Understanding the self-assembly mechanism of E2 protein cage and exploring its potential applications

Tao, Peng

2013

Tao, P. (2013). Understanding the self-assembly mechanism of E2 protein cage and exploring its potential applications. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/52237>

<https://doi.org/10.32657/10356/52237>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**UNDERSTANDING THE SELF-ASSEMBLY
MECHANISM OF E2 PROTEIN CAGE AND
EXPLORING ITS POTENTIAL APPLICATIONS**

PENG TAO

SCHOOL OF CHEMICAL AND BIOMEDICAL ENGINEERING

2013

**UNDERSTANDING THE SELF-ASSEMBLY
MECHANISM OF E2 PROTEIN CAGE AND
EXPLORING ITS POTENTIAL APPLICATIONS**

PENG TAO

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

2013

Chapter 2 © 2011 American Chemical Society

Chapter 3 © 2012 American Chemical Society

All other materials © 2012 Peng Tao

Acknowledgements

I want to dedicate some words to express my gratitude to those who helped me make this work possible. I would like to thank my supervisor Prof. Sierin Lim for her guidance and generous support through the years. The active lab environment and enormous encouragement from her helped me have a joyful Ph.D. life.

Many thanks to Prof. Hwankyu Lee, Dr. Nikodem Tomczak, and Dr. David Paramelle for their helps on problem solving and technical discussions.

My sincere appreciations to Prof. Luo Qian, Kathy, Prof. Susanna Leong, Prof. Jiang Rongrong, Prof. Chen Hongyu, Dr. Wang Xiujuan, Dr. Yu Shucong for their generous helps on instrument facilities and technical supports.

I would like to thank Mr. Aung Pyae, Miss. Chitra Devi D/O Subramaniam, and Miss. Yeo Kah Yan for their helps on facilities and consumable purchase.

Thanks to the past and present members of our BeANs lab for all their help and being good friends. Special thanks to Dr. Barindra Sana, Dr. Li Yan, Miss. Yu Kang, and Miss. Herlina Arianita Dewi for valuable discussions and the happy time we spent together.

I would like to thank my friends, Yu Ting, Yang Tianyi, Yuan Jifeng, Feng Xuesong, Chen Xue, Dong Jing, Wang Liang, Chen Yu, Zhou Yusi, Yu Yaolun, Yu Haiyang, Zhong Jidan, etc. for their friendships.

Thanks to School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore for providing me the opportunity and scholarship to pursue the Ph.D. degree.

Finally, I am deeply indebted to my family. My dear parents and sister are giving me their love and supports all the time. Special appreciations to my wife Dr. He Pengfei, for her selfless understanding and supports all these years. My Ph.D. study became easier for having her standing with me.

Curriculum Vitae

Peng Tao

Education:

2008 - Now	Ph.D in Bioengineering
School of Chemical and Biomedical Engineering,
Nanyang Technological University, Singapore
2003 - 2007	B.E. in Bioengineering
School of Life Science and Technology,
Xi'an Jiaotong University, China

Publications:

1. Peng T, Lim S (2011), Trimer-Based Design of pH-Responsive Protein Cage Results in Soluble Disassembled Structures, *Biomacromolecules* 12(19):3131-3138.
2. Peng T, Lee H, Lim S (2012), Isolating a Trimer Intermediate in the Self-assembly of E2 Protein Cage, *Biomacromolecules* 13(3):699–705.
3. Qiu H, Dong X, Sana B, Peng T, Chen P, Lim S (2013) Ferritin-templated synthesis and self-assembly of Pt nanoparticles on monolithic porous graphene network for electrocatalysis in fuel cell, *ACS Applied Materials and Interfaces* 5(3):782-787.
4. Li Y, Toyip RO, Peng T, Lim S. (2011). Encapsulation and Release Profile of Protein Cage from a Polymeric Matrix. *Nano LIFE*, 2(1):1250001.

Patent:

- Paramelle D, Tomczak N, Free P, Lim S, Peng T. Specific Internalization of Nanoparticles into Protein Cages.

Conference Proceedings:

Peng T, Tan SW, Dharmawan RE, Lim S. "Investigating the influence of ionic concentrations and subunit interactions on the self-assembly of E2 protein." In AIP Conference Proceedings, vol. 1502, p. 34. 2012.

