in Figure 4.8. It was found that these three models predicted similar behavior of the liquid-liquid co-existence curve, but the widths of the curves were different using the same λ' . The reduced liquid-liquid co-existence curves generated by RPA were of the same shape and width regardless

the value of λ' . For the curves that were calculated using the other two methods, the widths of the curves that were calculated using the same method varied with value of λ' , but the difference was not significant.

4.3.5 Linking Models to Experimental Data

The liquid-liquid co-existence curves obtained from the described three methods along with the cloud-point temperatures for lysozyme [35-37] and γ II-crystallin [133] as reported in the literature were illustrated in Figure 4.8. It was shown that the data for the liquid-liquid co-existence curves of the lysozyme and γ II-crystallin solutions at different solution conditions were on almost the same curve. The liquid-liquid co-existence curves that were calculated from RPA, the Gibbs free energy model proposed by Haas and Drenth, and the Gibbs ensemble Monte Carlo method were not of the same width as that of the experimental data. One of the possible reasons for the failure of these models to describe the liquid-liquid co-existence curve quantitatively may be that all these models were isotropic model, in which the spatial variation of the interactions between protein molecules was neglected. It was further supported by the better description of the phase diagram of protein solutions generated by the models that characterized the interactions of protein molecules with an anisotropic model [196]. In addition, the interactions between protein molecules in these models were assumed to be temperature independent. However, the fact that the protein interactions between protein molecules were mediated by the surrounding water molecules suggested the possibility to take the temperature dependence of the protein interaction into account [143].

In this study, the square-well potential of the RPA was modified to be temperature dependent to characterize the liquid-liquid co-existence curve of protein solutions. According to Lomakin *et al.* [143], the energy of the interaction between protein molecules was considered to be temperature dependent, as described in equation (4.16),

$$\varepsilon(T) = kT_c \hat{\varepsilon}_c [1 + \kappa'(T - T_c)/T_c] \quad (4.16)$$

in which, k was the Boltzmann constant, κ' was a constant, and $\hat{\varepsilon} = \varepsilon / kT_c$. Therefore,

$$\frac{kT}{\varepsilon(T)} = \frac{kT}{kT_c \hat{\varepsilon}_c [1 + \kappa'(T - T_c)/T_c]} \quad (4.17)$$

The scaled temperature T/T_c may be calculated as

$$\frac{T}{T_c} = \frac{(1 - \kappa')(\hat{\varepsilon}_c(T_c)/\hat{\varepsilon}(T))}{1 - \kappa'(\hat{\varepsilon}_c(T_c)/\hat{\varepsilon}(T))} \quad (4.18)$$

in which $\hat{\varepsilon}_c(T_c)$ and $\hat{\varepsilon}(T)$ can be obtained from the simulated data by RPA.

In this work, we found that a good fit between the experimental data and the modified RPA was achieved when $\kappa' = -6$. As shown in Figure 4.8, it characterized the experimental data more quantitatively than other models. Similar situation was found in the work reported by Lomakin *et al.* for γ -crystallin [143]. Indeed further investigations are needed to gain an insight into the dependence of κ' on temperature.

In addition, when square-well interaction potential is applied, the energy of protein interaction can be calculated from equation (4.19)

$$B_{22} = \frac{2\pi\sigma^3}{3} [\lambda^3 - e^{\varepsilon/kT} (\lambda^3 - 1)] \quad (4.19)$$

The energy of protein interactions for solutions at different temperature and precipitant concentrations was calculated from the experimentally measured second virial coefficient and summarized in Figure 4.9 and Figure 4.10. The energy of protein interactions ε/k was found to increase with the increasing of precipitant concentration and increase with the decreasing of temperature. A possible reason for the temperature dependence of the protein interaction was that the interactions between protein molecules were mediated by the surrounding water molecules [143]. Indeed, the dependence of the strength of the protein interactions on the precipitant concentration and temperature requires further investigation.

Figure 4.9. Dependence of ε/k for lysozyme calculated from the measured second virial coefficients on temperature at different sodium chloride solutions. The buffer is 0.1 M NaOAc buffer at pH 4.2.

Figure 4.10. Dependence of ε/k for bovine serum albumin calculated from the measured second virial coefficients on temperature at different PEG 1500 concentrations. The buffer is 0.1 M NaOAc buffer at pH 5.0.

4.4 Conclusions

The liquid-liquid co-existence curves for aqueous solutions of lysozyme and BSA in low density portion have been determined experimentally. As expected, the basic shapes of the curves were the same, however, the location of them varied with the concentrations of NaCl (for lysozyme) and PEG 1500 (for BSA) added. It was found that the location of the liquid-liquid co-existence curve shifted up in temperature as the concentrations of NaCl and PEG 1500 increased. We also presented the second virial coefficients for lysozyme and BSA in solutions of various precipitant concentrations and temperature. In order to correlate the cloud-point temperature and the second virial coefficient, the dimensionless cloud-point temperature was plotted versus the dimensionless second virial

coefficient. There was a systematic variation between the dimensionless cloud-point temperature and the dimensionless second virial coefficient over the protein second virial coefficient range studied. The data indicated that the cloud-point temperature was increased as the second virial coefficient became larger and more negative, which was consistent with the prediction that the higher the cloud-point temperature, the stronger the attractive protein-protein interactions. In this work, the random phase approximation (RPA) with a square-well potential was also applied to describe the cloud-point temperature data reported in literature by modifying the interactions between protein molecules to be temperature dependent. Although the physical reasonability of this simple model requires further investigation, it has been shown to provide a reasonably accurate description of the liquid-liquid co-existence curve for lysozyme and BSA solutions. The temperature dependence of the protein-protein interaction as calculated from the measured second virial coefficient provided a further support for this modification. The experimental data and modified model reported in this work has raised a number of interesting theoretical and experimental issues regarding the interactions between protein molecules and its correlation with the phase diagram of protein solution. The further investigation regarding these issues may provide a better insight into the phase separation of protein solution.

5 The Effects of Solution Parameters on the Nucleation and Growth of Lysozyme Crystals

5.1 Introduction

In Chapter 3 and 4, the phase diagram for lysozyme molecules in different practical conditions has been characterized, and a correlation was established to relate the phase behavior of protein molecules to protein interactions. Although protein crystals of good diffraction quality were found to grow from solutions that were of moderate protein interactions, an understanding of the influences of solution parameters on the nucleation and growth kinetics of protein crystals is also important in the control of the size and quality of protein crystals.

It is known that protein crystallization is affected by a lot of factors, such as protein concentration, precipitant type and concentration, temperature, pH, *etc.* Thus a detailed understanding of the protein crystallization process and the effects of these parameters on protein nucleation and growth kinetics is important for the successful production of protein crystals. Since nucleation is the necessary first step in protein crystallization process, many efforts have focused on this area. For example, the nucleation rates of lysozyme crystals were indirectly estimated from the induction time measured using light scattering [197] and the enthalpy of crystallization determined using microcalorimetry [176]. In other studies, nucleation rates were determined directly by measuring the number of lysozyme crystals produced per unit volume per unit time [19, 20, 52-54]. It was suggested that nucleation rates increased with supersaturation, but there was a broken dependence of nucleation rates on supersaturation when liquid-liquid phase separation

occurred. Subsequently, considerable research has been directed to studying the effect of the liquid-liquid phase separation and phase transition on nucleation in supersaturated protein solutions [18, 19, 21-26, 28, 35, 36, 52, 55]. The simplest mechanism proposed to explain the enhancement of nucleation by the protein-rich phase was the higher protein concentration in it [53]. The presence of a metastable liquid-liquid critical point was also proposed to change the nucleation pathway dramatically and enhance the nucleation by density fluctuation [25, 26]. Numerical analysis also indicated that the protein-rich phase could reduce the nucleation energy by wetting the crystalline nucleus [55]. However, the exact mechanism of the effects of liquid-liquid phase separation on protein nucleation is still not clear. For crystal growth, the influence of supersaturation on the lysozyme crystal growth rate and habit has also been investigated [79-81]. The growth rates were obtained by measuring the changes in the dimensions of a particular face of protein crystals using optical microscopy. The defects in lysozyme crystal have been investigated using electron microscopy [82], atomic force microscopy [83] and advanced optical microscopy [84].

As mentioned above, different aspects of protein crystallization have been investigated, but most of these studies just focused on one or two sides of the whole crystallization process. When a full picture of protein crystallization was sought, the experimental conditions and methods were normally inconsistent with each other. Furthermore, the interactions of many parameters are not well understood and highly nonlinear in their effects on other parameters, hence, there is still not a generally accepted theory to describe protein crystallization. A further understanding on the mechanism of protein nucleation and crystal growth is necessary for tuning protein crystallization in a controlled manner.

This work presented a full-picture investigation of protein crystallization, namely a detailed study on the phase behavior and crystallization mechanism of protein molecules including nucleation kinetics, crystal growth kinetics, crystal morphologies, solubility, second virial coefficient and cloud-point temperatures using consistent experimental conditions and methodologies. Raman microscopy was also attempted to probe the conformation changes and the related protein-protein interactions of lysozyme molecules in aqueous solutions of different lysozyme concentrations. It was found that at a fixed sodium chloride concentration, the nucleation rate increased with lysozyme concentration. However, there was a broken dependence of the nucleation rate on the protein concentration at 4% and 5% w/v NaCl, which was beyond the prediction of classical nucleation theory. This broken dependence was attributed to the liquid-liquid phase separation. On the basis of the optical microscopy and Raman spectroscopy analysis, it was suggested that nucleation was arrested in the protein-rich phase. Crystals nucleated from the protein-lean phase. Instead of causing the nucleation of new crystals, the protein-rich phase dissolved into the protein-lean phase to restore the local concentration. The influence of the overall protein concentration on the local concentration in the region in between the growing nucleus and the protein-rich phase was also analyzed to explain the increase of nucleation rate with the overall protein concentration when crystals nucleated from a liquid-liquid phase separation solution. In addition, the differences in the shapes and morphologies of lysozyme crystals could be explained by analyzing the growth kinetics of crystals and the strength of protein interactions. The systematic measurement and analysis in this work were expected to provide a further understanding of protein crystallization process.

5.2 Experimental Section

5.2.1 Materials

Hen egg white lysozyme (L7651, 3 × re-crystallized, lyophilized), sodium chloride ($\geq 99.5\%$), sodium acetate ($\geq 99.5\%$), acetic acid (HPLC grade), sodium hydroxide ($\geq 99.5\%$), toluene (HPLC grade), and paraffin oil (for IR spectroscopy) were purchased from Sigma-Aldrich, Singapore. All reagents were used without any further treatment. The buffer was prepared with ultrafiltered, ultrapure water from a Milli-Q[®] ultra-pure water purification system to result in a 0.1 M NaOAc/HOAc solution of pH 4.2.

5.2.2 Solution Preparation

Lysozyme solutions of different concentrations were prepared by dissolving a specific amount of lysozyme powder in the sodium acetate buffer. Sodium chloride was dissolved in the same buffer to prepare the precipitant solutions of desired concentrations. The pH of both solutions measured with a pH meter (Mettler Toledo SevenMulti) was adjusted by sodium hydroxide solution and acetic acid. The solutions were passed through 0.22 μm sterile filters to remove foreign nucleation sources before use. A Shimadzu UV 2450 spectrophotometer was applied to determine the concentration of the lysozyme solution by measuring the absorbance at 280 nm using an extinction coefficient of 2.64 [52].

5.2.3 Cloud-point Temperature Measurement

An optical microscope (Olympus BX 51) with a heating/cooling stage (Linkam THMS 600) was used to measure the cloud-point temperature. The stage was designed to observe the thermo-behavior of a sample under the microscope. Temperature could be controlled by liquid nitrogen with an accuracy of 0.1 $^{\circ}\text{C}$ in the range from -192 to 600 $^{\circ}\text{C}$, which

allowed a rapid equilibrium. The paraffin oil sealed sample was placed in this stage and observed visually by the optical microscope with CCD camera. The sample was equilibrated at 45 °C for 30 min. The solution was uniform at this temperature, as illustrated in Figure 5.1 (a). Then the temperature was lowered at a rate of 1 °C/min until the onset of liquid-liquid phase separation. Figure 5.1 (b) showed that the onset of liquid-liquid phase separation was associated with a separation of protein-rich droplets from the clear solution, which was in agreement with the phenomena observed by Tanaka *et al.* [18]. The temperature at the onset of the liquid-liquid phase separation was defined as the cloud-point temperature T_{cloud} . Each measurement was repeated for at least three times to assure the reproducibility.

Figure 5.1. Illustration of a typical result of cloud-point temperature determination for lysozyme solution (lysozyme, 128.4 mg/mL, NaCl 3% w/v) in 0.1 M sodium acetate buffer, pH 4.2: (a) 45 °C, the solution was uniform, (b) 12.2 °C, the protein-rich droplets separated from the solution.

5.2.4 Solubility and Second Virial Coefficient Measurement

The solubility of lysozyme in solutions of the conditions investigated in this work was measured in the method as described by Bhamidi *et al.* [54]. The solutions with excess lysozyme crystals were incubated at 10 °C for several weeks until the solution equilibrium was established. The supernatant in equilibrium was then filtered through a 0.1 μm Anotop sterile filter (inorganic membrane). The concentration of lysozyme in the filtered supernatant was determined using the Shimadzu UV spectrometer and noted as the solubility. These measurements were repeated for at least three times to verify the reproducibility of the data. The concentrations of the sodium chloride concentrations studied were 2%, 3%, 4% and 5% w/v. The second virial coefficients for lysozyme in

solutions of various sodium chloride concentrations in 0.1 M sodium acetate buffer at pH 4.2, 10 °C were determined using the method as described in section 3.2.4.

5.2.5 Observation of the Crystallization Processes by Optical Microscopy

Observation of the crystallization process was made by an Olympus BX51 polarized-light microscope with a CCD camera. A series of time-lapsed images of the crystallization process without the occurrences of the liquid-liquid phase separation, the liquid-liquid phase separation, and the subsequent crystallization process were taken using the CCD camera attached.

5.2.6 Nucleation Rates Measurement

In this work, micro-batch nucleation method, which allowed an exact initial supersaturation, was applied to measure the nucleation rates. Experiments were conducted in a micro quartz cell (15mm diameter × 2 mm depth) accommodated in a Linkam heating/cooling microscope stage (Linkam THMS 600). For the controllability of the operation conditions, the temperature was set at 10 °C. Crystallization solutions of different initial concentrations were prepared by manually mixing the lysozyme and NaCl solutions in equal volume in a 1.5 mL Eppendorf tube so that the crystallization batch contained the required concentrations of lysozyme and NaCl. A droplet of 0.5 µL premixed solution was deposited on the quartz cell and then sealed with paraffin oil. An optical microscope (Olympus BX 51) with a CCD camera was used in combination with a computer to record the images of the studied droplets at certain time intervals. At a later time, the number of crystals was counted. Due to the unique phenomena of protein crystallization compared with the crystallization of small molecules and the fact that

experiments were conducted in a micro vessel without agitation and the number density of crystals was not high, secondary nucleation, crystal agglomeration and breakage could be neglected. Therefore, it was reasonable to calculate the apparent nucleation rates based on the assumption that the stable nucleus grew into respective detectable crystals. As shown in Figure 5.2, the change in the number density of the crystals with the time from initial measurement gave the initial rate of nucleation. Each measurement was repeated for at least five times to ensure the reliability of the experimental data. Although the number of crystals in a droplet was a random variable, a good repeatability was found in the initial nucleation rate for the experiments at the same crystallization condition. This good repeatability in the initial nucleation rate indicated the applicability of the technique. The initial nucleation rates reported in this work were the average of five measurements.

Figure 5.2. A typical example of the measurement of nucleation rate. The number of crystals in one droplet of 0.5 μL was plotted against the time from initial measurement. The slope of the straight line was calculated to be the nucleation rate J .

Figure 5.3. Illustration for the measurement of the growth rates for (110) and (101) faces of tetragonal lysozyme crystals.

5.2.7 Crystal Growth Rate Measurement

Crystal growth rates for both the (101) and (110) faces of the tetragonal lysozyme crystals were measured using the same experimental set up as that applied to measure the cloud-point temperature and the nucleation rate. The method to determine the crystal growth rates was illustrated in Figure 5.3. A series of time-lapsed images of the growing crystals were taken by the CCD camera attached to the microscope. Then the size of the crystals was analyzed using the Analysis Pro software. The lengths of L_A and M were plotted against time from the observation, as shown in Figure 5.3. And the growth rates for the faces (110) and (101), G_{110} and G_{101} , could be calculated as described by Durbin and Feher [79].

$$G_{110} = 0.5 \frac{dL_A}{dt} \quad (5.1)$$

$$G_{101} = 0.45 \frac{dM}{dt} + 0.30 G_{110} \quad (5.2)$$

5.2.8 Raman Microscopy Measurement

A Renishaw inVia Reflex Raman spectrometer system was used in this study, equipped with a high powered Renishaw Diode laser, at least 300 mW at 785 nm, allowing measurement of the Raman spectrum from 100 cm⁻¹ to 3200 cm⁻¹, spectral resolution of 2 cm⁻¹, variable laser spots size from 1 to 300 μm. The solution samples were prepared in the way as described before to achieve the concentrations of 50.6 mg/mL, 88.4 mg/mL, 126.3 mg/mL and 177.1 mg/mL, respectively. A 10 μL lysozyme droplet was deposited on a silicon wafer and then placed on the stage of Raman microscope for Raman spectra collection. The samples were irradiated by the Diode laser operating at 300 mW. For the solutions samples, 100 scans were loaded. The scan range was from 550 cm⁻¹ to 2000 cm⁻¹ with the resolution of 2 cm⁻¹.