Conference Presentations:

1. Peng T, Lim S, Modifying Protein Cage for Controlled Release Applications, 2012 World Congress- Medical Physics and Biomedical Engineering, 26-31th, June, 2012, Beijing, China.
2. Peng T, Lim S, Probing the Self-Assembly Mechanism of E2 Protein, International Conference on Nanotechnology - Research and Commercialization (ICONT 2011), 6-9th, Jun, 2011, Sabah, Malaysia.
3. Peng T, Lim S, Probing Important Interactions in the Self-Assembly of E2 Protein Cage and its Potential Application, BES 5th scientific meeting, 28th, May, 2011, Singapore.
4. Peng T, Tan S, Dharmawan R, Tay T, Lim S, Determining self-assembly mechanism of a protein nanocage, International Conference on Cellular & Molecular Bioengineering (ICCMB2), 2-4th, Aug, 2010, Singapore.
5. Peng T, Tan S, Dharmawan R, Tay T, Lim S, Probing the Self-assembly Mechanism of E2 Core Protein, 1st Nano Today Conference, 2-5th, Aug, 2009, Singapore.

Summary

Dihydrolipoamide acetyltransferase (E2 protein) is part of pyruvate dehydrogenase multi-enzyme complex from *Bacillus stearothermophilus*. However, we don't consider its enzymatic activity in this work. E2 protein used in this work is a protein cage derived and is composed of 60 subunits. Structurally, the 60 subunits self-assemble to form a 25-nm hollow caged structure. Previous works have demonstrated its potential application as nanocapsule in drug delivery to encapsulate small molecules in its inner cavity, i.e. cancer drugs.

The researches that we have been done in this work are summarized as followings:

To better control the release of the encapsulated drugs at target sites during drug delivery, modulating the self-assembly process of E2 protein under different environmental stimuli is desirable. Therefore, understanding of the self-assembly mechanism of E2 protein is required. By truncating the C-terminal α -helix of E2 protein, the association between trimeric structures are eliminated and the self-assembly was halted at the trimer intermediate state. Molecular dynamic simulation reaffirms that the E2 protein adopts trimer intermediate prior to its fully assembled caged structure and that the C-terminus plays a critical role in mediating the self-assembly from monomer to trimer and to 60-mer.

Upon identifying the importance of the C-terminus in the trimer and subsequently the cage assembly, other functionalities were introduced to enhance the release of molecular cargos (e.g. cancer drugs) from the cavity of the E2 protein cage. To this goal, amino acid histidines are introduced at critical sites at the interfaces between trimer structures. Histidine has a unique property that it is charged at acidic pH while

remains uncharged at neutral and basic pH-s. The proximity of two or more histidines at the subunit interfaces induces repulsion and hence their separation at acidic pH, subsequently leading to the soluble and non-denatured disassembly of the E2 protein.

Incorporation of a unique GALA peptide (peptide with repeating amino acids: Glu-Ala-Leu-Ala) with pH-sensitive helix-to-coil transition at low and high pH, respectively, to substitute the original C-terminal α -helix induces an interesting assembly characteristic. As a result, the self-assembly of E2 protein become pH-responsive and present in inversed direction. At neutral or high pH-s, the extended random coil state of GALA peptide dissociates the E2 protein cage. As the pH is lowered, E2 protein re-assembled due to the formation of α -helix from GALA peptide. Moreover, the pH-inducible self-assembly of this engineered E2 protein is reversible.

The well-defined inner cavity provides the possibility to utilize the E2 protein cage as a nanoreactor for inorganic nanoparticle synthesis for potential application as an imaging contrast agent. Iron-binding peptides were incorporated to the interior surface of E2 protein without jeopardizing its self-assembly and caged structure. The engineered proteins are able to nucleate irons. The subsequent oxidation of the iron molecules in the constrained cavities of the E2 protein results in the formation of uniform nanoparticles.

The results in this study suggest that E2 protein is expandable beyond its native function to serve as a controllable multi-display platform in biomedical as well as other nanotechnological applications.

Table of Contents

ACKNOWLEDGEMENTS	I
CURRICULUM VITAE.....	III
SUMMARY	V
LIST OF FIGURES	XI
LIST OF TABLES	XV
LIST OF ACRONYMS AND ABBREVIATIONS	XVI
CHAPTER 1	1
INTRODUCTION.....	1
1.1 PROTEIN CAGES	2
1.2 SELF-ASSEMBLY OF PROTEIN CAGES	4
1.3 OVERVIEW OF DRUG DELIVERY SYSTEMS	6
<i>1.3.1 Controlled Drug Release.....</i>	<i>8</i>
1.4 PROTEIN-BASED DRUG DELIVERY SYSTEMS	11
<i>1.4.1 Non-Caged Protein as DDS.....</i>	<i>12</i>
<i>1.4.2 Protein cage as DDS</i>	<i>13</i>
1.5 PROTEIN CAGE AS A TEMPLATE FOR NANOPARTICLE SYNTHESIS	15
1.6 THE MODEL PROTEIN CAGE: E2 PROTEIN	19
1.7 STATEMENT OF PURPOSES	24
CHAPTER 2.....	27
MATERIALS AND METHODS	27
2.1 MATERIALS.....	28
2.2 GENE EXPRESSION AND PROTEIN PURIFICATION	28
2.3 MOLECULAR MASS DETERMINATION	30
2.4 DYNAMIC LIGHT SCATTERING	30
2.5 CIRCULAR DICHROISM.....	31
2.6 TRANSMISSION ELECTRON MICROSCOPY	31
2.7 SIZE EXCLUSION CHROMATOGRAPHY	32
2.8 PROTEIN CROSS-LINKING	32

2.9 MOLECULAR DYNAMIC SIMULATION.....	33
2.10 IRON MINERALIZATION AND CHARACTERIZATIONS	34
2.11 STABILITY EVALUATION OF IRON MINERALIZATION	34
CHAPTER 3.....	35
TRIMER-BASED DESIGN OF PH-RESPONSIVE PROTEIN CAGE RESULTS IN SOLUBLE DISASSEMBLED STRUCTURES.....	35
3.1 ABSTRACT	36
3.2 INTRODUCTION	37
3.3 RESULTS AND DISCUSSION	40
3.3.1 <i>Design and Construction of pH-Sensitive Mutant Proteins.....</i>	40
3.3.2 <i>Mutant Proteins Show Correct Assemblies</i>	42
3.3.3 <i>The Secondary Structures of the Mutant Proteins are Altered at Physiological pH.....</i>	44
3.3.4 <i>Intra-Trimer Modified Cage Retains its Correct Assembled Structure at pH 5.0</i>	46
3.3.5 <i>Inter-Trimer Modified Cages Show Aggregations at pH 5.0</i>	48
3.3.6 <i>Disassembly from Inter-Trimer Interface does not Denature E2 Subunits and is Irreversible</i>	49
3.3.7 <i>Cross-Linking Verifies Non-Denatured Disassembly and Suggests Trimer as Building Block.....</i>	51
3.4 CONCLUSIONS.....	54
CHAPTER 4.....	55
ISOLATING A TRIMER INTERMEDIATE IN THE SELF-ASSEMBLY OF E2 PROTEIN CAGE.....	55
4.1 ABSTRACT	56
4.2 INTRODUCTION	57
4.3 RESULTS AND DISCUSSIONS.....	62
4.3.1 <i>Design and Construction of Mutant Protein</i>	62
4.3.2 <i>E2-ΔC9 is Present as Both Monomer and Trimer</i>	63
4.3.3 <i>E2-ΔC9 Shows Dynamic Transitions Between Monomer and Trimer</i>	67
4.3.4 <i>Molecular Dynamics Simulations Support the Importance of Interactions between Trimers.....</i>	70