A study of lysozyme secondary structure was carried out in the amide I band spectral region (1600 cm⁻¹ ~ 1700 cm⁻¹), amide III band spectral region (1200 cm⁻¹ ~ 1300 cm⁻¹) and tyrosine doublet band spectral region (800 cm⁻¹ ~ 900 cm⁻¹). Amide I and III band regions provide the information on the conformation changes on the backbone of lysozyme. The side chain conformation change was revealed by the tyrosine doublet ratio. All the spectra were acquired and processed using the Renishaw WiRE 2.0 software. The tyrosine doublet ratio was calculated in the way that the intensity of the peak at higher

Raman shift ($850 \text{ cm}^{-1} \sim 860 \text{ cm}^{-1}$) was divided by that of the peak at lower Raman shift ($830 \text{ cm}^{-1} \sim 840 \text{ cm}^{-1}$).

5.3 Theoretical Models

5.3.1 Analysis of Measured Nucleation Rates by Classical Nucleation Theory (CNT)

In classical nucleation theory, the nucleus was assumed to be of spherical shape and the nucleation rate J was given by [165, 198]

$$J = \frac{2v\sqrt{kT\gamma}}{h} N_1 \exp\left(\frac{-\Delta G_a}{kT}\right) \exp\left(\frac{-16\pi\gamma^3 v^2}{k^3 T^3 (\ln S)^2}\right) \quad (5.3)$$

in which v is the molecular volume, k is the Boltzmann's constant, T is the absolute temperature, γ is the surface energy per unit area of the nuclei, h is the Planck's constant, N_1 is the number density of monomers in the solution, ΔG_a is the energy barrier to diffusion from bulk solution to the cluster and S is the supersaturation ratio. The supersaturation ratio can be calculated via

$$\ln S = \ln(c/c_e) \quad (5.4)$$

where c is the solute concentration and c_e is the solubility. The interfacial energy can be directly related to the strength and the range of protein interaction through the correlation as developed for a square well interaction potential as [198, 199]

$$\frac{\gamma\sigma^2}{kT} = \gamma' = \frac{\sqrt{3}}{[1 + b(\lambda'-1)]^2} \frac{\varepsilon}{kT} \quad (5.5)$$

where b is a constant smaller than 1, σ is the diameter of protein molecule, λ' is the range of protein interaction and ε is the strength of protein interaction. For lysozyme molecule, the diameter σ was known to be 3.4 nm [182]. The strength of protein-protein interaction can be calculated from the equation (4.19).

5.3.2 Analysis of the Local Concentration Profile in the Region in between the Growing Crystals and Protein-rich Droplets

According to the work reported by Heijna *et al.* [24], the retracting protein-rich phase was regarded as an iso-concentration line in a diffusion system. Due to the thin geometry of the system, convection can be neglected. The mass transport process in the system follows Fick's second law. In addition, the liquid-liquid phase separation and the subsequent crystallization process as observed by optical microscopy suggested that the growing nucleus could be assumed to be a cylinder. The cross section of the cylindrical nucleus in a semi-spherical droplet was illustrated in Figure 5.4. By considering a cylinder in which diffusion is everywhere radial, the concentration is a function of the diffusion radius r' and time t only. The concentration diffusion equation can be written as [200]

$$\frac{\partial C'(r',t)}{\partial t} = D \frac{\partial^2 C'(r',t)}{\partial r'^2} + \frac{D}{r'} \frac{\partial C'(r',t)}{\partial r'} \quad (5.6)$$

The solution of this equation depends on the initial condition and boundary conditions. In this study, the boundary conditions before the disappearing of all the protein-rich droplets were applied. In this case, at $r' = r_{LL}$, the concentration was constant and in equilibrium with the protein-rich phase. Therefore,

$$C'(r_{LL}) = C_L \quad (5.7)$$

in which C_L was the concentration of the protein-lean phase. At $r' = r_n$, there were protein molecules incorporating into the nucleus. Thus

$$\frac{\partial C'}{\partial r'} + \frac{\beta C_n}{D} (C_{cr} - C') = 0 \quad (5.8)$$

where C_n was the protein concentration in the nucleus and C_{cr} was the crossover concentration which was around 7.7 times of the solubility [201]. Since both the boundary

conditions were constant before the disappearing of the protein-rich phase, the system could be assumed to be in a quasi-steady state. Hence, equation (5.8) becomes

$$0 = D \frac{\partial^2 C'(r')}{\partial r'^2} + \frac{D}{r'} \frac{\partial C'(r')}{\partial r'} \quad (5.9)$$

In combination of these boundary conditions, the concentration distribution in the “buffer zone” [23], the region in between the nucleus and the protein-rich droplet is

$$C'(r') = \frac{C_L \left(1 + \frac{\beta C_n}{D} r_n \ln \left(\frac{r'}{r_n} \right) \right) + \frac{\beta C_n}{D} r_n C_{tr} \ln \left(\frac{r_{LL}}{r'} \right)}{\frac{\beta C_n}{D} r_n \ln \left(\frac{r_{LL}}{r'} \right) + 1} \quad (5.10)$$

According to the well known thermodynamic equilibrium of protein-rich and protein-lean phase, the difference in the overall protein concentration did not affect the concentrations of protein-rich and protein-lean phases but resulted in the differences in the relative amounts of protein-rich to protein-lean phase. This difference in the relative amounts of protein-rich to protein-lean phase resulted in the difference in the positions of the liquid-liquid phase separation boundary r_{LL} . The smaller the overall protein concentration, the larger the boundary r_{LL} is. The schematic concentration profiles in the “buffer zone” with different protein concentrations were plotted in Figure 5.5. It was shown that the local concentrations in the “buffer zone” increased with the increase of the overall protein concentration.

Figure 5.4. Schematic illustration of a crystal nucleating and growing from a liquid-liquid phase separation solution.

Figure 5.5. Local concentration profile calculated at different initial overall concentrations assuming the initial moment of the system to be in a quasi-steady state.

5.4 Results and Discussion

5.4.1 Cloud-point Temperatures for Lysozyme Solutions

The liquid-liquid co-existence boundaries for lysozyme at four different NaCl concentrations in 0.1 M sodium acetate buffer at pH 4.2 were determined by measuring the cloud-point temperatures of the solutions at different conditions. The cloud-point temperatures of lysozyme at four different NaCl concentrations were systematically measured and presented in Figure 5.6. The results indicated that the location of the liquid-liquid co-existence boundary depended on the concentration of NaCl. The liquid-liquid co-existence boundaries shifted up in temperature as the concentration of NaCl varied from 2% to 5% w/v. At a constant NaCl concentration, the cloud-point temperature increased with the lysozyme concentration in the scope of the research work. As shown in Figure 5.6, it was also found that the cloud-point temperatures for solutions with 2% and 3% w/v NaCl investigated in this work were smaller than 10 °C. While the cloud-point temperatures for solutions with lysozyme concentrations larger than around 46.5 and 30 mg/mL for 4% and 5% w/v NaCl were higher than 10 °C.

5.4.2 Protein Crystallization without the Liquid-liquid Phase Separation

For solutions whose cloud-point temperatures were smaller than 10 °C, lysozyme crystals nucleated and grew from the solution without the occurrence of the liquid-liquid phase separation as illustrated in Figure 5.7. As shown in Figure 5.7(a), a lysozyme crystal nucleated from the solution containing 23 mg/mL and 3% w/v NaCl at 10 °C after 15 minutes from the initial observation. Then the crystal continued to grow as shown from Figure 5.7 (b) to Figure 5.7 (d).

Figure 5.6. Cloud-point temperature as a function of lysozyme concentration at four sodium chloride solutions in 0.1 M sodium acetate buffer, at pH 4.2, 10 $^{\circ}\text{C}$.

Figure 5.7. Images of lysozyme crystals growing from solution consisting of 23 mg/mL lysozyme, 3% w/v NaCl, in 0.1 M sodium acetate buffer at pH 4.2, 10 °C. The pictures were taken at (a) 15 min, (b) 25 min, (c) 35 min and (d) 45 min after initial observation.

In this work, the number of crystals formed in sample solution was counted at certain time intervals and plotted versus time from the initial measurement. As shown in Figure 5.2, the nucleation rate was determined from the slope of the plot. In fact, the supersaturation of protein changed after a period of nucleation. However, the supersaturation changed insignificantly in the initial stage of crystallization [54]. Therefore, the supersaturation of protein could be assumed to be constant during the initial stage of the system, which made the investigation of the initial nucleation kinetics to be reasonable. The nucleation rate data for lysozyme crystals obtained at different NaCl concentrations at pH 4.2, 10 °C were summarized as a function of protein concentration in Figure 5.8. The data reported were the average of five measurements with 95% confidence limits based on the scatter of the data. As expected, the nucleation rates increased with the protein concentration at 2% and 3% w/v NaCl. However, at 4% and 5% w/v NaCl, the nucleation rate data split into two regimes. The dependence broke at the protein concentrations of around 48.8 mg/mL and 33.8 mg/mL for 4% and 5% w/v NaCl, respectively. In both

regimes, the nucleation rates increased monotonically with the overall protein concentration.

Figure 5.8. Dependence of nucleation rate J on the lysozyme concentration at four sodium chloride concentrations in 0.1 M sodium acetate buffer at pH 4.2, 10 °C. Dash lines are fitted with the classical nucleation theory.

The dependence of nucleation rate on lysozyme concentration at fixed sodium chloride concentration obtained experimentally in this work was modeled using equation (5.3) as described by classical nucleation theory and illustrated in Figure 5.8. It was observed from

Figure 5.8 that the classical nucleation theory predicted the dependence of the nucleation rate on lysozyme concentration at fixed temperature and sodium chloride concentrations well.

The effects of sodium chloride concentration on the nucleation rates at fixed lysozyme concentration were investigated by linking classical nucleation theory with the protein-protein interaction as obtained from the measured second virial coefficient. The strength of protein-protein interaction ε/kT for lysozyme in sodium chloride solutions at 10 °C can be calculated from the measured second virial coefficient using equation (4.19) and listed in Table 5.1. The different values of ε/kT were calculated from the same second virial coefficient with different values of λ' . As shown in Table 5.1, the ε/kT increased with sodium chloride concentration regardless of the range of interaction. By substituting the ε/kT data and the solubility data as summarized in Table 5.1 into equation (5.3) and equation (5.5), $\gamma^3/(k^3T^3(\ln S)^2)$ was found to decrease with the increase of the concentration of sodium chloride as illustrated in Figure 5.9. Therefore $\exp(-16\pi\gamma^3v^2/(k^3T^3(\ln S)^2))$ and $\sqrt{\gamma}$ increased with sodium chloride concentration. Since ΔG_a was the energy barrier to diffusion from bulk solution to the cluster, it was related to ε . The increase in ε would decrease ΔG_a and result in the increase of $\exp(-\Delta G_a/kT)$ [198]. Hence it could be concluded that the nucleation rate would increase with sodium chloride concentration at fixed lysozyme concentration and temperature, which was in a good agreement with the experimental data determined in this work.

Figure 5.9. Dependence of $\gamma^3 / (k^3 T^3 (\ln S)^2)$ on the sodium chloride concentration at fixed lysozyme concentrations.

Table 5.1. Second virial coefficient (B_{22}), solubility (c_e) and protein Interaction ($\frac{\epsilon}{kT_{283.15}}$) in various sodium chloride solutions at pH 4.2, T=10°C.

C(NaCl) % w/v	c_e (mg/mL)	$B_{22} \times 10^4$ (mol.mL/g ²)	$\frac{\epsilon}{kT_{283.15}}$	$\frac{\epsilon}{kT_{283.15}}$	$\frac{\epsilon}{kT_{283.15}}$
2	9.47	-3.89	1.81 ^a	1.54 ^b	1.34 ^c
3	3.22	-5.94	2.06 ^a	1.77 ^b	1.55 ^c
4	1.89	-6.82	2.15 ^a	1.86 ^b	1.63 ^c
5	1.33	-7.69	2.23 ^a	1.93 ^b	1.71 ^c

^a The range of protein-protein interaction $\lambda' = 1.15$.

^b The range of protein-protein interaction $\lambda' = 1.20$.

^c The range of protein-protein interaction $\lambda' = 1.25$.

The morphologies of lysozyme crystals were also studied under different concentrations of NaCl and lysozyme. Lysozyme crystals grown from sodium chloride solutions were found to be of either tetragonal or orthorhombic structures when crystallization occurred in the solid-liquid co-existence region [185]. However, orthorhombic lysozyme crystals were only obtained when the temperature was higher than 25 °C [51]. Since all the experiments in this work were carried at 10 °C, only tetragonal lysozyme crystals were obtained when the crystallization occurred in the solid-liquid co-existence region. But the shapes of the tetragonal lysozyme crystals obtained from the solutions without the occurrence of liquid-liquid phase separation were different, as shown in zone II and zone III in Figure 5.10. This may be due to the difference in the growth rate of the (110) and (101) faces of the lysozyme crystals under the same conditions as illustrated in Figure 5.11. For the crystals of the shape as shown in zone II, the growth rate of the (101) face was larger than that of (110) face, while in zone III, the growth rate of the (101) face was smaller than that of (110) face. The trend that the (101) face growth rates were larger than the (110) face growth rates at low supersaturations and smaller at high supersaturations may be attributed to the differences in the growth units of these two faces and the differences in the concentrations of these respective growth units at different supersaturations. The analysis of the tetragonal lysozyme crystals using periodic bond chain (PBC) suggested that the crystallizing unit for the (101) face was tetramer and that for the (110) face was octamer [174]. At low supersaturation, the number of the tetramer was larger than that of the octamer, which led to the (101) face growing faster than that of the (110) face, and the number of octamer was larger than that of the tetramer at high supersaturation, thus the (110) face growing faster than that of the (101) face [202].

Figure 5.10. Crystallization window of lysozyme crystals obtained in various crystallization conditions in 0.1 M sodium acetate buffer at pH 4.2, 10 °C.

Figure 5.11. Growth rates of the (110) and (101) faces of the tetragonal lysozyme crystals in 0.1 M sodium acetate buffer, at pH 4.2, 10 °C in various crystallization conditions. The lines are drawn for the guide to an eye.

Figure 5.12. Images of lysozyme crystals growing from a solution consisting of 55.7 mg/mL lysozyme, 4% w/v NaCl in 0.1 M sodium acetate buffer at pH 4.2, 10 °C. These pictures were taken at (c) 10 min, (d) 20 min, (e) 30 min and (f) 40 min after initial observation. (a) The solution was uniform right after mixing at 45 °C. (b) Protein-rich droplets separated from the solution at 10 °C.

5.4.3 Liquid-liquid Phase Separation and the Subsequent Crystallization Process

In Figure 5.12, a series of optical micrographs showed a typical process of the liquid-liquid phase separation and the subsequent crystallization. As shown in Figure 5.12 (a), protein-rich droplets separated from the solution after cooling down the solution containing 51.5 mg/mL lysozyme and 4% w/v NaCl at 10 °C. Figure 5.12 (b) to Figure 5.12 (f) showed that the crystals nucleated and grew from the protein-lean phase. Around the crystal, a “buffer zone” [23] was developed without the presence of protein-rich droplets. As time progressed, new crystals nucleated and grew from the “buffer zone”. During the process, the protein-rich phase dissolved into the “buffer zone” as the nucleating and growing crystals depleted their surroundings. The polarizer of the microscope was also crossed to support that the protein-rich phase did not cause nucleation in this work.

Experimental observation in Figure 5.12 indicated that the protein-rich droplets did not cause the nucleation of new crystals but dissolved into the protein-lean phase to locally restore the equilibrium concentration of this phase in contact with the dense phase [18, 22, 24, 203-205]. The suppression of the nucleation in the protein-rich phase was attributed to the high viscosity of the lysozyme-rich phase [192] or the gel formation of lysozyme at higher concentrations [37, 38, 203, 206]. In order to further probe the suppression of lysozyme crystal nucleation in the protein-rich phase, Raman microscope was applied to investigate the conformation changes and the related protein-protein interactions of lysozyme molecules in aqueous solutions of different lysozyme concentrations. The Raman spectra of lysozyme solutions of different concentrations were collected and compared in Figure 5.13. Spectra comparison was made primarily in amide I band region

($1600\text{ cm}^{-1} \sim 1700\text{ cm}^{-1}$), amide III band region ($1200\text{ cm}^{-1} \sim 1300\text{ cm}^{-1}$) and tyrosine doublet band region ($800\text{ cm}^{-1} \sim 900\text{ cm}^{-1}$). Although the peak position shifted to a little extent in some spectra, there were still in good agreements in the amide I band regions among these four spectra, suggesting that lysozyme molecules retained their backbone conformations in solutions of different concentrations. The conclusion was further supported by the good agreement among these spectra in the amide III band regions. The tyrosine doublet intensity ratio was applied to study the side chain conformation change of lysozyme molecules. After baseline correction, the absolute intensities of both peaks of the tyrosine doublet were recorded and doublet ratio of all samples were calculated and listed in Table 5.2. The tyrosine doublet ratio was found to increase with the lysozyme concentration. By extending this tendency to the highly concentrated solution, such as protein-rich phase, it may be concluded that the tyrosine doublet ratio of the protein-rich phase was larger than that of the protein-lean phase. Referring to the work reported by Wu [206], it was assumed that when a tyrosine residue was buried inside the molecules, it only participated in intramolecular hydrogen bonding and served as a hydrogen bond donor only. If it was exposed at the outer surface of a protein molecule, the tyrosine residue could be involved in the intermolecular hydrogen bonding as both donor and acceptor at the same time. In the later case, protein interactions were apparently stronger than the former case. Therefore, an increase in the tyrosine doublet ratio might suggest an increase in the interactions between protein molecules. As for the protein solutions of different concentrations, the tyrosine doublet ratio increased with the concentration. This may suggest that the protein interactions in the protein-rich phase were larger than that in the protein-lean phase.