4.3.5 C-Terminus Mediates the Self-Assembly from Trimer to 60-mer.....	72
4.4 CONCLUSIONS	75
CHAPTER 5	77
DESIGN OF REVERSIBLE INVERSED PH-RESPONSIVE E2 PROTEIN	
CAGE	77
5.1 ABSTRACT	78
5.2 INTRODUCTION	79
5.3 RESULTS AND DISCUSSIONS	83
5.3.1 Construction of GALA Incorporated E2 Protein	83
5.3.2 The Incorporation of GALA Affects E2 Assembly at pH 7.0.....	84
5.3.3 pH-Responsive Self-Assembly of E2-GALA.....	87
5.3.4 E2-GALA Self-assembles at pH 4.0	89
5.3.5 Reversible pH-responsive self-assembly.....	92
5.4 CONCLUSIONS	94
CHAPTER 6	95
DESIGNING NON-NATIVE IRON-BINDING SITE ON E2 PROTEIN CAGE	
.....	95
6.1 ABSTRACT	96
6.2 INTRODUCTION	97
6.3 RESULTS AND DISCUSSIONS	100
6.3.1 Design and Construction of Ferritin-like Catalytic Domain in E2 Protein	
.....	100
6.3.2 E2 Proteins Assembled Correctly upon Incorporation of Iron-Binding	
Peptides.....	101
6.3.3 Iron Mineralization does not affect the Protein Structures	102
6.3.4 Iron Mineralization within E2 Protein Cages.....	105
6.3.5 Mutant E2 Proteins Show High Iron-Binding Capacities	107
6.4 CONCLUSION	109
CHAPTER 7	111
CONCLUSIONS AND FUTURE DIRECTIONS.....	111
7.1 CONCLUSIONS.....	112

7.2 FUTURE DIRECTIONS	114
REFERENCES.....	117
APPENDICES	139
A.1 CROSS-LINKING OF E2 PROTEIN.....	140
A.2 DENATURANT EFFECTS ON E2 PROTEIN SELF-ASSEMBLY	142
A.3 IRON MINERALIZATION SUPPORTING CHARACTERIZATIONS	145
A.4 SITE-DIRECTED MUTAGENESIS PROTOCOL	147
A.5 SDS POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE).....	149
A.6 NEGATIVE STAINING OF TEM SAMPLES:	151

List of Figures

Figure 1.1	Three well-defined surfaces of a protein cage.	3
Figure 1.2	Commonly used VLPs in drug delivery research.	4
Figure 1.3	Schematic of the cell endocytosis during intracellular drug delivery.	10
Figure 1.4	Model for active-site coupling in a hypothetical E1E2E3 complex.	20
Figure 1.5	Schematic of pH-mediated disassembly of N-terminus truncated E2 caged protein.	22
Figure 3.1	Molecular structures highlighting identified key amino acids and relevant protein domains.	39
Figure 3.2	SDS-PAGE showing correctly produced and purified proteins. (1) E2-WT, (2) E2-(2+2)H, (3) E2-4H.	43
Figure 3.3	Electron micrographs showing the structures of wild type and mutants E2 at physiological pH 7.4. All proteins presented correctly assembled spherical structures.	44
Figure 3.4	Far-UV circular dichroism showing molar ellipticity versus wavelength for E2-WT, E2-(2+2)H, and E2-4H at pH 7.4.	46
Figure 3.5	Electron micrographs of E2-WT and mutants at pH 5.0.	47
Figure 3.6	Comparisons of far-UV circular dichroism spectra of	48

mutant E2 proteins at pH 7.4 and 5.0.

Figure 3.7	Representative DLS measurements showing the hydrodynamic diameters of E2-4H at pH 7.4 (blue) and 5.0 (red).	49
Figure 3.8	SDS-PAGE and electron micrographs of E2-WT and mutants E2-(2+2)H and E2-4H after cross-linking.	53
Figure 4.1	Molecular structures highlighting the truncated C-terminal α -helix of E2 protein.	61
Figure 4.2	SEC profiles of E2- Δ C9 at different concentrations.	64
Figure 4.3	SDS-PAGE analysis of (1) E2-WT, (2) trimer fraction, and (3) monomer fraction of E2- Δ C9.	64
Figure 4.4	Representative DLS scan shows the hydrodynamic diameter of about 8 nm and 25 nm for trimer (blue) and wild-type 60-mer (red) structures of E2 protein, respectively.	65
Figure 4.5	CD spectra of E2-WT and fractions of E2- Δ C9 from SEC.	66
Figure 4.6	Relations between trimer percentages and concentrations of (A) E2- Δ C9 and (B) Redistributed trimer fraction based on the integrated signals of SEC profiles in Figure 3.2 and Figure 3.7, respectively.	68
Figure 4.7	SEC profiles of collected trimer fractions at different concentrations.	69