Figure 5.13. FT-Raman spectra of (a) 50.6 mg/mL, (b) 88.4 mg/mL, (c) 126.3 mg/mL and (d) 177.1 mg/mL lysozyme solutions.

Table 5.2. Tyrosine doublet ratio for lysozyme samples.

sample	doublet position (cm ⁻¹)	doublet ratio
50.6 mg/mL lysozyme solution	856, 836	0.50
88.4 mg/mL lysozyme solution	855, 839	1.19
126.3 mg/mL lysozyme solution	855, 835	1.40
177.1 mg/mL lysozyme solution	856, 838	1.72

The differences in the protein interactions together with the self-assembly of protein molecules may be used to explain the suppression of the nucleation in the protein-rich phase. It is known that protein nucleation is a concentration-dependent self-association process, and operations over a larger supersaturation range indicate that there is a correspondingly large range in the association state of the solute [207]. It should be viewed as a transition along two order parameters; density and structure [208]. Only the structure that lowers the total free energy of the system may become a nucleus. The monomer assembly requires that each monomer must be positioned as it comes into the attachment site. Molecules that are not oriented correctly cannot be adjusted easily *in situ* but must be dissociated and re-associated [196, 207]. This argument explains why protein crystals favor some space-groups over the others [196]. In order to nucleate from the solution, all bonds must be broken at first and then realigned to attach to the correct position, which may suggest a conformation change. The differences of the tyrosine doublet ratio may suggest differences in the packing structures of the protein molecules in the protein-lean and protein-rich phase. The relatively smaller of the doublet ratio in the protein-lean phase made it more easily for protein molecules to adjust their conformation to rearrange their

association structure, which lead to the nucleation of the protein crystal. On the contrary, in protein-rich phase, the stronger interaction may take a certain time for the protein molecules to change their conformation and orientation. This may serve as a reason of the suppression of the nucleation in the protein-rich phase.

Figure 5.14. Schematic illustration of the liquid-liquid co-existence curve for lysozyme-NaCl solution.

As indicated in Figure 5.6 and Figure 5.8, the nucleation rate data split into two regimes at around 48.8 mg/mL and 33.8 mg/mL for 4% and 5% w/v NaCl, respectively. In the solid-liquid co-existence regime, the nucleation rate increased with the lysozyme concentration. Extremely high nucleation rates were observed close to the liquid-liquid co-existence boundary. Further increase in the lysozyme concentration would cause the occurrence of liquid-liquid phase separation and the decrease in the nucleation rate. This discontinuity could be attributed to the occurrence of liquid-liquid phase separation. According to the low density portion of the liquid-liquid co-existence curve as shown in

Figure 5.6, liquid-liquid phase separation occurred at these two conditions. The concentrations of the protein-lean phase were 46.5 mg/mL and 30.0 mg/mL for 4% w/v and 5% w/v NaCl solutions at 10 °C, respectively. Thus, the concentration of the protein-lean phase was lower than the overall concentration, which would decrease the nucleation rate. In addition, the averaging of the nucleate rate in the protein-lean phase over the total solution volume would further decrease the apparent nucleation rate.

Additionally, the nucleation rates in the liquid-liquid co-existence region were expected to be constant regardless of the overall concentration, which was in contradictory with the experimental data in this work. According to the schematic phase diagram of protein solutions, as shown in Figure 5.14, the solution separated into protein-rich and protein-lean phases in the liquid-liquid co-existence region. At the same temperature, the differences of in the overall protein concentration did not affect the concentrations of either the protein-rich or protein-lean phase [23]. In the conditions of this work, the nucleation of lysozyme crystals was found to be arrested in the protein-rich phase. Lysozyme crystals were found to nucleate from the protein-lean phase. If the concentration of protein lean-phase was applied in equation (5.3), the nucleation rates in the liquid-liquid co-existence region should be constant regardless of the overall concentration. However, it is well recognized that the differences in the overall protein concentrations resulted in the difference of the relative amount of protein-rich phase to protein-lean phase [23]. The mass transport in the system resulted in the differences of the local concentrations in the region in between the growing nucleus and the protein-rich phase when the overall protein concentrations were different. The local concentration in the region in between the growing crystals and protein-rich droplets was found to increase with the overall protein

concentration, as shown in Figure 5.5. Therefore, the increase of nucleation rate of lysozyme crystals from the liquid-liquid phase separation solution with the overall lysozyme concentration could be considered to be reasonable by taking the local concentration in this region into account.

Moreover, it was reported by George and Wilson [39] that the values of the B_{22} for the solutions in crystallization conditions lied within a narrow range of $-1 \times 10^{-4} \sim -8 \times 10^{-4} \text{ mol} \cdot \text{mL} \cdot \text{g}^{-2}$. The solutions whose B_{22} values that were out of this range would either result in no crystallization or disordered aggregates. As summarized in Table 5.1, the second virial coefficients for the solutions in this work were in the range of optimal crystallization, which was further supported by the crystals obtained in zone II and III as shown in Figure 5.10. The tetragonal crystals of rough surfaces as shown in zone IV, Figure 5.10 were formed via the process in Figure 5.12. Due to the buffering effects of the protein-rich droplet, the concentration of protein-lean phase was kept constant before the disappearing of all the droplets [23]. The resulted high driving force may result in the surfaces of the crystals got rounded and convex [209], as shown in Figure 5.12. The further growth of the crystal would result in the decrease in the protein concentration after the disappearing of the protein-rich droplets. The rounded crystals became faceted to form the tetragonal crystals of rough surface as the decrease of the protein concentration, as shown in zone IV Figure 5.10. The crystals of sea-urchin morphology, which were not of the suitable quality for X-ray diffraction [53] may be attributed to the strong attractive protein-protein interaction. This could be supported by investigating the sea-urchin morphologies of the crystals obtained at solutions of 7% w/v sodium chloride. For

solutions of 7% w/v sodium chloride, the second virial coefficient ($-9.82 \times 10^{-4} \text{ mol} \cdot \text{mL} \cdot \text{g}^{-2}$) was out of the range of optimal crystallization. The protein-protein attractive interaction was too strong for the lysozyme molecules to adjust their orientations to form the crystals that are of suitable quality for X-ray diffraction.

5.5 Conclusions

In summary, the nucleation kinetics of lysozyme crystals at various lysozyme and NaCl concentrations has been investigated using an initial nucleation rate measurement method. In consistence with the classical nucleation theory, the nucleation rate increased with the protein concentration when crystallization happened in the solid-liquid co-existence region. Depending on the solution conditions, the crystals obtained were of different shapes and morphologies. The different shapes of the crystals obtained in the solid-liquid co-existence region could be attributed to the differences in the growth rates of the (110) and (101) faces of tetragonal crystals. The differences in the growth rates of the (110) and (101) faces could be explained by the PBC analysis that the growth units for the (101) face was tetramer and that for the (110) face was octamer. At low supersaturation, the number of the tetramer was larger than that of the octamer, which led to the (101) face growing faster than that of the (110) face, and the number of octamer was larger than that of the tetramer at high supersaturation, thus the (110) face grew faster than that of the (101) face. The broken dependence of nucleation rate on protein concentration around the liquid-liquid co-existence boundary and the appearance of the sea-urchin morphology crystal may be due to the occurrence of the liquid-liquid phase separation. Some further support came from the scenario of the liquid-liquid phase separation and the subsequent crystallization process as observed by optical microscopy. In the experimental conditions of this work, it

was found that the protein-rich phase did not cause the nucleation of new crystals but dissolved into the “buffer zone” to restore the local concentrations. In addition, the suppression of the nucleation in the protein-rich phase was probed in terms of protein-protein interactions, which was determined by Raman microscopy and the self-assembly of the protein molecules. The relatively stronger protein-protein interaction in the protein-rich phase could make it difficult to adjust the conformation and orientation of the protein molecules to rearrange their association structure to achieve the nucleation of the protein crystals. Furthermore, the concentration distribution in the “buffer zone” was derived to explain the increase of the nucleation rate in the liquid-liquid phase separation solution. The thorough investigation not only provided a further understanding of protein crystallization from protein-protein interaction to final crystal morphology but also a guideline for the future study on protein crystallization and the tuning of protein crystallization in a controlled way.

Part of this work has been published on Crystal Growth & Design (Vol 10, pp. 548-558, 2010).

6 Influence of the Roughness, Topography and Physicochemical Properties of Chemically Modified Surfaces on the Heterogeneous Nucleation of Protein Crystals

6.1 Introduction

Based on the work as discussed in the previous chapters, it can be concluded that the crystals that are of good diffraction qualities are always obtained from the solutions that are of low supersaturation. The excess supersaturation may result in the growth rates that are too fast for the molecules to adjust their orientations and thus cause defects in protein crystals. Therefore, the approaches that can induce protein nucleation at low supersaturation are advantages in the control of the quality of protein crystals. Since heterogeneous nucleation can decrease the free energy barrier for nucleation and thus grow protein crystals in solutions that are of relatively lower supersaturation compared with homogeneous nucleation, many academic efforts have been made on this issue.

As summarized in Table 2.3, Chapter 2, various materials have been tested as heterogeneous nucleation surfaces for protein crystallization in the past few decades. The first attempt to control protein nucleation on heterogeneous surfaces was carried out by McPherson and Shlichta through examining the applicability of mineral surfaces as potential heterogeneous nucleants on various proteins [56, 59]. After this initiative, mineral substrates [57, 58], poly-L-lysine coated surfaces [61, 62] and lipid layers [63-65] were applied to control nucleation as epitaxial nucleants. These epitaxial nucleants were found to influence the nucleation of protein crystals by modifying crystal habit, promoting unit cell properties or involving a lattice match [17]. Other than these epitaxial nucleants, chemically modified mica [66, 67], silanized polystyrene flat-bottom wells [68] and

polymeric film surfaces [69] were also used as heterogeneous nucleants to induce the crystallization of protein crystals at low starting protein concentrations. Their effectiveness was attributed to the promotion of molecular collisions and clustering with the due symmetry for the formation of the crystal nuclei caused by the non-specific attractive and local interactions between protein and the foreign surfaces. Besides, the presence of surface microstructures was revealed to be another factor that can influence the heterogeneous nucleation of protein crystals. In previous studies, porous silicon [70-72], porous glass surfaces [169] and porous poly(vinylidene fluoride) membranes [170] were successfully applied to induce the heterogeneous nucleation of protein crystals by the presence of a specific surface structure.

These results as summarized above suggest that different surfaces may affect protein heterogeneous nucleation through different mechanisms [57, 170] and none of the surfaces studied has been shown to be a universal heterogeneous nucleant. In order to design the heterogeneous nucleation protocol rationally, a clear understanding of the heterogeneous nucleation mechanism is necessary.

The aim of this work is to provide a further understanding of the mechanism of heterogeneous nucleation by investigating the influences of surface structure and chemical properties of foreign nucleation surfaces on the heterogeneous nucleation of protein crystal. A number of lysozyme micro-batch crystallization experiments were carried out on a series of chemically modified glass slide surfaces. Compared with the nucleation on bare glass slides (BG), protein nucleation was promoted on the surfaces that were treated using (3-aminopropyl) triethoxysilane (APTES), poly(2-hydroxyethyl methacrylate) (P2HEMA)

and poly-L-glutamic acid (PLG). Whereas poly(methyl methacrylate) (PMMA) and poly(4-vinyl pyridine) (P4VP) coated surfaces were found to suppress the heterogeneous nucleation of protein crystals. The measurements of the surface roughness and topography and the contact angles between nucleating solutions and the surfaces suggested that these surfaces affected the heterogeneous nucleation in different ways. Classical nucleation theory (CNT) has also been applied to derive a mathematical correlation to relate the free energy barrier for heterogeneous nucleation to surface roughness and the contact angle between the nucleating solution and the surface. The theoretical analysis was found to be quite consistent with the experimental data.

6.2 Theory

According to the classical nucleation theory [165], the particle is assumed to be of spherical shape. The homogeneous nucleation rate is given by [68]

$$J = A^* \exp\left[\frac{-\Delta G_{\text{hom}}}{RT}\right] = A^* \exp\left[-\left(\frac{16\pi M^2 \gamma^3 N_A}{3R^3 T^3 \rho^2}\right) \ln^{-2} S\right] = A^* \exp[-B' \ln^{-2} S] \quad (6.1)$$

in which J , γ , S and A^* are the nucleation rate, the interfacial energy, the supersaturation, and the collision factor, respectively. For a system experiencing mononuclear mechanism, the induction time is inversely proportional to the nucleation rate,

$$t_{\text{ind}} = 1/JV \quad (6.2)$$

where V is the volume of crystallization solution [68]. The relationship between induction time and supersaturation can be obtained by substituting equation. (6.2) to equation (6.1),

$$\ln t_{\text{ind}} = -\ln(A^* V) + B' \ln^{-2} S \quad (6.3)$$

Since the surface energy of nucleus on the substrate was known to be lowered upon interfacial contact, a heterogeneous nucleation on a foreign surface was considered to be energetically less demanding [210]. Therefore,

$$\Delta G_{het} = \phi' \Delta G_{hom} \quad (6.4)$$

$$B'_{het} = \phi' B'_{hom} \quad (6.5)$$

where ϕ' is the ratio of the Gibbs free energy of heterogeneous nucleation to that of homogeneous nucleation. Thus,

$$0 \leq \phi' \leq 1 \quad (6.6)$$

Schematic plots for homogeneous and heterogeneous nucleation rates with different ratios (ϕ') is presented in Figure 6.1.

When a heterogeneous nucleation takes place on a smooth, perfectly planar, and chemically homogeneous surface, the ratio of the Gibbs free energy of heterogeneous nucleation to that of homogeneous nucleation becomes

$$\frac{\Delta G_{het}}{\Delta G_{hom}} = \left(\frac{1}{2} - \frac{3}{4} \cos \theta' + \frac{1}{4} \cos^3 \theta' \right) \quad (6.7)$$

where θ' is the contact angle between nucleus and surface. However, equation (6.7) is not applicable when nucleation occurs on a rough surface. As shown in Figure 6.2, the nucleating solution deposited on a rough surface can be assumed to be of spherical cap shape. The rough surface is assumed to be composed of a series of uniform cones. R' is the main radius of the spherical cap, r'' , h' and n are the radius, height and number of the cones, respectively.

Figure 6.1. Schematic plots for homogeneous and heterogeneous nucleation rates with different coefficients (ϕ').

Figure 6.2. Geometry of a sphere cap shaped nucleating solution on the rough surface.

Geometrical relations of the spherical cap have

$$\alpha' = \frac{r'}{R'} \quad (6.8)$$

$$\beta' = \frac{h'}{R'} \quad (6.9)$$

$$A_L = 2\pi R'^2 (1 - \cos \theta') \quad (6.10)$$

$$A_{SL} = \pi R'^2 \left(\sin^2 \theta' - n \alpha'^2 + n \alpha' \sqrt{\alpha'^2 + \beta'^2} \right) \quad (6.11)$$

$$V = \frac{\pi}{3} R'^3 (1 - \cos \theta')^2 (2 + \cos \theta') + \pi R'^3 \beta' \sin^2 \theta' - \frac{n}{3} \pi R'^3 \alpha'^2 \beta' \quad (6.12)$$

in which A_{SL} , A_L and V are the solid-liquid interfacial area, surface area of the liquid spherical cap, and the volume of the nucleating solution when a droplet of nucleation solution is deposited on the surface. Referring to the work of Curcio *et al.*, the total free energy change due to the formation of a spherical cap droplet of a crystallization solution on the rough surface is [170]

$$\Delta G_{het} = \left(\frac{-\Delta\mu}{\Omega} \right) V + \gamma_L A_L - (\gamma_S - \gamma_i) A_{SL} \quad (6.13)$$

where Ω is the molar volume, and

$$\Delta\mu \approx kT \ln S \quad (6.14)$$

According to Young's equation [211],

$$\gamma_S - \gamma_i = \gamma_L \cos \theta_Y \quad (6.15)$$

in which γ_L , γ_i , γ_S , and θ_Y are the nucleus-liquid interfacial energy, the nucleus-substrate interfacial energy, the liquid-substrate interfacial energy, and the Young contact angle. As Young's equation is based on an ideal surface, Wenzel equation [212] is employed here, and gives

$$\cos \theta_Y = \frac{\cos \theta'}{\Gamma_w} = \frac{\cos \theta' \sin^2 \theta'}{(\sin^2 \theta' - n \alpha'^2 + n \alpha' \sqrt{\alpha'^2 + \beta'^2})} \quad (6.16)$$

where θ' and Γ_w are the apparent contact angle and roughness ratio, respectively.

Therefore

$$\Delta G_{het} = \left(\frac{-\Delta\mu}{\Omega} \right) \left[\frac{\pi}{3} R'^3 (1 - \cos \theta')^2 (2 + \cos \theta') + \pi R'^3 \beta' \sin^2 \theta' - \frac{n}{3} \pi R'^3 \alpha'^2 \beta' \right] + 2\gamma_L \pi R'^2 (1 - \cos \theta') - \gamma_L \pi R'^2 \cos \theta' \sin^2 \theta' \quad (6.17)$$

The radius of a critical cluster nucleated on a rough surface can be obtained by differentiating as

$$\frac{\partial \Delta G_{het}}{\partial R'} = 0 \quad (6.18)$$

Then, the radius R^* of a critical cluster nucleated on a rough surface is

$$R^* = \left(\frac{2\gamma_L \Omega}{\Delta\mu} \right) \left[\frac{2(1 - \cos \theta') - \cos \theta' \sin^2 \theta'}{(1 - \cos \theta')^2 (2 + \cos \theta') + 3\beta' \sin^2 \theta' - n \alpha'^2 \beta'} \right] \quad (6.19)$$

For $\theta' = 180^\circ$, $\cos \theta' = -1$, $n = 0$, equation (6.19) reduces to the classical value given by

Volmer theory for an ideal surface

$$R^* = \frac{2\gamma_L \Omega}{\Delta\mu} \quad (6.20)$$

The critical variation of the free Gibbs energy for heterogeneous nucleation is

$$\Delta G_{het}^* = \frac{4\pi\gamma_L^3 \Omega^2}{3(\Delta\mu)^2} \cdot \frac{[2(1 - \cos \theta') - \cos \theta' \sin^2 \theta']^3}{[(1 - \cos \theta')^2 (2 + \cos \theta') + 3\beta' \sin^2 \theta' - n \alpha'^2 \beta']^2} \quad (6.21)$$

For a critical cluster nucleating in the homogenous phase

$$\Delta G_{hom}^* = \frac{16}{3} \pi \gamma_L^3 \left(\frac{\Omega}{\Delta\mu} \right)^2 \quad (6.22)$$

Dividing equation (6.21) by equation (6.22) gives

$$\frac{\Delta G_{het}^*}{\Delta G_{hom}^*} = \frac{1}{4} \cdot \frac{[2(1 - \cos \theta') - \cos \theta' \sin^2 \theta']^3}{[(1 - \cos \theta')^2 (2 + \cos \theta') + 3\beta' \sin^2 \theta' - n\alpha'^2 \beta']^2} \quad (6.23)$$

Hence, the energy barrier to heterogeneous nucleation taking place on a rough surface depends on the geometric structure and the number of the cones. If $n = 0$, $\alpha' = 0$ and $\beta' = 0$, equation (6.23) can be converted into equation (6.7) for heterogeneous nucleation on an ideal surface.

6.3 Experimental Section

6.3.1 Materials

Hen egg white lysozyme (L7651, 3 × re-crystallized), sodium chloride (≥99.5%), sodium acetate (≥99.5%), acetic acid (HPLC grade), sodium hydroxide (≥98.0%), paraffin oil (for IR spectroscopy), poly-L-glutamic acid (PLG, MW=2000-15000), sulfuric acid (A.R. grade), hydrogen peroxide solution (ACS reagent, 90% w/v in water), (3-aminopropyl)-triethoxysilane (APTES, 99%), poly(2-hydroxyethyl methacrylate) (P2HEMA, MW = 20000), poly(methyl methacrylate) (PMMA, medium molecular weight), poly(4-vinyl-pyridine) (P4VP, MW = 60000), *N,N*-dimethylformamide (DMF, HPLC grade), and chloroform (HPLC grade) were purchased from Sigma-Aldrich. Nickel(□) nitrate hexahydrate (96%) was obtained from Riedel-deHaën. All reagents were used directly without any further purification. The buffer was prepared with ultrafiltered, deionized water from a Milli-Q[®] ultra-pure water purification system.

6.3.2 Preparation of Glass Slide Surfaces

Microscope glass slides were firstly cut into small pieces (0.5×0.5 cm). The obtained glass pieces were carefully cleaned using piranha solution (the mixture of sulfuric acid and

hydrogen peroxide solution in the volume ratio of 3:1), and then rinsed by ultra-pure water and dried in a dust free oven.

The silanized glass slides were prepared using the protocol as described by Tang *et al.* [67]. The freshly cleaned glass slides were firstly washed using 0.1 M nickel nitrate solution and ultra-pure water. Subsequently, the cleaned glass slides were immersed in aqueous solution of 1.0% v/v APTES for 5 min. The solution was removed and the silanized glass slides were rinsed with ultra-pure water extensively and dried in the dust free vacuum oven at 393 K for 2.5 h.

The poly-L-glutamic acid and poly(2-hydroxyethyl methacrylate) coated glass slides were prepared as described by Matsushita *et al.* [213]. The polymer solutions were prepared by dissolving poly-L-glutamic acid and poly(2-hydroxyethyl methacrylate) in phosphate buffered saline (PBS) at the concentration of 1 mg/mL respectively. The freshly cleaned glass slides were immersed in the solutions and incubated at 310 K for 1 h. Then the treated glass slides were rinsed by ultra-pure water and dried in the dust free vacuum oven at 310 K.

Spin coater (Cost Effective Equipment) was applied to coat PMMA and P4VP onto the surfaces of glass slides. The concentrations of PMMA/chloroform and P4VP/DMF solutions were 9 mg/mL and 15 mg/mL, respectively. The samples were then spun in open air for 60 s at a speed of 3,000 rpm [214]. Films were allowed to dry in the air, and then placed in a vacuum oven for 1 h to remove the residual solvents.

6.3.3 Contact Angle Measurement

Contact angle is generally defined as the angle at which a liquid/vapor interface meets a solid surface. In this study, contact angles for both water and protein crystallization solutions on the studied surfaces were measured using FTA-200 (First Ten Angstroms) contact angle meter at room temperature. A droplet of water or protein crystallization solution was dropped onto the investigated surface by an automatic micro-syringe at a rate of $5 \mu\text{L/s}$. After 1 minute's stabilizing, a digital camera was used to record the shape of the droplet. The angle between the baseline of the droplet and the tangent at the droplet boundary was the contact angle. The contact angle reported in this work was the average of five measurements.

6.3.4 Atomic Force Microscopy (AFM) Measurement

A Nanoscope III atomic force microscope (AFM) (Digital Instruments) was applied to characterize the structural characteristics of the surfaces investigated. The images of the surfaces investigated were obtained by applying the tapping mode. The measurements were carried out at room temperature. The scanner sizes were $5 \times 5 \mu\text{m}$ and $1 \times 1 \mu\text{m}$. Nanoscope software was applied to calculate both the average roughness and root mean square roughness. The average roughness and root mean square roughness was defined as the arithmetic average of the absolute values of the surface height deviation measured from the mean plane surface [66] and the standard deviation of the surface profile from the mean plane surface, respectively.

6.3.5 Induction Time Measurement

Hen egg white lysozyme was dissolved in 0.1 M sodium acetate buffer at pH 4.8. Precipitant solution was prepared by dissolving a specific amount of sodium chloride in the same buffer. The pH was adjusted by sodium hydroxide solution and acetic acid. The concentration of lysozyme solution was determined by measuring the absorbance at 280 nm using extinction coefficient 2.64 by a Shimadzu UV 2450 spectrophotometer. The sodium chloride concentration calculated on the basis of the amount of sodium chloride and the solution volume. The solutions were filtered through 0.22 μm sterile filters to remove the foreign nucleation sources for nucleation experiments. The crystallization solution was prepared by mixing the prepared protein solution and the sodium chloride solution in the ratio of 1:1 by volume.

Heterogeneous nucleation experiments were performed on the following surfaces: BG, APTES, PLG, P2HEMA, PMMA and P4VP surfaces. The nucleation experiments were conducted at 277 K with the concentrations of lysozyme and sodium chloride of 20 mg/mL and 2% w/v, respectively. The temperature was controlled by Linkam heating/cooling stage (THMS 600). The surface investigated was placed in the sample holder of the heating stage, then a droplet of 2 μL pre-mixed protein and NaCl solution was deposited on the surface and sealed with paraffin oil.

The induction time was determined visually by an Olympus BX 51 optical microscope attached with a CCD camera. The induction time was defined as the elapsed time between the micro-droplet deposition and the appearance of the first crystal of about 5 μm in diameter. The induction time reported in this paper was the average of five measurements.

The final number and size of protein crystals produced in each batch nucleation experiment were analyzed microscopically as well.

6.4 Results

6.4.1 Contact Angle

The contact angles for water and crystallization solution on the surfaces studied are measured and compared in Figure 6.3. The average values of the contact angles for water on the surfaces studied are: 29.78° for BG, 45.10° for P2HEMA, 47.95° for APTES, 53.74° for PLG and 68.09° for PMMA. And the contact angles for crystallization solution on these surfaces are: 35.40° for BG, 46.83° for P2HEMA, 50.72° for APTES, 54.70° for PLG and 72.21° for PMMA. Each contact angle reported here is the average of five measurements. The stabilizing time is one minute. In fact, contact angles for water and crystallization solution on the above five surfaces slightly decrease with the time (data not shown), which can be attributed to the effect of evaporation [215, 216]. As to the P4VP coated glass slide, the contact angle for water on its surface is 58.91°, while the contact angle for crystallization solution on its surface significantly decreases with the time, as shown in Figure 6.4. Typical illustration images of the decay in contact angle are presented in Figure 6.5.

Figure 6.3. Contact angles for water and crystallization solution (lysozyme, 20 mg/mL; NaCl, 2% w/v) on various surfaces with 1 min's stabilizing time.

Figure 6.4. The time dependence of the contact angles for water and crystallization solution (lysozyme, 20 mg/mL; NaCl, 2% w/v) on the surfaces of P4VP and PMMA.

Figure 6.5. Illustration of the time dependence of the contact angles for crystallization solution (lysozyme, 20 mg/mL; NaCl, 2% w/v) on the surface of P4VP coated glass slide: (a) 52.65° (0 min), (b) 32.51° (1 min), (c) 17.22° (12 min).

6.4.2 Surface Topography

In heterogeneous nucleation, surface roughness is an important parameter to affect nucleation mechanism. In this work, both root mean square (RMS) and average roughness (R_a) were utilized to characterize the surface roughness. The average roughness and RMS of different surfaces are summarized and listed in Table 6.1. The three-dimensional images of BG, P2HEMA, APTES, PLG, PMMA, and P4VP surfaces with the scanner size of $1 \times 1 \mu\text{m}$ are shown in Figure 6.6. Figure 6.6 (a, d, e) presents that the BG, PMMA and APTES surfaces are relatively smooth. A few plateaux and eddies exist on the PMMA surface, which results in larger values of roughness. Figure 6.6 (b, c) shows that a large number of hillocks exist on the surfaces of P2HEMA and PLG. Differently, the size and height of the hillocks on the surface of PLG are larger than that on the P2HEMA surface. The concave parts of the surface of P2HEMA are not as smooth as that on the PLG surface. There are a number of hillocks exist on the surfaces of P4VP surface, as shown in Figure 6.6 (f).

The three-dimensional images of these surfaces with the scanner size of $5 \times 5 \mu\text{m}$ are compared in Figure 6.7. It can be seen that the P4VP surface is full of hillocks and cockles while the BG surface is relatively uniform with sparse spikes. PMMA and APTES surfaces are relatively smooth. Be consistent with the topography obtained with the scanner size of $1 \times 1 \mu\text{m}$, there are a lot of hillocks exists on the surfaces of P2HEMA and PLG surfaces. But the size and heights of the hillocks on these two surfaces are different.

Figure 6.6. Three-dimensional AFM images of the investigated surfaces of (a) bare glass slide, (b) poly(2-hydroxyethyl methacrylate) coated glass slide, (c) poly-L-glutamic acid coated glass slide, (d) poly(methyl methacrylate) coated glass slide, (e) ATPES-silanized glass slide, and (f) poly(4-vinyl pyridine) coated glass slide. The scanner size is 1×1μm.

Figure 6.7. Three-dimensional AFM images of the investigated surfaces of (a) bare glass slide, (b) poly(2-hydroxyethyl methacrylate) coated glass slide, (c) poly-L-glutamic acid coated glass slide, (d) poly(methyl methacrylate) coated glass slide, (e) ATPES-silanized glass slide, and (f) poly(4-vinyl pyridine) coated glass slide. The scanner size is 5×5μm.

TABLE 6.1. Root mean square (RMS) and the average roughness (R_a) of chemically modified glass slides, as derived from AFM analysis.

	range I ^a		range II ^b	
	RMS (nm)	R_a (nm)	RMS (nm)	R_a (nm)
BG	0.298	0.239	3.199	1.405
P2HEMA	2.060	1.634	4.156	3.004
APTES	0.398	0.313	3.819	1.529
PLG	5.829	4.411	10.440	6.138
PMMA	1.896	1.294	4.010	2.956
P4VP	1.301	0.723	2.419	1.943

^a Roughness calculated over a square of 1 μm surface by Nanoscope software.

^b Roughness calculated over a square of 5 μm surface by Nanoscope software.

6.4.3 Induction Time

The measured induction times of heterogeneous nucleation on various surfaces are compared in Figure 6.8. Compared with BG surface, the induction times of heterogeneous nucleation on P2HEMA, APTES and PLG surfaces are decreased, which means that the heterogeneous nucleation is promoted by these surfaces. However, the induction times are increased when nucleation is conducted on PMMA and P4VP surfaces, indicating that the heterogeneous nucleation is suppressed. The suppression is further supported by the number density of the crystals observed after the cessation of nucleation. For instance, in comparison with the crystallization on the BG surface, the number density of crystals grown on the P4VP coated surface is remarkably decreased, as shown in Figure 6.9.

Figure 6.8. The induction times of protein crystallization (20 mg/mL lysozyme; 2% w/v NaCl; 4 °C) on various surfaces. Error bars denote standard deviation ($n_s = 5$).

Figure 6.9. Lysozyme crystals grown on the surfaces of (a) bare glass slide and (b) poly(4-vinyl pyridine) coated glass slide. The crystallization conditions of the two images were the same, which were conducted at 4 °C and with the concentration of lysozyme 20 mg/mL and the concentration of NaCl 2% w/v.

6.5 Discussion

The hydrophobic property of surfaces investigated can be characterized by the contact angles between water and the surfaces. Since nucleation experiments were carried out in sealed systems, the time dependence of the contact angle caused by evaporation can be neglected. For the convenience for comparison, the contact angle values measured after one minute's stabilizing can be used to characterize the investigated surfaces. As shown in Figure 6.3, the contact angles for lysozyme solution on the investigated surfaces are generally higher than those for water on the investigated surfaces. The largest discrepancy (18.9%) has been observed on BG surface. Moderate variations have also been found on PMMA (6.1%) and APTES (5.8%), respectively. The tendency can be applicable to other two surfaces, though the extent of the differences between the contact angles for water and

lysozyme solution on them is reduced (<5%). According to the work of Curcio *et al.* [170], lysozyme was of the same degree of hydrophobicity as the poly(vinylidene fluoride) (PVDF) membrane, which was prepared by a nonsolvent-induced phase inversion process from a homogeneous solution (*N,N*-dimethylacetamide (DMA) 77.5 wt %, poly(vinylpyrrolidone) (PVP) 2.5 wt % and PVDF polymer pellet (KYNAR 460) 20 wt %). The contact angle for water on the PVDF membrane was about 76 °. In the present study, the contact angles for water on the investigated surfaces range from 27 to 69 ° (Figure 6.3), which suggest that the investigated surfaces in this work possess higher degree of hydrophilicity than lysozyme. Although there are numerous exceptions, proteins tend to adsorb more extensively at the hydrophobic than at hydrophilic surface [217-219]. Because of the adsorption of lysozyme on the studied surfaces, the hydrophobicity of the surfaces tends to increase, resulting in the contact angles for lysozyme solution on the surfaces higher than those for water [220], as shown in Figure 6.3.

According to our theoretical analysis, the free energy barrier for heterogeneous nucleation increases with the increase of the contact angle for nucleating solution on the surface. As for P2HEMA, APTES and PMMA surfaces, a qualitative increase of the induction time with the increase of the contact angle is consistent with the tendency of the heterogeneous nucleation barrier ΔG_{het}^* to increase at higher contact angles.