Figure 4.8	Snapshots at the beginning (0ns, top) and the end (20ns, bottom) of simulations of (A) E2-WT, (B) E2- Δ C9, and (C) E2-5H.	71
Figure 4.9	Secondary structures of E2-WT (top) and E2- Δ C9 (bottom) as a function of time during the simulation process.	71
Figure 4.10	Transmission electron micrographs of (A) E2-WT and (B) E2-5H showing correct assembled structures at pH 7.4.	73
Figure 4.11	Snapshot of the interactions between two E2-WT monomers at intertrimer interface.	74
Figure 5.1	Molecular structures highlighting the substitution of C-terminal α -helix in E2 protein with GALA peptide.	82
Figure 5.2	SEC profiles showing the oligomeric states of E2-GALA at different pH-s.	85
Figure 5.3	SDS-PAGE verifies the E2 protein compositions of E2-GALA at different pH-s. E2-WT is used as a control.	85
Figure 5.4	CD spectra of E2-GALA at different pH-s. E2-GALA show characteristic profiles of α -helix rich protein with minima at 208 and 222 nm at all pH-s.	87
Figure 5.5	Representative DLS scans show the hydrodynamic diameters of E2-GALA at different pH-s.	89
Figure 5.6	Electron micrographs of E2-GALA at different pH-s.	91
Figure 5.7	Reversibility analysis of E2-GALA.	93

Figure 6.1	PyMol representation of E2 protein cage for iron mineralization.	98
Figure 6.2	Correct assemblies of E2-LFer and E2-LE6 before iron-loading indicated by DLS and TEM.	102
Figure 6.3	Colors of E2 protein solutions before and after iron-loading.	103
Figure 6.4	Comparisons of DLS profiles for (A) E2-LFer and (B) E2-LE6 before and after iron-loading; Correlograms are shown on the right.	104
Figure 6.5	Electron micrographs of iron-mineralized mutant E2 protein cages.	105
Figure 6.6	SEC profiles indicating co-elutions of protein cages and mineralized iron cores: (A) E2-LFer, and (B) E2-LE6.	107
Figure A.1.1	Crystallographic structure of trimers and highlighted lysine residues.	140
Figure A.2.1	The effect of denaturants on hydrodynamic diameters of E2-WT.	143
Figure A.2.2	SEC profiles of E2-WT at different concentrations of GuHCl.	144
Figure A.3.1	UV-vis absorbance scans of E2-LFer and E2-LE6 solutions.	145
Figure A.3.2	PyMol representing the locations of RDGE loops around the pore on E2 protein.	146

List of Tables

Table 3.1	Oligonucleotides for mutant plasmids construction. Mutation sites are in bold, genes are in uppercase, and restriction enzyme sites are underlined.	42
Table 3.2	Hydrodynamic diameters (in nm) of WT and mutant E2 proteins in sodium phosphate buffer at pH 7.4 and 5.0.	44

List of Acronyms and Abbreviations

BSA	Bovine serum albumin
CCMV	Cowpea chlorotic mottle virus
CD	Circular dichroism
CLP	Caged-like proteins
CPMV	Cowpea mosaic virus
CPV	Canine parvovirus
DDS	Drug delivery system
DLS	Dynamic light scattering
E1	Pyruvate decarboxylase
E3	Dihydrolipoamide dehydrogenase
E2	Dihydrolipoamide acetyltransferase
E2-WT	Wild-type E2
E2-(2+2)H	E2 protein with 4 Histidine clusters at intra-trimer interface
E2-4H	E2 protein with 4 histidine pairs at inter-trimer interface
E2-5H	E2 protein with 5 histidine pairs at inter-trimer interface
E2-ΔN	E2 protein with N-terminus truncation
E2-ΔC9	E2 protein with 9 amino acids truncated at C-terminus
E2-GALA	E2 protein incorporated with GALA peptide at C-terminus
E2-LFer	E2 protein incorporated with ferritin-mimicking iron-binding peptide
E2-LE6	E2 protein incorporated with 6 glutamic acids
EDTA	Ethylene Diamine Tetraacetic Acid
EGF	Epidermal growth factor