On the other hand, some of hydrophobic surfaces have been shown to promote nucleation, which is attributed to possible preferential binding of the hydrophobic residues on the protein surface with the nucleant surface [72]. The tendency of lysozyme to adsorb on more hydrophobic surface will result in a higher local supersaturation on the more

hydrophobic surface. APTES, P2HEMA and PLG surfaces are of more hydrophobic property than BG. This may serve as another reason for the promotion of heterogeneous nucleation on these three surfaces compared with BG surfaces. Besides hydrophobic interaction, other factors, *e.g.* surface structural characteristics, electrostatic interaction, *etc.*, can affect the heterogeneous nucleation of protein crystals.

According to our theoretical analysis equation (6.23), the thermodynamic term $\Delta G_{het}^* / \Delta G_{hom}^*$ is a function of surface roughness. $\Delta G_{het}^* / \Delta G_{hom}^*$ decreases with the increasing of the surface roughness, as shown in Figure 6.10. The surface roughness measured by atomic force microscopy indicates that BG surface is flatter than P2HEMA and PLG surfaces. The induction time of the heterogeneous nucleation of lysozyme crystals on the BG surface is larger than that on these three surfaces. The result is consistent with the theoretical prediction as shown in equation (6.23) and Figure 6.10.

Figure 6.10. $\Delta G_{het}^* / \Delta G_{hom}^*$ ratio as a function of the contact angle on different roughness.

However, the roughness measured by AFM only may not be sufficient enough to account for the effect of surface structure on the heterogeneous nucleation. This is because the statistical measurements of the roughness are only sensitive to vertical rather than horizontal structures. These statistical parameters are simple and reliable, but they only give information along the vertical direction, and hence cannot fully characterize the surface [221]. It has been reported that material with a rough surface can trap protein molecules and encourage them to nucleate and form crystals [67]. As illustrated in Figure 6.11, protein molecules may adsorb onto the surface randomly and migrate over the surface until the first monolayer is established when nucleation occurred on an ideal surface. Then the second layer grows slowly onto the first layer [218, 222]. During that assembling process, a few molecules in the adsorption layer may form small aggregates [223]. The aggregates, in which protein molecules form suitable bond angles [224] then grow into critical clusters, which would grow into nucleus and later crystals. When heterogeneous nucleation takes place on a rough surface, as shown in Figure 6.12, the hillocks on the surface may block the lateral migration of the adsorbed protein molecules, thus the protein molecules are trapped in between the convexes and concaves and packed into compact structure. Nucleation will happen via the compact critical cluster comprising of molecules forming suitable bond angles with their neighbors. The molecules in randomly packed compact structure may form fractal cluster which, cannot work as a nucleus for crystal growth. Besides, the trapping of molecules on the rough surface may result in a relatively higher local supersaturation, which would increase the possibility of nucleation compared with on an ideally flat surface.

Figure 6.6(c) shows that the surface of PLG coated glass slide is covered with regular, smooth hillocks. Figure 6.6(b) exhibits dense irregular convexes and concaves on the P2HEMA surface. This suggests that the surface topography can also affect the heterogeneous nucleation of protein crystals. One probable reason is that the surface with irregular topography may pack molecules in different ways. The resulted diversity in the bond angle for the protein molecules with its neighbors may increase the probability of nucleation as well. Moreover, electrostatic interaction between protein molecules and the surface probably plays an important role in heterogeneous nucleation. The nucleation can be promoted, when protein molecules are oppositely charged with the surface and suppressed when they are of the same signs. At $\text{pH} = 4.8$, the lysozyme molecule is positively charged. The BG surface carries some unstable negative charge [225]. While the APTES, PLG and P2HEMA surfaces are positively charged as a result of their amino and/or hydroxyl groups. But it is interesting to note that the induction time of nucleation on the positively charged surfaces is decreased in comparison with that on the negatively charged BG surface. The results suggest that the electrostatic interaction may not be the main cause of the decrease of the induction time, which is consistent with the work reported by Fermani *et al.* [69].

Figure 6.11. Schematic illustration of the processes of protein aggregation and crystallization on ideally flat surface. (a) Lysozyme molecules are dispersed in solution, (b) The adsorption of lysozyme molecules: the adsorption of lysozyme molecules may be reversible; lysozyme molecules may arrive at the surface randomly and migrate over the surface to form the absorption layer, (c) Lysozyme molecules form multi-layers layer by layer, (d) Some of the molecules form small aggregates, (e) The formation of quasi-critical clusters. (f) The formation of crystals from the quasi-critical clusters.

Figure 6.12. Schematic illustration of protein crystallization and the formation of large cluster on rough surface. (a) Lysozyme molecules are dispersed in solution. (b) The adsorption of lysozyme molecules from the solution to the surface; the presence of hillocks on the surface may block the lateral migration of the molecules. (c) The blocked molecules are packed into compact aggregates. (d) The molecules that are packed with bond angles corresponding to protein crystals form quasi-critical clusters; the molecules that are randomly packed form fractal clusters. (e) Quasi-critical clusters nucleate and grow into crystals, while the fractal clusters grow into larger clusters.

Although P4VP coated glass slide surface is of higher roughness (Figure 6.6 and Figure 6.7), the heterogeneous nucleation is suppressed notably on its surface compared with the bare glass slide. In addition, the contact angle of crystallization solution on its surface decreases significantly with the time (Figure 6.4), which is different from other chemically modified surfaces. Polyelectrolytes were generally assumed to have an extended conformation in water because of the electrostatic repulsion among the charge groups [226]. As shown in Figure 6.13, the pyridine ring in P4VP is positively charged when P4VP contacts with the crystallization solution. The extended conformation in the solution may be the reason for the significant decay of the contact angle for crystallization solution. Similarly, the suppressed nucleation on the P4VP surface can be attributed to the electrical repulsive interaction between the P4VP surface and the positively charged lysozyme molecules. That is, the pyridine ring in P4VP molecule is of a better capacity to carry the positive charge compared with the hydroxyl and amino groups. The increased positive charge that appears on the P4VP surface increases the electrostatic repulsion between lysozyme and P4VP, and thus lysozyme molecules can be stabilized by the polycation chains provided by P4VP [227]. Therefore, the nucleation was inhibited on the P4VP surface.

Figure 6.13. Schematic illustration of the swelling of P4VP in contact with the crystallization solution (lysozyme, 20 mg/mL; NaCl, 2% w/v).

6.6 Conclusions

A series of chemically modified glass slides has been tested for the heterogeneous nucleation of lysozyme crystals. In comparison with on bare glass slide, the heterogeneous nucleation on P2HEMA, PLG, and APTES surfaces was promoted, whereas it was suppressed on P4VP and PMMA surfaces. Based on the contact angle and AFM measurements, it was found that surface structural characteristics, chemical and physical

interactions between the surfaces and protein molecules had influenced on the heterogeneous nucleation of protein crystals to different extents. And various surfaces may affect the heterogeneous nucleation in different mechanisms. Both the experimental data and the theoretical analysis suggested that the energetic barrier to heterogeneous nucleation was found to decrease at lower contact angles, and to increase at lower roughness. However, the surface roughness alone was not able to account for the differences of the induction time of the heterogeneous nucleation on different surfaces. Surfaces with irregular surface topographies can increase the probability of heterogeneous nucleation. In addition, the specific interactions between protein molecules and the surfaces probably play a role in the heterogeneous nucleation. For example, the repulsive electrostatic interaction between protein molecules and the surface would stabilize protein molecules and inhibit nucleation. The present work provides a further understanding of the mechanism of the heterogeneous nucleation of protein crystals, and an inspiration to design heterogeneous nucleant.

Part of results in this chapter has been reported on Journal of Physical Chemistry B (Vol 111, pp. 13971-13978, 2007)

7 Conclusions and Future Directions

7.1 Conclusions

The present work systematically studied the phase behavior and crystallization mechanisms of protein molecules. In particular, the crystallization of lysozyme in sodium chloride solutions were investigated by measuring the solubility, the second virial coefficient, the cloud-point temperature, the nucleation rate and the crystal growth rate. In addition, the influences of roughness, topography and physicochemical properties of chemically modified surfaces on the heterogeneous nucleation of protein crystals were examined by measuring the induction time, the contact angle, and the surface roughness. Systematic measurement and analysis in this work were expected to provide a further understanding of protein crystallization process and a general guideline to search for the optimal crystallization conditions.

In Chapter 3, the effects of pH, temperature, salt type and ionic strength on the solubility and second virial coefficient of lysozyme in various salt solutions were investigated. Crystallization trials in the corresponding conditions were performed in parallel. Additionally, crystals obtained in the crystallization trials were characterized using optical microscopy and Raman spectroscopy. It was found that the solubility decreased with increasing salt concentration, decreasing temperature, and the increase of pH from 4.2 to 5.2. It was also observed that there was a significant correlation between the solubility and the second virial coefficient. The solubility decreased as the second virial coefficient became large and more negative. Crystals of high quality were obtained in solutions of moderate attractive protein interactions. No crystals formed when the attractive interactions were not strong enough. On the other hand, attractive interactions

that were too strong resulted in the formation of crystals that were not of good diffraction quality. The salts of different anions and cations were observed to be of different abilities to precipitate protein from aqueous solutions. The incorporation of different anions in protein crystals resulted in the crystals that were of different morphologies. The experimental data were expected to extend previous studies on the correlation between interaction and phase behavior of protein molecules.

In Chapter 4, the low density region of liquid-liquid co-existence curves for lysozyme solution with 2%, 3%, 4% and 5% w/v sodium chloride at pH 4.2 and bovine serum albumin solutions with 60 mg/mL, 70 mg/mL and 80 mg/mL polyethylene glycol 1500 at pH 5.0 was determined by measuring the cloud-point temperature. For both lysozyme and bovine serum albumin solutions, the location of the liquid-liquid co-existence curve shifted up in temperature with the increase of the concentration of the precipitant, but the basic shapes of the curves were the same. At a fixed precipitant concentration, the cloud-point temperature monotonically increased with the concentration of protein. In order to correlate the cloud-point temperature and the second virial coefficient, the second virial coefficients for lysozyme and bovine serum albumin in various precipitant concentrations and temperatures were measured. The dimensionless cloud-point temperatures were plotted versus the dimensionless second virial coefficient. A strong correlation was found to exist between the cloud-point temperature and the second virial coefficient. In addition, the random phase approximation, in conjunction with a square-well potential was applied to describe the cloud-point temperatures reported in literature. A better representation of the experimental data was accomplished by assuming the protein interactions to be temperature dependent. The temperature dependence of the second virial coefficient was

measured to provide a support for this assumption. Although much work is still required to test the physical reasonability of this assumption, experimental data and modified model had raised a number of interesting theoretical and experimental issues regarding the interactions between protein molecules and its correlation with the phase behavior of protein molecules. The further investigation regarding these issues may provide a better insight into the phase transition of protein molecules.

In Chapter 5, a detailed study on the phase behavior and mechanism of lysozyme crystallization was presented. The nucleation and crystal growth rates, crystal morphologies, the solubility, the second virial coefficient, and the cloud-point temperatures under different solution conditions were experimentally measured and theoretically analyzed. In agreement with the classical nucleation theory, the nucleation rates were found to increase with the protein and sodium chloride concentrations when the crystallization occurred in the solid-liquid co-existence region. The different shapes of obtained crystals were caused by the different growth rates of the (110) and (101) faces of tetragonal crystals. The broken dependence of nucleation rate on protein concentration around the liquid-liquid co-existence boundary could be attributed to the liquid-liquid phase separation. Under our experimental conditions, it was found that the protein-rich phase did not cause the nucleation of new crystals but dissolved into the “buffer zone” to restore the local concentrations. In addition, the suppression of the nucleation in the protein-rich phase was probed in terms of protein-protein interactions determined by Raman microscopy and the self-assembly of the protein molecules. A relatively stronger protein-protein interaction in the protein-rich phase could make it difficult to adjust the conformation and orientation of protein molecules to rearrange their association structure

to achieve the nucleation of protein crystals. Furthermore, the concentration distribution in the region between the growing nucleus and the protein-rich phase was derived to explain the increase of nucleation rate in the liquid-liquid co-existence region. These new observations and analysis were expected to provide further understanding and guideline for protein crystallization.

In Chapter 6, the influence of some factors on the heterogeneous nucleation of lysozyme on a series of chemically modified surfaces has been investigated. Micro-batch crystallization experiments were conducted on the microscope glass slides that were treated with poly-L-glutamic acid, poly(2-hydroxyethyl methacrylate), poly(methyl methacrylate), poly(4-vinyl pyridine), and (3-aminopropyl)triethoxysilane. An optical microscope with a heating/cooling stage was employed to measure the induction time of the heterogeneous nucleation. Surface topography and roughness were characterized by AFM. Contact angles for crystallized solution on the investigated surfaces were measured by a contact angle meter. From the theoretical analysis, the energetic barrier to the heterogeneous nucleation was found to increase at higher contact angles, and to decrease at higher roughness. Experimentally, a qualitative increase of the induction time of the heterogeneous nucleation on poly(2-hydroxyethyl methacrylate), (3-aminopropyl)triethoxysilane and poly (methyl methacrylate) surfaces with the contact angles was observed. Such surfaces as poly(2-hydroxyethyl methacrylate) and poly-L-glutamic acid, which were of higher roughness, were shown to promote the heterogeneous nucleation. In addition, the surface with specific topography was expected to increase the possibility of the formation of a critical nucleus. Finally, poly(4-vinyl pyridine) surface appeared to suppress the heterogeneous nucleation due to the electrostatic interaction

between lysozyme and poly(4-vinyl pyridine) molecules. The experimental data and theoretical analysis in this work provide a further understanding of the heterogeneous nucleation mechanism of protein molecules and some useful suggestions in the progress of searching the universal heterogeneous nucleants.

7.2 Suggestions for Future Work

Although a systematic investigation of the phase behavior and protein crystallization mechanism for lysozyme has been established, these conclusions drawn in this study are required to be extended to other proteins.

It is known that membrane proteins play important roles in a lot of biological processes, such as a transporter for the cell. Thus, a detailed understanding on the three-dimensional structures of membrane proteins is essential in the understanding of the fundamental mechanisms of these processes. However, the structure data for membrane protein are limited due to the difficulty to crystallize membrane proteins. They are insoluble in water. In order to crystallize membrane protein, usage of detergent is required. Although many academic efforts have been made on the mechanism of membrane protein crystallization, they mainly focus on the water soluble proteins such as lysozyme. In order to facilitate the structure determination of membrane protein, it is necessary to carry out the research on the mechanism of the crystallization of membrane proteins and the approach to screen for optimal crystallization conditions for membrane proteins.

So far, a lot of materials have been tested as heterogeneous nucleants. It has been found that there are various factors that influence the heterogeneous nucleation of protein

molecules in different ways. In order to search for the universal heterogeneous nucleant, more efforts are required to study the specific effects of different parameters, such as the electrostatic interaction, the hydrophobic interaction, and the microstructure of the surfaces, *etc* on the heterogeneous nucleation of protein crystals.

With the intensive investigations on the nucleation mechanisms of protein crystals, some proteins that could not be crystallized have been crystallized successfully. All too often, crystals obtained are of poor quality. In the future, protocols on the post-crystallization treatments are desired to improve the quality of the obtained crystals.

In the view point of drug design, it is necessary to understand the interaction between the candidate compound and the target protein. Thus, the investigation on the protein-ligand complex crystallization mechanism is also important in the rational design of the crystallization of crystals of high quality.

The search of the ways to improve the physical, chemical and biological properties of protein crystals is also of great interest in the study of protein crystallization.

8 References

- [1] O. Annunziata, A. Payne, Y. Wang, Solubility of lysozyme in the presence of aqueous chloride salts: common-ion effect and its role on solubility and crystal thermodynamics, *J. Am. Chem. Soc.* 130 (2008) 13347-13352.
- [2] M.L. Brader, M. Sukumar, A.H. Pekar, D.S. McClellan, R.E. Chance, D.B. Flora, A.L. Cox, L. Irwin, S.R. Myers, Hybrid insulin cocrystals for controlled release delivery, *Nat. Biotechnol.* 20 (2002) 800-804.
- [3] H.P. Merkle, A. Jen, A crystal clear solution for insulin delivery, *Nat. Biotechnol.* 20 (2002) 789-790.
- [4] S. Stolnik, K. Shakesheff, Formulations for delivery of therapeutic proteins, *Biotechnol. Lett.* 31 (2009) 1-11.
- [5] A.L. Margolin, M.A. Navia, Protein crystals as novel catalytic materials, *Angew. Chem. Int. Ed.* 40 (2001) 2204-2222.
- [6] E. Cacioppo, S. Munson, M.L. Pusey, Protein solubilities determined by a rapid technique and modification of that technique to a micro-method, *J. Crystal Growth* 110 (1991) 66-71.
- [7] E. Cacioppo, M.L. Pusey, The solubility of the tetragonal form of hen egg white lysozyme from pH 4.0 to 5.4, *J. Crystal Growth* 114 (1991) 286-292.
- [8] J.B. Cole, M.L. Bryan, W.P. Bryan, Thermodynamics of solution of lysozyme crystals, *Arch. Biochem. Biophys.* 130 (1969) 86-91.
- [9] S.B. Howard, P.J. Twigg, J.K. Baird, E.J. Meehan, The solubility of hen egg-white lysozyme, *J. Crystal Growth* 90 (1988) 94-104.
- [10] M. Ataka, M. Asai, Systemic studies on the crystallization of lysozyme: determination and use of phase diagrams, *J. Crystal Growth* 90 (1988) 86-93.