ELP	Elastin-like-protein
EPR effect	Enhanced vascular permeability and retention effect
GALA	Synthetic amino acid repeats of Glu-Ala-Leu-Ala
GuHCl	Guanidine hydrochloride
HIV	Human immunodeficiency virus
IPTG	Isopropyl β -D-thiogalactopyranoside
LIP	Lipoyl domain
MALDI-TOF MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MDS	Molecular dynamic simulations
OD	Optical density
PDH	Pyruvate dehydrogenase
pE2-WT	Plasmid pET-11a inserted with wild-type E2 gene
PEG	Polyethylene glycol
PSBD	Peripheral subunit binding domain
RDV	Rice dwarf virus
RES	Reticuloendothelial system
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SLP	Silk-like-protein
STIV	Sulfolobus turreted icosahedral virus
TEM	Transmission electron microscopy
V _e	Elution volume
VLP	Virus-like particles
VMD	Visual molecular dynamics

Chapter 1

Introduction

1.1 Protein Cages

Protein cages are protein-based hierarchical architectures which self-assemble from defined number of subunits into hollow spherical structures. Protein cages are derived from living organisms and play crucial biological functions. Viruses use their capsids to recognize specific host cell as well as to protect and deliver viral nucleic genome to the host cells.^{1, 2} Ferritins serve to store and control the release of irons in a physiological system.^{3, 4} Heat shock proteins work as molecular chaperones to protect other proteins from denaturation.⁵ Interestingly, despite their unique biological functions, all protein cages have similarities in structure which is folded and assembled from numbers of subunits to form spherically symmetrical structures. The complexity and self-assembly nature of the protein cage make three well-defined surfaces available for modifications: interior surface, exterior surface, and the interfaces between subunits (Figure 1.1).⁶

Figure 1.1 Three well-defined surfaces of a protein cage.⁶ (A) Cryo-reconstruction of the *Sulfolobus turreted* icosahedral virus (STIV). (B) Schematic illustration of the three interfaces available in a protein cage. Both figures are used to highlight the structural features of protein cage architectures that can be chemically and genetically modified to introduce therapeutic functionalities.

The ferritins, a family of protein cages, play a key role in iron sequestration in living organisms.^{3, 4} The first ferritin was isolated from horse spleen, and then its crystal structure was determined.⁷ Since then, ferritins from diverse organisms including animals, plants, and bacteria have been isolated and crystallized.⁸⁻¹⁰ While the primary amino acid sequences of ferritins from different organisms show little homology, the structural homology (at the 2°, 3°, and 4° levels) is highly conserved.¹¹ Structurally, ferritins are composed of 24 homologous subunits to self-assemble into a spherical hollow cage with octahedral symmetry. The outer diameter of the cage is about 12 nm, and the diameter of the inner cage is about 8 nm.¹²

Viral capsids are the structural protein shells of the viruses. Typically, virus capsids are self-assembled from repeating subunits to form highly symmetrical and homogenous architectures.¹³ Virus capsids can occur in a range of shapes and sizes, from 18-500 nm for icosahedral structure and > 2 μm in length for filamentous or rod-

shaped viruses.¹⁴ Their varied sizes within nanometer level make the VLPs has potential as nanoscale platforms. After removal of genetic materials, the only viral protein shells left are referred to as virus-like particles (VLP). The most understood VLPs at the moment include cowpea mosaic virus (CPMV),^{15, 16} cowpea chlorotic mottle virus (CCMV),^{17, 18} canine parvovirus (CPV),¹⁹ and MS2 bacteriophage²⁰ (Figure 1.2).

Figure 1.2 Commonly used VLPs in drug delivery research. Their average diameters are: CPMV: 30 nm, CCMV: 27.8 nm, CPV: 25.4 nm, and MS2: 26.8 nm. Virus structure models were obtained from <http://viperdb.scripps.edu>.

1.2 Self-assembly of Protein Cages

The native conformation of a protein is determined by its amino acids sequence, with combination of the solvent, the salts concentration, and the temperature.^{21, 22} Interactions, including hydrophobic interactions, dipole-dipole interactions, ion-dipole interactions, and ion-ion interactions, within proteins and between amino acids and the mediums determine the assembly and conformation of protein complex.²³

Resolving the self-assembly mechanisms of protein supramolecular complexes is challenging especially the structures are highly symmetrical and homo-oligomeric, such as protein cages.¹² Protein cages are always believed to adopt assembly-intermediate in their self-assembly process. However, most intermediates are transient. Formation of an intermediate requires collective binding events that are regulated by balanced competitions of interactions between individual subunits.²⁴ It is difficult to experimentally analyze the interactions that are critical to the self-assembly process, thus further understand the self-assembly mechanism. Therefore, most investigations are focused on modifying the interactions between subunits, thus further understand the self-assembly processes.

Many works have been done on investigating the self-assembly of ferritins. Each ferritin subunit is made up of a four-helix (ABCD) bundle with a short fifth helix (E) at the C-terminus. The role of E helix plays in ferritin self-assembly is protein specific. Removal of the E helix from bacterioferritin (BFR) will lead to destabilized protein that can only assemble into a dimer.²⁵ However, another research proposes that human ferritin H-chain can assemble into a cage with the E helix flipped out or inside the inner cavity.²⁶ Based on this statement, Choi et al. substitute the E-helix with artificial GALA peptide, which has pH-sensitive coil-to-helix transition, to control the self-assembly of human ferritin.²⁷ Since the spatial arrangements between subunits are influenced by conformational change of GALA peptide, the engineered ferritin self-assembles at neutral pH, but disassembles at acidic pH. In another hand, the N-terminus points outside of the protein cage and has little subunit-subunit overlap in most ferritins. Therefore, N-terminus deletion human H-chain ferritin still self-assembles into protein cage and catalyses iron oxidation.²⁸

The interactions between and within the subunits provide important contributions to the formation of VLPs. In some virus, the N-terminus of the subunit is involved in modulating the self-assembly of the capsids.²⁹⁻³¹ Wild-type CCMV has a T=3 capsids with 180 subunits arranged as 90 dimers.²⁹ The N-terminus of CCMV is required as switch for mediating the T=3 self-assembly. The removal of N-terminus results in mutant CCMV capsids with different self-assembly manners: T=3 capsids of 90 dimers, T=2 capsids of 60 dimers, and T=1 capsids of 30 dimers.²⁹ In another virus tomato bushy stunt virus (TBSV), the removal of the N-terminus affects the interactions between subunits and result in smaller capsids size.³¹ Similar characteristic is also observed in virus-like particles (VLP) such as the E2 protein in this investigation,³² the dihydrolipoyl acetyltransferase component of pyruvate dehydrogenase from *Bacillus stearothermophilus*. The removal of N-terminus imparts no influence on self-assembly of E2 protein at physiological pH. However, the intra-trimer interfaces are exposed without N-terminus protection. The acidic solvent accesses to the interface and triggers the disassembly of E2 protein at pH 5.0.^{32, 33}

Investigation on self-assembly mechanism of protein cages set fundamentals to understand protein folding and protein-protein interactions. Moreover, it is instrumental in the protein engineering rational design to form more stable and functional non-native nanostructures, with the ultimate goal to utilize them as vehicles for delivery systems and in novel templated materials synthesis.

1.3 Overview of Drug Delivery Systems

The current drug delivery system (DDS) for chemotherapy spread drugs via the circulatory system which leads to some potential limitations, such as low drug

efficiencies as a result of rapid clearance from the blood stream,³⁴ or undesired side effects due to unnecessary reactions or over-accumulation of drugs at inappropriate sites.³⁵ Hence, it is necessary to develop other means to improve the drug delivery efficiency to protect the patients from overdose and side effects.

An ideal DDS is suggested to be able to overcome drug intractable insolubility, protect drugs from degradation, direct drugs to target sites, and provide controlled drug release profiles.³⁶ DDS are aimed to maximize the therapeutic potential and to minimize potential immunogenicity. The characteristics of biocompatibility, biodegradability, biodistribution, therapeutic cargo capacity, and the ability to incorporate other functionalities which may facilitate targeted delivery or controlled release manner of drug delivery systems have to be considered for future applications. The design and development of the ideal DDS will require interdisciplinary approaches that combine pharmaceuticals, polymer science, chemistry, and biotechnology.

Since the 1980s, nanotechnology has been emerging as areas of scientific and health researches, and has attracted much attention. A wide range of research fields, such as electronics, material science, information technology, and biomedical engineering, have been developing prosperously with applied nanotechnology. Among these, an important application is the development of nanoscale DDS. To minimize drug loss and degradation, and to prevent harmful side effects, DDS are designed to encapsulate drug within nanoscale capsules.