- [11] M.M. Ries-Kautt, A.F. Ducruix, Relative effectiveness of various ions on the solubility and crystal growth of lysozyme, *J. Biol. Chem.* 264 (1989) 745-748.
- [12] E.L. Forsythe, R.A. Judge, M.L. Pusey, Tetragonal chicken egg white lysozyme solubility in sodium chloride solutions, *J. Chem. Eng. Data* 44 (1999) 637-640.
- [13] V. Mikol, R. Giegé, Phase diagram of a crystalline protein: determination of the solubility of concanavalin A by a microquantitation assay, *J. Crystal Growth* 97 (1989) 324-332.
- [14] R.A. Judge, M.R. Johns, E.T. White, Solubility of ovalbumin in ammonium sulfate solutions, *J. Chem. Eng. Data* 41 (1996) 422-424.
- [15] R.C. DeMattei, R.S. Feigelson, The solubility dependence of canavalin on pH and temperature, *J. Crystal Growth* 110 (1991) 34-40.
- [16] R.J. Gray, W.B. Hou, A.B. Kudryavtsev, L.J. DeLucas, A new approach to the measurement of protein solubility by Michaelson interferometer, *J. Crystal Growth* 232 (2001) 10-16.
- [17] Y. Suzuki, G. Sasaki, K. Visuri, K. Tamura, K. Nakajima, S. Yanagiya, Significant decrease in the solubility of glucose isomerase crystals under high pressure, *Cryst. Growth Des.* 2 (2002) 321-324.
- [18] S. Tanaka, M. Ataka, K. Ito, Pattern formation and coarsening during metastable phase separation in lysozyme solutions, *Phys. Rev. E* 65 (2002) 051804.
- [19] O. Galkin, P.G. Vekilov, Direct determination of the nucleation rates of protein crystals, *J. Phys. Chem. B* 103 (1999) 10965-10971.
- [20] O. Galkin, P.G. Vekilov, Nucleation of protein crystals: critical nuclei, phase behavior, and control pathways, *J. Crystal Growth* 232 (2001) 63-76.

- [21] O. Galkin, P.G. Vekilov, Control of protein crystal nucleation around the metastable liquid-liquid phase boundary, *Proc. Natl. Acad. Sci.* 12 (2000) 6277-6281.
- [22] S. Tanaka, M. Yamamoto, K. Ito, R. Hayakawa, M. Ataka, Relation between the phase separation and the crystallization in protein solutions, *Phys. Rev. E* 56 (1997) R67-R69.
- [23] P.S. Chow, X.Y. Liu, J. Zhang, R.B.H. Tan, Spherulitic growth kinetics of protein crystals, *Appl. Phys. Lett.* 81 (2002) 1975-1977.
- [24] M.C.R. Heijna, W.J.P. van Enkevort, E. Vlieg, Crystal growth in a three-phase system: diffusion and liquid-liquid phase separation in lysozyme crystal growth, *Phys. Rev. E* 76 (2007) 011604.
- [25] P.R. ten Wolde, D. Frenkel, Enhancement of protein crystal nucleation by critical density fluctuations, *Science* 277 (1997) 1975-1978.
- [26] P.R. ten Wolde, D. Frenkel, Enhanced protein crystallization around the metastable critical point, *Theor. Chem. Accounts* 101 (1999) 205-208.
- [27] C. Haas, J. Drenth, W.W. Wilson, Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution, *J. Phys. Chem. B* 103 (1999) 2808-2811.
- [28] N.M. Dixit, C.F. Zukoski, Crystal nucleation rates for particles experiencing short-range attractions: applications to proteins, *J. Colloid Interface Sci.* 228 (2000) 359-371.
- [29] J.J. Grigsby, H.W. Blanch, J.M. Prausnitz, Cloud-point temperatures for lysozyme in electrolyte solutions: effect of salt type, salt concentration and pH, *Biophys. Chem.* 91 (2001) 231-243.
- [30] E.J. Park, Y.C. Bae, Cloud-point temperatures of lysozyme in electrolyte solutions by thermooptical analysis technique, *Biophys. Chem.* 109 (2004) 169-188.

- [31] J. Lu, K. Carpenter, R.J. Li, X.J. Wang, C.B. Ching, Cloud-point temperature and liquid-liquid phase separation of supersaturated lysozyme solution, *Biophys. Chem.* 109 (2004) 105-112.
- [32] D. Vivarès, F. Bonneté, Liquid-liquid phase separations in urate oxidase/PEG mixtures: characterization and implications for protein crystallization, *J. Phys. Chem. B* 108 (2004) 6498-6507.
- [33] S. Grouazel, J. Perez, J.P. Astier, F. Bonneté, S. Veessler, BPTI liquid-liquid phase separation monitored by light and small-angle X-ray scattering, *Acta Crystallogr., Sect. D* 58 (2002) 1560-1563.
- [34] S. Grouazel, F. Bonneté, J.P. Astier, N. Ferté, J. Perez, S. Veessler, Exploring bovine pancreatic trypsin inhibitor phase transitions, *J. Phys. Chem. B* 110 (2006) 19664-19670.
- [35] C. Ishimoto, T. Tanaka, Critical behavior of a binary mixture of protein and salt water, *Phys. Rev. Lett.* 39 (1977) 474-477.
- [36] V.G. Taratuta, A. Holschbach, G.M. Thurston, D. Blankschtein, G.B. Benedek, Liquid-liquid phase separation of aqueous lysozyme solutions: effects of pH and salt identity, *J. Phys. Chem.* 94 (1990) 2140-2144.
- [37] M. Muschol, F. Rosenberger, Liquid-liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization, *J. Chem. Phys.* 107 (1997) 1953-1962.
- [38] Y. Wang, O. Annunziata, Comparison between protein-polyethylene glycol (PEG) interactions and the effect of PEG on protein-protein interactions using the liquid-liquid phase transition, *J. Phys. Chem. B* 111 (2007) 1222-1230.

- [39] A. George, W.W. Wilson, Predicting protein crystallization from a dilute solution property, *Acta Crystallogr., Sect. D* 50 (1994) 361-365.
- [40] O.D. Velev, E.W. Kaler, A.M. Lenhoff, Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen, *Biophys. J.* 75 (1998) 2682-2697.
- [41] D.F. Rosenbaum, C.F. Zukoski, Protein interactions and crystallization, *J. Crystal Growth* 169 (1996) 752-758.
- [42] M. Farnum, C. Zukoski, Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor, *Biophys. J.* 76 (1999) 2716-2726.
- [43] F. Bonneté, S. Finet, A. Tardieu, Second virial coefficient: variations with lysozyme crystallization conditions, *J. Crystal Growth* 196 (1990) 403-414.
- [44] F. Bonneté, D. Vivarès, Interest of the normalized second virial coefficient and interaction potentials for crystallizing large macromolecules, *Acta Crystallogr., Sect. D* 58 (2002) 1571-1575.
- [45] P.M. Tessier, A.M. Lenhoff, S.I. Sandler, Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography, *Biophys. J.* 82 (2002) 1620-1631.
- [46] A.C. Dumetz, A.M. Snellinger-O'Brien, E.W. Kaler, A.M. Lenhoff, Patterns of protein-protein interactions in salt solutions and implications for protein crystallization, *Protein Sci.* 16 (2007) 1867-1877.
- [47] A.C. Dumetz, A.M. Chockla, E.W. Kaler, A.M. Lenhoff, Effects of pH on protein-protein interactions and implications for protein phase behavior, *Biochim. Biophys. Acta* 1784 (2008) 600-610.

- [48] P.M. Tessier, J. Jinkoji, Y.C. Cheng, J.L. Prentice, A.M. Lenhoff, Self-interaction nanoparticle spectroscopy: a nanoparticle-based protein interaction assay, *J. Am. Chem. Soc.* 130 (2008) 3106-3112.
- [49] B. Guo, S. Kao, H. McDonald, A. Asanov, L.L. Combs, W.W. Wilson, Correlation of second virial coefficients and solubilities useful in protein crystal growth, *J. Crystal Growth* 196 (1999) 424-433.
- [50] B.W. Berger, C.M. Gendron, A.M. Lenhoff, E.W. Kaler, Effects of additives on surfactant phase behavior relevant to bacteriorhodopsin crystallization, *Protein Sci.* 15 (2006) 2682-2696.
- [51] R.A. Judge, R.S. Jacobs, T. Frazier, E.H. Snell, M.L. Pusey, The effect of temperature and solution pH on the nucleation of tetragonal lysozyme crystals, *Biophys. J.* 77 (1999) 1585-1593.
- [52] O. Galkin, P.G. Vekilov, Are nucleation kinetics of protein crystals similar to those of liquid droplets?, *J. Am. Chem. Soc.* 122 (2000) 156-163.
- [53] V. Bhamidi, E. Skrzypczak-Jankun, C.A. Schall, Dependence of nucleation kinetics and crystal morphology of a model protein system on ionic strength, *J. Crystal Growth* 232 (2001) 77-85.
- [54] V. Bhamidi, S. Varanasi, C.A. Schall, Measurement and modeling of protein crystal nucleation kinetics, *Cryst. Growth Des.* 2 (2002) 395-400.
- [55] C. Haas, J. Drenth, The interface between a protein crystal and an aqueous solution and its effects on nucleation and crystal growth, *J. Phys. Chem. B* 104 (2000) 368-377.
- [56] A. McPherson, P. Shlichta, Heterogeneous and epitaxial nucleation of protein crystals on mineral surfaces, *Science* 239 (1988) 385-387.

- [57] W.L. Kimble, R.W. Rousseau, A. Sambanis, Lysozyme crystallization by vapor diffusion: characterization and modeling in the absence and presence of exogenous minerals, *J. Crystal Growth* 147 (1995) 165-171.
- [58] W.L. Kimble, T.E. Paxton, R.W. Rousseau, A. Sambanis, The effect of mineral substrates on the crystallization of lysozyme, *J. Crystal Growth* 187 (1998) 268-276.
- [59] A. McPherson, P.J. Shlichta, Facilitation of the growth of protein crystals by heterogeneous/epitaxial nucleation, *J. Crystal Growth* 85 (1987) 206-214.
- [60] A. McPherson, P. Shlichta, The use of heterogeneous and epitaxial nucleants to promote the growth of protein crystals, *J. Crystal Growth* 90 (1988) 47-50.
- [61] L. Rong, H. Komatsu, M. Natsuisaka, S. Yoda, Expitaxial nucleation of protein crystal on poly-L-lysine modified surface, *Jpn. J. Appl. Phys.* 40 (2001) 6677-6678.
- [62] L. Rong, H. Komatsu, S. Yoda, Control of heterogeneous nucleation of lysozyme crystals by using poly-L-lysine modified substrate, *J. Crystal Growth* 235 (2002) 489-493.
- [63] R.D. Kornberg, S.A. Darst, Two-dimensional crystals of proteins on lipid layers, *Curr. Opin. Struct. Biol.* 1 (1991) 642-646.
- [64] A.C. Ku, S.A. Darst, C.R. Robertson, A.P. Gast, R.D. Kornberg, Molecular analysis of two-dimensional protein crystallization, *J. Phys. Chem.* 97 (1993) 3013-3016.
- [65] S.A. Darst, A.M. Edwards, Epitaxial growth of protein crystals from two-dimensional crystals on lipid layers, *Curr. Opin. Struct. Biol.* 5 (1995) 640-644.
- [66] G. Falini, S. Fermani, G. Conforti, A. Ripamonti, Protein crystallization on chemically modified mica surfaces, *Acta Crystallogr., Sect. D* 58 (2002) 1649-1652.
- [67] L. Tang, Y.B. Huang, D.Q. Liu, J.L. Li, K. Mao, L. Liu, Z.J. Cheng, W.M. Gong, J. Hu, J.H. He, Effects of the silanized mica surface on protein crystallization, *Acta Crystallogr., Sect. D* 61 (2005) 53-59.

- [68] T.E. Paxton, A. Sambanis, R.W. Rousseau, Influence of vessel surfaces on the nucleation of protein crystals, *Langmuir* 17 (2001) 3076-3079.
- [69] S. Fermani, G. Falini, M. Minnucci, A. Ripamonti, Protein crystallization on polymeric film surfaces, *J. Crystal Growth* 224 (2001) 327-334.
- [70] N.E. Chayen, E. Saridakis, R. El-Bahar, Y. Nemirovsky, Porous silicon: an effective nucleation-inducing material for protein crystallization, *J. Mol. Biol.* 312 (2001) 591-595.
- [71] N.E. Chayen, E. Saridakis, R.P. Sear, Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium, *Proc. Natl. Acad. Sci.* 103 (2006) 597-601.
- [72] S. Stolyarova, E. Saridakis, N.E. Chayen, Y. Nemirovsky, A model for enhanced nucleation of protein crystals on a fractal porous substrate, *Biophys. J.* 91 (2006) 3857-3863.
- [73] P.A. Bancel, V.B. Cajipe, F. Rodier, J. Witz, Laser seeding for biomolecular crystallization, *J. Crystal Growth* 191 (1998) 537-544.
- [74] D.Y. Zhu, Y.Q. Zhu, Y. Xiang, D.C. Wang, Optimizing protein crystal growth through dynamic seeding, *Acta Crystallogr., Sect. D* 61 (2004) 772-775.
- [75] A. D'Arcy, A.M. Sweeney, A. Haber, Using natural seeding material to generate nucleation in protein crystallization experiments, *Acta Crystallogr., Sect. D* 59 (2003) 1343-1346.
- [76] A. D'Arcy, A. Mac Sweeney, A. Haber, Modified microbatch and seeding in protein crystallization experiments, *J. Synchrotron Rad.* 11 (2004) 24-26.

- [77] M. Taleb, C. Didierjean, C. Jelsch, J.P. Mangeot, B. Capelle, A. Aubry, Crystallization of proteins under an external electric field, *J. Crystal Growth* 200 (1999) 575-582.
- [78] G. Sazaki, E. Yoshida, H. Komatsu, T. Nakada, S. Miyashita, K. Watanabe, Effects of a magnetic field on the nucleation and growth of protein crystals, *J. Crystal Growth* 173 (1997) 231-234.
- [79] S.D. Durbin, G. Feher, Crystal growth studies of lysozyme as a model for protein crystallization, *J. Crystal Growth* 76 (1986) 583-592.
- [80] E. Forsythe, F. Ewing, M.L. Pusey, Studies on tetragonal lysozyme crystal growth rates, *Acta Crystallogr., Sect. D* 50 (1994) 614-619.
- [81] E. Forsythe, M.L. Pusey, The effects of temperature and NaCl concentration on tetragonal lysozyme face growth rates, *J. Crystal Growth* 139 (1994) 89-94.
- [82] S.D. Durbin, G. Feher, Studies of crystal growth mechanisms of proteins by electron microscopy, *J. Mol. Biol.* 212 (1990) 763-774.
- [83] S.D. Durbin, W.E. Carlson, Lysozyme crystal growth studied by atomic force microscopy, *J. Crystal Growth* 122 (1992) 71-79.
- [84] G. Sazaki, K. Tsukamoto, S. Yai, M. Okada, K. Nakajima, In situ observation of dislocations in protein crystals during crystal growth by advanced optical microscopy, *Cryst. Growth Des.* 5 (2005) 1729-1735.
- [85] H. Takashima, High-resolution protein structure determination by NMR, *Annu. Rep. NMR. Spect.* 59 (2006) 235-273.
- [86] A.G. Tzacos, C.R.R. Grace, P.J. Lukavsky, R. Riek, NMR techniques for very large proteins and RNAs in solution, *Annu. Rev. Biophys. Biomol. Struct.* 35 (2006) 319-342.

- [87] M. Mueller, S. Jenni, N. Ban, Strategies for crystallization and structure determination of very large macromolecular assemblies, *Curr. Opin. Struct. Biol.* 17 (2007) 572-579.
- [88] B.A. Maniasetty, A.P. Turnbull, S. Panjikar, K. Büssow, M.R. Chance, Automated technologies and novel techniques to accelerate protein crystallography for structural genomics, *Proteomics* 8 (2008) 612-625.
- [89] S. Subramaniam, J.L.S. Milne, Three-dimensional electron microscopy at molecular resolution, *Annu. Rev. Biophys. Biomol. Struct.* 33 (2004) 141-155.
- [90] M. Auer, Three-dimensional electron cryo-microscopy as a powerful structural tool in molecular medicine, *J. Mol. Med.* 78 (2000) 191-202.
- [91] A. Böckmann, High-resolution solid-state MAS NMR of protein-Crh as an example, *Magn. Reson. Chem.* 45 (2007) s24-s31.
- [92] <http://www.rcsb.org/pdb/home/home.do>.
- [93] S.P. Williams, L.F. Kuyper, K.H. Pearce, Recent applications of protein crystallography and structure-guided drug design, *Curr. Opin. Chem. Biol.* 9 (2005) 371-380.
- [94] S. Thaisrivongs, D.L. Romero, R.A. Tommasi, M.N. Janakiraman, J.W. Strohbach, S.R. Turner, C. Biles, R.R. Morge, P.D. Johnson, P.A. Aristoff, P.K. Tomich, J.C. Lynn, M.M. Horng, K.T. Chong, R.R. Hinshaw, W.J. Howe, B.C. Finzel, K.D. Watenpaugh, Structure-based design of HIV protease inhibitors: 5,6-dihydro-4-hydroxy-2-pyrones as effective, nonpeptidic inhibitors, *J. Med. Chem.* 39 (1996) 4630-4642.
- [95] H.I. Skulnick, P.D. Johnson, P.A. Aristoff, J.K. Morris, K.D. Lovasz, W.J. Howe, K.D. Watenpaugh, M.N. Janakiraman, D.J. Anderson, R.J. Reischer, T.M. Schwartz, L.S. Banitt, P.K. Tomich, J.C. Lynn, M.M. Horng, K.T. Chong, R.R. Hinshaw, L.A. Dolak,

E.P. Seest, F.J. Schwende, B.D. Rush, G.M. Howard, L.N. Toth, K.R. Wilkinson, T.J. Kakuk, C.W. Johnson, S.L. Cole, R.M. Zaya, G.L. Zipp, P.L. Possert, R.J. Dalga, W.Z. Zhong, M.G. Williams, K.R. Romines, Structure-based design of nonpeptidic HIV protease inhibitors: the sulfonamide-substituted cyclooctopyranones, *J. Med. Chem.* 40 (1997) 1149-1164.

[96] S. Thaisrivongs, J.W. Strohbach, Structure-based discovery of Tipranavir Disodium (PNU-140690E): a potent, orally bioavailable, nonpeptidic HIV protease inhibitor, *Biopolymers* 51 (1999) 51-58.

[97] D.L.N.G. Surleraux, H.A. de Kock, W.G. Verschueren, G.M.E. Pille, L.J.R. Maes, A. Peeters, S. Vendeville, S.D. Meyer, H. Azijin, R. Pauwels, M.P. de Bethune, N.M. King, M. Prabu-Jeyabalan, C.A. Schiffer, P.B.T.P. Wigerinck, Design of HIV-1 protease inhibitors active on multidrug-resistant virus, *J. Med. Chem.* 48 (2005) 1965-1973.

[98] M. Anderson, J.F. Beattie, G.A. Breault, J. Breed, K.F. Byth, J.D. Culshaw, R.P.A. Ellston, S. Green, C.A. Minshull, R.A. Norman, R.A. Pauptit, J. Stanway, A.P. Thomas, P.J. Jewsbury, Imidazol[1,2-a]pyridines: a potent and selective class of cyclin-dependent kinase inhibitors identified through structure-based hybridisation, *Bioorg. Med. Chem. Lett.* 13 (2003) 3021-3026.

[99] N. Foloppe, L.M. Fisher, R. Howes, P. Kierstan, A. Potter, A.G.S. Robertson, A.E. Surgenor, Structure-based design of novel Chk1 inhibitors: insights into hydrogen bonding and protein-ligand affinity, *J. Med. Chem.* 48 (2005) 4332-4345.

[100] P. Furet, P. Imbach, M. Noorani, J. Koeppler, K. Laumen, M. Lang, V. Guagnano, P. Fuerst, J. Roesel, J. Zimmermann, C. García-Echeverría, Entry into a new class of potent proteasome inhibitors having high antiproliferative activity by structure-based drug design, *J. Med. Chem.* 47 (2004) 4810-4813.

- [101] Q. Li, A. Claiborne, T.M. Li, L. Hasvold, V.S. Stoll, S. Muchmore, C.G. Jakob, W. Gu, J. Cohen, C. Hutchins, D. Frost, S.H. Rosenberg, H.L. Sham, Design, synthesis, and activity of 4-quinolone and pyridone compounds as nonthiol-containing farnesyltransferase inhibitors, *Bioorg. Med. Chem. Lett.* 14 (2004) 5367-5370.
- [102] M.S. Coumar, J.S. Leou, P. Shukla, J.S. Wu, A.K. Dixit, W.H. Lin, C.Y. Chang, T.W. Lien, U.K. Tan, C.H. Chen, J.T.A. Hsu, Y.S. Chao, S.Y. Wu, H.P. Hsien, Structure-based drug design of novel Aurora kinase A inhibitors: structural basis for potency and specificity, *J. Med. Chem.* 52 (2009) 1050-1062.
- [103] S.J. Stachel, C.A. Coburn, T.G. Steele, K.G. Jones, E.F. Loutzenhister, A.R. Gregro, H.A. Rajapakse, M.T. Lai, M.C. Crouthamel, M. Xu, K. Tugusheva, J.E. Lineberger, B.L. Peitrak, A.S. Espeseth, X.P. Shi, E. Chen-Dodson, M.K. Holloway, S. Munshi, A.J. Simon, L. Kuo, J.P. Vacca, Structure-based design of potent and selective cell-permeable inhibitors of human β -secretase (BACE-1), *J. Med. Chem.* 47 (2004) 6447-6450.
- [104] R. Machauer, S. Veenstra, J.M. Rondeau, M. Tintelnot-Blomley, C. Betschart, U. Neumann, P. Paganetti, Structure-based design and synthesis of macrocyclic peptidomimetic β -secretase (BACE-1) inhibitors, *Bioorg. Med. Chem. Lett.* 19 (2009) 1361-1365.
- [105] W.G. Harter, H. Albrect, K. Brady, B. Caprathe, J. Dunbar, J. Gilmore, S. Hays, C.R. Kostlan, B. Lunney, N. Walker, The design and synthesis of sulfonamides as caspase-1 inhibitors, *Bioorg. Med. Chem. Lett.* 14 (2004) 809-812.
- [106] M.L. Sznajdman, C.D. Haffner, P.R. Maloney, A. Fivush, E. Chao, D. Goreham, M.L. Sierra, C. LeGrumelec, H.E. Xu, V.G. Montana, M.H. Lambert, T.M. Willson, W.R.O. Jr, D.D. Sternbach, Novel selective small molecule agonists for peroxisome

proliferator-activated receptor delta (PPARdelta)-synthesis and biological activity, *Bioorg. Med. Chem. Lett.* 13 (2004) 1517-1521.

[107] T. Terasaka, T. Kinoshita, M. Kuno, N. Seki, K. Tanaka, I. Nakanishi, Structure-based design, synthesis, and structure-activity relationship studies of novel non-nucleoside adenosine deaminase inhibitors, *J. Med. Chem.* 47 (2004) 3730-3743.

[108] A. Takami, M. Iwakubo, Y. Okada, T. Kawata, H. Odai, N. Takahashi, K. Shindo, K. Kimura, Y. Tagami, M. Miyake, K. Fukushima, M. Inagaki, M. Amano, K. Kaibuchi, H. Iijima, Design and synthesis of Rho kinase inhibitors (I), *Bioorg. Med. Chem.* 12 (2004) 2115-2137.

[109] T. Terasaka, T. Kinoshita, M. Kuno, I. Nakanishi, A highly potent non-nucleoside adenosine deaminase inhibitor: efficient drug discovery by intentional lead hybridization, *J. Am. Chem. Soc.* 126 (2004) 34-35.

[110] T. Terasaka, H. Okumura, K. Tsuji, T. Kato, I. Nakanishi, T. Kinoshita, Y. Kato, M. Kuno, N. Seki, Y. Naoe, T. Inoue, K. Tanaka, K. Nakamura, Structure-based design and synthesis of non-nucleoside, potent, and orally bioavailable adenosine deaminase inhibitors, *J. Med. Chem.* 47 (2004) 2728-2731.

[111] A.C. Pierce, E. ter Haar, H.M. Binch, D.P. Kay, S.R. Patel, P. Li, CH...O and CH...N hydrogen bonds in ligand design: a novel quinazolin-4-ylthiazol-2-ylamine protein kinase inhibitor, *J. Med. Chem.* 48 (2005) 1278-1281.

[112] M.S. South, B.L. Case, R.S. Wood, D.E. Jones, M.J. Hayes, T.J. Girard, R.M. Lachance, N.S. Nicholson, M. Clare, A.M. Stevens, R.A. Stegeman, W.C. Stallings, R.G. Kurumbail, J.J. Parlow, Structure-based drug design of pyrazinone antithrombotics as selective inhibitors of the tissue factor VIIa complex, *Bioorg. Med. Chem. Lett.* 13 (2003) 2319-2325.

- [113] V. Stoll, K.D. Stewart, C.J. Maring, S. Muchmore, V. Giranda, Y.Y. Gu, G. Wang, Y.W. Chen, M.H. Sun, C. Zhao, A.L. Kennedy, D.L. Madigan, Y.B. Xu, A. Saldivar, W. Kati, G. Laver, T. Sowin, H.L. Sham, J. Greer, D. Kempf, Influenza neuraminidase inhibitors: structure-based design of a novel inhibitor series, *Biochemistry* 42 (2003) 718-727.
- [114] S. Venkatraman, F.G. Njoroge, V.M. Girijavallabhan, V.S. Madison, N.H. Yao, A.J. Prongay, N. Butkiewicz, J. Pichardo, Design and synthesis of depeptidized macrocyclic inhibitors of hepatitis C NS3-4A protease using structure-based drug design, *J. Med. Chem.* 48 (2005) 5088-5091.
- [115] T.M. Bergfors, Protein crystallization: techniques, strategies, and tips, International University Line, La Jolla, California, 1999.
- [116] A. McPherson, Introduction to protein crystallization, *Methods* 34 (2004) 254-265.
- [117] L. Stewart, R. Clark, C. Behnke, High-throughput crystallization and structure determination in drug discovery, *Drug Discov. Today* 7 (2002) 187-196.
- [118] B. Zheng, C.J. Gerdt, R.F. Ismagilov, Using nanoliter plugs in microfluidics to facilitate and understand protein crystallization, *Curr. Opin. Struct. Biol.* 15 (2005) 548-555.
- [119] M.H.J. Hagen, D. Frenkel, Determination of phase diagrams for the hard-core attractive Yukawa system, *J. Chem. Phys.* 101 (1994) 4093-4097.
- [120] C. Liu, N. Asherie, A. Lomakin, J. Pande, O. Ogun, G.B. Benedek, Phase separation in aqueous solutions of lens γ -crystallins: special role of γ_s , *Proc. Natl. Acad. Sci.* 93 (1996) 377-382.
- [121] P.L. San Biagio, M.U. Palma, Spinodal lines and Flory-Huggins free-energies for solutions of human hemoglobins HbS and HbA, *Biophys. J.* 60 (1991) 508-512.

- [122] N. Asherie, Protein crystallization and phase diagram, *Methods* 34 (2004) 266-272.
- [123] G. Sazaki, K. Kurihara, T. Nakada, S. Miyashita, H. Komatsu, A novel approach to the solubility measurement of protein crystals by two-beam interferometry, *J. Crystal Growth* 169 (1996) 355-360.
- [124] F. Rosenberg, S.B. Howard, J.W. Sowers, T.A. Nyce, Temperature dependence of protein solubility-determination and application to crystallization in X-ray capillaries, *J. Crystal Growth* 129 (1993) 1-12.
- [125] D.D. Dunuwila, K.A. Berglund, ATR FTIR spectroscopy for in situ measurement of supersaturation, *J. Crystal Growth* 179 (1997) 185-193.
- [126] P.A. Darcy, J.M. Wiencek, Estimating lysozyme crystallization growth rates and solubility from isothermal microcalorimetry, *Acta Cryst. D* 54 (1998) 1387-1394.
- [127] L.F. Haire, D.M. Blow, A novel spin filter method for the measurement of solubility, *J. Crystal Growth* 232 (2001) 17-20.
- [128] E. Ruckenstein, I.L. Shulgin, Effects of salts and organic additives on the solubility of proteins in aqueous solutions, *Adv. Colloid Interface Sci.* 123-126 (2006) 97-103.
- [129] R.J. Siezen, M.R. Fisch, C. Slingsby, G.B. Benedek, Opacification of γ -crystallin solutions from calf lens in relation to cold cataract formation, *Proc. Natl. Acad. Sci. USA* 82 (1985) 1701-1705.
- [130] Q. Chen, P.G. Vekilov, R.L. Nagel, R.E. Hirsch, Liquid-liquid phase separation in hemoglobins: distinct aggregation mechanisms of the $\beta 6$ mutants, *Biophys. J.* 86 (2004) 1702-1712.
- [131] O. Galkin, K. Chen, R.L. Nagel, R.E. Hirsch, P.G. Vekilov, Liquid-liquid phase separation in solutions of normal and sickle cell hemoglobin, *Proc. Natl. Acad. Sci.* 99 (2002) 8479-8483.

- [132] G.M. Thurston, Liquid-liquid phase separation and static light scattering of concentrated ternary mixtures of bovine α and γ B crystallins, *J. Chem. Phys.* 124 (2006) 134909.
- [133] J.A. Thomson, P. Schurtenberger, G.M. Thurston, G.B. Benedek, Binary liquid phase separation and critical phenomena in a protein/water solution, *Proc. Natl. Acad. Sci.* 84 (1987) 7079-7083.
- [134] M.L. Broide, C.R. Berland, J. Pande, O.O. Ogun, G.B. Benedek, Binary-liquid phase separation of lens protein solutions, *Proc. Natl. Acad. Sci.* 88 (1991) 5660-5664.
- [135] C.R. Berland, G.M. Thurston, M. Kondo, M.L. Broide, J. Pande, O. Ogun, G.B. Benedek, Solid-liquid phase boundaries of lens protein solutions, *Proc. Natl. Acad. Sci.* 89 (1992) 1214-1218.
- [136] C. Liu, A. Lomakin, G.M. Thurston, D. Hayden, A. Pande, J. Pange, O. Ogun, N. Asherie, G.B. Benedek, Phase separation in multi-component aqueous-protein solutions, *J. Phys. Chem.* 99 (1995) 454-461.
- [137] P. Katsonis, S. Brandon, P.G. Vekilov, Corresponding-states laws for protein solutions, *J. Phys. Chem. B* 110 (2006) 17638-17644.
- [138] V. Vlachy, H.W. Blanch, J.M. Prausnitz, Liquid-liquid phase separations in aqueous solutions of globular proteins, *AIChE J.* 39 (1993) 213-223.
- [139] Y.C. Chiew, D. Kuehner, H.W. Blanch, J.M. Prausnitz, Molecular thermodynamics for salt-induced protein precipitation, *AIChE J.* 41 (1995) 2150-2159.
- [140] F.W. Tavares, D. Bratko, A. Striolo, H.W. Blanch, J.M. Prausnitz, Phase behavior of aqueous solutions containing dipolar proteins from second-order perturbation theory, *J. Chem. Phys.* 120 (2004) 9859-9869.

- [141] C. Haas, J. Drenth, Understanding protein crystallization on the basis of phase diagram, *J. Crystal Growth* 196 (1999) 388-394.
- [142] D.L. Pagan, J.D. Gunton, Phase behavior of short-ranged square-well model, *J. Chem. Phys.* 122 (2005) 184515.
- [143] A. Lomakin, N. Asherie, G.B. Benedek, Monte Carlo study of phase separation in aqueous protein solutions, *J. Chem. Phys.* 104 (1996) 1646-1656.
- [144] P.M. Tessier, A.M. Lenhoff, Measurements of protein self-association as a guide to crystallization, *Curr. Opin. Biotechnol.* 14 (2003) 512-516.
- [145] B.L. Neal, D. Asthagiri, O.D. Velev, A.M. Lenhoff, E.W. Kaler, Why is the osmotic second virial coefficient related to protein crystallization?, *J. Crystal Growth* 196 (1999) 377-387.
- [146] J.J. Grigsby, H.W. Blanch, J.M. Prausnitz, Effect of secondary structure on the potential of mean force for poly-L-lysine in the α -helix and β -sheet conformations, *Biophys. Chem.* 99 (2002) 107-116.
- [147] K. Bartling, A. Sambanis, R.W. Rousseau, Dependence of Apoferritin Crystal Growth on Temperature and Cadmium Concentration, *Cryst. Growth Des.* 7 (2007) 569-575.
- [148] M. Muschol, F. Rosenberg, Interactions in undersaturated and supersaturated lysozyme solutions: static and dynamic light scattering results, *J. Chem. Phys.* 103 (1995) 10424-10432.
- [149] A.M. Kulkarni, A.P. Chatterjee, K.S. Schweizer, C.F. Zukoski, Effects of polyethylene glycol on protein interactions, *J. Chem. Phys.* 113 (2000) 9863-9873.

- [150] W. Liu, D. Bratko, J.M. Prausnitz, H.W. Blanch, Effect of alcohols on aqueous lysozyme-lysozyme interactions from static light-scattering measurements, *Biophys. Chem.* 107 (2004) 289-298.
- [151] D.J. Winzor, M. Deszczynski, S.E. Harding, P.R. Wills, Nonequivalence of second virial coefficients from sedimentation equilibrium and static light scattering studies of protein solutions, *Biophys. Chem.* 128 (2007) 46-55.
- [152] J. Bloustine, V. Berejnov, S. Fraden, Measurements of protein-protein interactions by size exclusion chromatography, *Biophys. J.* 85 (2003) 2619-2623.
- [153] W. Kadima, A. McPherson, M.F. Dunn, F. Jurnak, Precrystallization aggregation of insulin by dynamic light scattering and comparison with canavalin, *J. Crystal Growth* 110 (1991) 188-194.
- [154] M. Boyer, M.O. Roy, M. Jullien, Dynamic light scattering study of precrystallizing ribonuclease solutions, *J. Crystal Growth* 167 (1996) 212-220.
- [155] Y. Georgalis, A. Zouni, W. Saenger, Dynamics of protein precrystallization cluster formation, *J. Crystal Growth* 118 (1992) 360-364.
- [156] Y. Georgalis, W. Saenger, Time-resolved light scattering studies on protein precrystallization fractal clusters, *Adv. Colloid Interface Sci.* 46 (1993) 165-183.
- [157] Y. Georgalis, A. Zouni, W. Eberstein, W. Saenger, Formation dynamics of protein precrystallization fractal clusters, *J. Crystal Growth* 126 (1993) 245-260.
- [158] Y. Georgalis, P. Umbach, J. Raptis, W. Saenger, Lysozyme aggregation studied by light scattering. I. influence of concentration and nature of electrolytes, *Acta Crystallogr., Sect. D* 53 (1997) 691-702.

- [159] Y. Georgalis, P. Umbach, J. Raptis, W. Saenger, Lysozyme aggregation studied in light scattering. II. variations of protein concentration, *Acta Crystallogr., Sect. D* 53 (1997) 703-712.
- [160] J.B. Bishop, W.J. Fredericks, S.B. Howard, T. Sawada, Dynamic light scattering analysis of solutions from which lysozyme crystals grow, *J. Crystal Growth* 122 (1992) 41-49.
- [161] S. Tanaka, M. Yamamoto, K. Kawashima, K. Ito, R. Hayakawa, M. Ataka, Kinetic study on the early stage of the crystallization process of two forms of lysozyme crystals by photon correlation spectroscopy, *J. Crystal Growth* 168 (1996) 44-49.
- [162] W. Kadima, A. McPherson, M.F. Dunn, F. Journak, Characterization of precrystallization aggregation of canavalin by dynamic light scattering, *Biophys. J.* 57 (1990) 125-132.
- [163] V. Mikol, P. Vincendon, G. Eriani, E. Hirsch, R. Giegé, Diagnostic of protein crystallization by dynamic light scattering; an application to an aminoacyl-tRNA synthetase, *J. Crystal Growth* 110 (1991) 195-200.
- [164] A. Auersch, W. Littke, L. Pang, W. Burchard, Static and dynamic light scattering on solutions of precrystalline β -galactosidase, *J. Crystal Growth* 110 (1991) 201-207.
- [165] J.W. Mullin, *Crystallization*, 4 ed., Oxford, Butterworth Heinemann, 2001.
- [166] D. Tsekova, S. Dimitrova, C.N. Nanev, Heterogeneous nucleation (and adhesion) of lysozyme crystals, *J. Crystal Growth* 196 (1999) 226-233.
- [167] T. Kubo, Y. Uchiyama, T. Mizushima, H. Hondoh, T. Nakada, Orientation of tetragonal lysozyme crystals nucleated on fatty acid thin films, *J. Crystal Growth* 269 (2004) 535-541.

- [168] T. Kubo, Y. Uchiyama, T. Mizushima, H. Hondoh, T. Nakada, Influence of the crystallization solution on orientations of lysozyme crystals nucleated on fatty acid thin films, *J. Crystal Growth* 275 (2005) e1431-e1436.
- [169] L. Rong, H. Komatsu, I. Yoshizaki, A. Kadowaki, S. Yoda, Protein crystallization by using porous glass substrate, *J. Synchrotron Rad.* 11 (2004) 27-29.
- [170] E. Curcio, E. Fontananova, G. Di Profio, E. Drioli, Influence of the structural properties of poly(vinylidene fluoride) membranes on the heterogeneous nucleation rate of protein crystals, *J. Phys. Chem. B* 110 (2006) 12438-12445.
- [171] Y.X. Liu, X.J. Wang, J. Lu, C.B. Ching, Influence of the roughness, topography, and physicochemical properties of chemically modified surfaces on the heterogeneous nucleation of protein crystals, *J. Phys. Chem. B* 111 (2007) 13971-13978.
- [172] A. Cacciuto, S. Auer, D. Frenkel, Onset of heterogeneous crystal nucleation in colloidal suspensions, *Nature* 428 (2004) 404-406.
- [173] C. Thaller, L.H. Weaver, G. Eichele, E. Wilson, R. Karlsson, J.N. Jansonius, Repeated seeding technique for growing large single crystals of proteins, *J. Mol. Biol.* 147 (1981) 465-469.
- [174] A. Nadarajah, M.L. Pusey, Growth mechanism and morphology of tetragonal lysozyme crystals, *Acta Crystallogr., Sect. D* 52 (1996) 983-996.
- [175] M.R. Li, A. Nadarajah, M.L. Pusey, Modeling the growth rates of tetragonal lysozyme crystals, *J. Crystal Growth* 156 (1995) 121-132.
- [176] P.A. Darcy, J.M. Wiencek, Identifying nucleation temperatures for lysozyme via differential scanning calorimetry, *J. Crystal Growth* 196 (1999) 243-249.
- [177] A. Kulkarni, C. Zukoski, Depletion interactions and protein crystallization, *J. Crystal Growth* 232 (2001) 156-164.

- [178] H.C. Jr, M. Allaman, J. Wiencek, J. Kaplan, P.J. Loll, Static light scattering studies of OMPF porin: implications for integral membrane protein crystallization, *Protein Sci.* 9 (2008) 1559-1566.
- [179] P.E. Pjura, A.M. Lenhoff, S.A. Leonard, A.G. Gittis, Protein crystallization by design: chymotrypsinogen without precipitants, *J. Mol. Biol.* 300 (2000) 235-239.
- [180] S. Ruppert, S.I. Sandler, A.M. Lenhoff, Correlation between the osmotic second virial coefficient and the solubility of proteins, *Biotech. Prog.* 17 (2001) 182-187.
- [181] D.F. Rosenbaum, A. Kulkarni, S. Ramakrishnan, C.F. Zukoski, Protein interactions and phase behavior: sensitivity to the form of the pair potential, *J. Chem. Phys.* 111 (1999) 9882-9890.
- [182] A.J. Sophianopoulos, C.K. Rhodes, D.N. Holcomb, K.E. van Holde, Physical studies of lysozyme. I. characterization, *J. Biol. Chem.* 237 (1962) 1107-1112.
- [183] Y. Park, G. Choi, Effects of pH, salt type and ionic strength on the second virial coefficients of aqueous bovine serum albumin solutions, *Korean, J. Chem. Eng.* 26 (2009) 193-198.
- [184] S. Finet, F. Skouri-Panet, M. Casselyn, F. Bonneté, A. Tardieu, The Hofmeister effect as seen by SAXS in protein solutions, *Curr. Opin. Colloid Interface Sci.* 9 (2004) 112-116.
- [185] M.C.R. Heijna, W.J.P. van Enkevort, E. Vlieg, Growth inhibition of protein crystals: a study of lysozyme polymorphs, *Cryst. Growth Des.* 8 (2008) 270-274.
- [186] P.M. Aker, J.X. Zhang, W. Nichols, Nitrate ion detection in aerosols using morphology-dependent stimulated Raman scattering, *J. Chem. Phys.* 110 (1999) 2202-2207.

- [187] M.L. Broide, T.M. Tominc, M.D. Saxowsky, Using phase transitions to investigate the effect of salts on protein interactions, *Phys. Rev. E* 53 (1996) 6325-6335.
- [188] K. Demoruelle, B. Guo, S. Kao, H.M. McDonald, D.B. Nikic, S.C. Holman, W.W. Wilson, Correlation between the osmotic second virial coefficient and solubility for equine serum albumin and ovalbumin, *Acta Crystallogr., Sect. D* 58 (2002) 1544-1548.
- [189] C. Haas, J. Drenth, The protein-water phase diagram and the growth of protein crystals from aqueous solution, *J. Phys. Chem. B* 102 (1998) 4226-4232.
- [190] P.M. Tessier, V.J. Verruto, S.I. Sandler, A.M. Lenhoff, Correlation of diafiltration sieving behavior of lysozyme-BSA mixtures with osmotic second virial cross-coefficients, *Biotech. Bioeng.* 87 (2004) 303-310.
- [191] D.N. Petsev, X.X. Wu, O. Galkin, P.G. Vekilov, Thermodynamic functions of concentrated protein solutions from phase equilibria, *J. Phys. Chem. B* 107 (2003) 3921-3926.
- [192] R.M.L. Evans, W.C.K. Poon, M.E. Cates, Role of metastable states in phase ordering dynamics, *Europhys. Lett.* 38 (1997) 595-600.
- [193] M.J. Grimson, Small-angle scattering from colloidal dispersions, *J. chem. Soc., Faraday Trans.* 79 (1983) 817-832.
- [194] J.P. Hansen, I.R. McDonald, *Physics of simple liquids*, Academic Press, London, 1986.
- [195] N.F. Carnahan, K.E. Starling, Thermodynamic Properties of a Rigid-Sphere Fluid, *J. Chem. Phys.* 53 (1970) 600-603.
- [196] A. Lomakin, N. Asherie, G.B. Benedek, Aeolotopic interactions of globular proteins, *Proc. Natl. Acad. Sci.* 96 (1999) 9465-9468.

- [197] A.M. Kulkarni, C.F. Zukoski, Nanoparticle crystal nucleation: influence of solution conditions, *Langmuir* 18 (2002) 3090-3099.
- [198] V. Bhamidi, S. Varanasi, C.A. Schall, Protein crystal nucleation: is the pair interaction potential the primary determinant of kinetics?, *Langmuir* 21 (2005) 9044-9050.
- [199] R.P. Sear, Classical nucleation theory for the nucleation of the solid phase of spherical particles with a short-ranged attraction, *J. Chem. Phys.* 111 (2001) 2001-2007.
- [200] J. Crank, *The Mathematics of Diffusion*, 2 ed., Oxford University Press, New York, 1975.
- [201] S. Gorti, E.L. Forsythe, M.L. Pusey, Kinetic roughening and energetics of tetragonal lysozyme crystal growth, *Cryst. Growth Des.* 4 (2004) 691-699.
- [202] H.L. Cui, Y. Yu, W.C. Chen, Q. Kang, Study of growth mechanism of lysozyme crystal by batch crystallization method, *Chin. Chem. Lett.* 17 (2006) 101-104.
- [203] P.G. Vekilov, Dense liquid precursor for the nucleation of ordered solid phases from solution, *Cryst. Growth Des.* 4 (2004) 671-685.
- [204] S. Gorti, J. Konnert, E.L. Forsythe, M.L. Pusey, Effects of kinetic roughening and liquid-liquid phase transition on lysozyme crystal growth velocities, *Cryst. Growth Des.* 5 (2005) 535-545.
- [205] Y.G. Kuznetsov, A.J. Malkin, A. McPherson, The liquid protein phase in crystallization: a case study-intact immunoglobulins, *J. Crystal Growth* 232 (2001) 30-39.
- [206] L. Wu, MSc Thesis, Mississippi State University, 1997.
- [207] M.L. Pusey, A. Nadarajah, A model for tetragonal lysozyme crystal nucleation and growth, *Cryst. Growth Des.* 2 (2002) 475-483.
- [208] V. Talanquer, D.W. Oxtoby, Crystal nucleation in the presence of a metastable critical point, *J. Chem. Phys.* 109 (1998) 223-227.

- [209] M.C.R. Heijna, F.F.M. van Wamel, W.J.P. van Enkevort, E. Vlieg, PEG-induced morphologically unstable growth of tetragonal hen egg-white lysozyme crystals, *Cryst. Growth Des.* 7 (2007) 1999-2008.
- [210] S.J. Bonafede, M.D. Ward, Selective nucleation and growth of an organic polymorph by ledge-directed epitaxy on a molecular crystal substrate, *J. Am. Chem. Soc.* 117 (1995) 7853-7861.
- [211] A.K. Panwar, S.K. Barthwal, S. Ray, Effect of evaporation on the contact angle of a sessile drop on solid substrates, *J. Adhesion Sci. technol.* 17 (2003) 1321-1329.
- [212] G. Wolansky, A. Marmur, Apparent contact angles on rough surfaces: the Wenzel equation revisited, *Colloid and Surf. A: Physicochem. Eng. Asp.* 156 (1999) 381-388.
- [213] T. Matsushita, K. Nakano, Y. Nishikura, K. Higuchi, A. Kiyota, R. Ueoka, Spheroid formation and functional restoration of human fetal hepatocytes on poly-amino acid-coated dishes after serial proliferation, *Cytotechnology* 42 (2003) 57-66.
- [214] C.B. Walsh, E.I. Franses, Ultrathin PMMA films spin-coated from toluene solutions, *Thin Solid Films* 429 (2003) 71-76.
- [215] G. McHale, S.M. Rowan, M.I. Newton, M.K. Banerjee, Evaporation and the wetting of a low-energy solid surface, *J. Phys. Chem. B* 102 (1998) 1964-1967.
- [216] H.Y. Erbil, R.A. Meric, Evaporation of sessile drops on polymer surfaces: ellipsoidal cap geometry, *J. Phys. Chem. B* 101 (1997) 6867-6873.
- [217] M. Malmstein, Formation of adsorbed protein layers, *J. Colloid Interface Sci.* 207 (1988) 186-199.
- [218] J.J. Ramsden, Puzzles and paradoxes in protein adsorption, *Chem. Soc. Rev.* 24 (1995) 73-78.

- [219] W. Norde, Adsorption of proteins from solution at the solid-liquid interface, *Adv. Colloid Interface Sci.* 25 (1986) 267-340.
- [220] L. Palacio, J.I. Calvo, P. Prádanos, A. Hernández, P. Väisänen, M. Nyström, Contact angles and external protein adsorption onto UF membranes, *J. Membr. Sci.* 152 (1999) 189-201.
- [221] S.J. Fang, S. Haplepete, W. Chen, C.R. Helms, H. Edwards, Analyzing atomic force microscopy images using spectral methods, *J. Appl. Phys.* 82 (1997) 5891-5898.
- [222] D.T. Kim, H.W. Blanch, C.J. Radke, Direct imaging of lysozyme adsorption onto mica by atomic force microscopy, *Langmuir* 18 (2002) 5841-5850.
- [223] S.T. Yau, P.G. Vekilov, Quasi-planar nucleus structure in apoferritin crystallization, *Nature* 406 (2000) 494-497.
- [224] K.Q. Zhang, X.Y. Liu, In situ observation of colloidal monolayer nucleation driven by an alternating electric field, *Nature* 429 (2004) 739-742.
- [225] D. Tsekova, S. Popova, C. Nanev, Nucleation rate determination by a concentration pulse technique: application on ferritin crystals to show the effect of surface treatment of a substrate, *Acta Crystallogr., Sect. D* 58 (2002) 1588-1592.
- [226] N. Yasumoto, N. Kasahara, A. Sakaki, M. Satoh, Ion-specific swelling behaviors of partially quaternized poly(4-vinyl pyridine) gel, *Colloid Polym. Sci.* 284 (2006) 900-908.
- [227] V.A. Kabanov, V.P. Evdakov, M.I. Mustafaev, A.D. Antipina, Cooperative interaction of serum albumin with quaternized poly-4-vinyl pyridine and structure of the complexes, *Mol. Biol. (Mosk)* (1977) 582-597.

9 List of Publications

Yingxin Liu, Xiujuan Wang and Chi Bun Ching, Towards Further Understanding of Protein Crystallization: Phase Diagram, Protein-Protein Interaction, Nucleation Kinetics and Growth Kinetics, **Crystal Growth & Design**, 10, (2010), 548 - 558.

Xiujuan Wang, Xia Yang, Yingxin Liu and Chi Bun Ching, Intrinsic MSZW Characteristics of Racemic Species: Implication for Chiral Crystallization, **AIChE Journal**, 54, (2008), 2281- 2292.

Ying-Xin Liu, Xiu-Juan Wang, Jie Lu, Chi-Bun Ching, Influence of the Roughness, Topography and Physicochemical Properties of Chemically Modified Surfaces on the Heterogeneous Nucleation of Protein Crystals, **The Journal of Physical Chemistry B**, 111, (2007), 13971-13978.

J. Lu, X.-J. Wang, Y.-X. Liu and C.-B. Ching, Thermal and FTIR Investigation of Freeze-dried Protein-excipient Mixtures, **Journal of Thermal Analysis and Calorimetry**, 89, (2007), 913-919.

Conference Presentations

Yingxin Liu, Xiujuan Wang and Chi Bun Ching, Systematic Investigation of the Nucleation Mechanism of Lysozyme Crystallization in the Liquid-liquid Co-existence Region, AIChE Annual Meeting 2008, Philadelphia, USA.

Yingxin Liu, Xiujuan Wang, Jie Lu and Chi Bun Ching, Effects of Ionic Strength and Liquid-liquid Phase Separation on the Nucleation of Protein Crystals, AICHE Annual Meeting 2007, Salt Lake City, USA.