The size of DDS is an important parameter to directly influence drug delivery efficacy. Microparticles can be easily captured and cleared from the vascular circulation by the lymphatic system. In contrast, DDS with nanoscale sizes in a range

less than 50 nm can escape from lymphatic system and be distributed in the extracellular and intracellular spaces. The nanoscale DDS can efficiently reach cancer tissues which usually contain more porous blood vessels compared to normal tissues. Besides considerable surface-to-volume ratios that allow more functional modifications with larger drug-loading capabilities, the unique narrow size distributions may lead to more uniform and predictable responses. The inherent advantages make nanoscale DDS attractive for further investigations, while their disadvantages can be addressed through different approaches. Different diseases may require tissue-specific or patient-specific drug delivery strategies. More researches are focused on developing new DDS with specific features and advantages to be utilized in particular disease applications.

Based on the composition of polymeric materials, nanoscale DDS can be categorized into non-protein-based and protein-based. Non-protein-based DDS include lipid based (e.g. micelle and liposome) and polymer-based (e.g. PLGA nanosphere and dendrimer) while protein-based DDS include monomeric protein (e.g. BSA) and protein cages.

1.3.1 Controlled Drug Release

To achieve maximum therapeutic efficiency, the drugs should be released in a controlled manner to the precise sites of activity within the body. Most drug molecules need to be dissolved in aqueous environment so that they can freely diffuse before acting on their target receptors. Controlled-release DDS protect drug molecules from the aqueous environment for a programmed period of time. The potential

advantages of controlled release can be accounted to: (i) localized release at target cell types to avoid adverse side effects on healthy cells, (ii) decreased drug dose and number of dosage, (iii) maintained drug concentration within a desired level by controlling the drug loading.³⁷ Depending on different compositional materials, DDS can be imparted with different drug release profiles. To achieve controlled release, DDS are usually designed to respond to specific stimuli. The commonly used stimuli include magnetic signals, electrical signals, ultrasonic signals, pH changes, or temperature changes.³⁸

Many potential therapeutic treatments require drugs to be delivered into cytosolic or nuclear compartment of cells. During a typical pathway for intracellular drug delivery, drug carrier will be taken in through cell endocytosis, macropinocytosis, or phagocytosis, and then target to lysosomes, where the pH is lowered to 5.0³⁹ (Figure 1.1). Furthermore, hypoxic condition of cancerous tissue also results in slightly acidic conditions within the surrounding area.⁴⁰ Therefore, other than targeting to desired cells, in order to control the burst release of drugs in lysosome or within the vicinity of cancerous cells, the DDS with pH-sensitive close-to-open mechanism will be favourable in this particular drug delivery pathway.

Figure 1.3 Schematic of the cell endocytosis for intracellular drug delivery. Drug carrier encounters physiological pH 7.4 at extracellular matrix, which subsequently acidifies at pH 5.0 in lysosome. Drugs will be released in lysosome and finally translocated or diffuse to cytosol.

To control the release of drugs in response to cellular stimuli when drug is delivered through endocytic intracellular pathway which is coupled to pH change, many works have been done on introducing pH-responsive triggers to DDS. For example, cross-linked poly(DEAEMA-co-PEGMA) hydrogel was imparted with pH-responsive swelling behaviour to deliver drugs into dendritic cells.⁴¹ Proteins are also suitable platforms to engineer pH-responsive property due to their defined subunit interfaces and highly modifiable amino acids residues. Various genetic modifications have been introduced to caged-like protein-based DDS, consequently, the open-close behaviour will be controlled by pH change.^{27, 32, 33, 42} Designing pH-responsive self-assembly of E2 protein cage is one of the main objective in this thesis.

1.4 Protein-Based Drug Delivery Systems

There are many on-going studies focused on non-protein-based DDS, including micelles,^{43, 44} liposomes,⁴⁵⁻⁴⁷ dendrimers,⁴⁸⁻⁵⁰ nanospheres,^{51, 52} and hydrogels.^{53, 54} Some of them have made great advancements towards clinical application, for example, due to its favourable systemic response (i.e. biodegradability and biocompatibility), poly(lactic-co-glycolic acid) (PLGA) has been approved as therapeutic device by FDA. Protein polymers have been developed and evaluated for use as drug and gene delivery systems for over the last two decades.⁵⁵ Naturally existing proteins or synthetic proteins have been used as templates to adsorb or encapsulate therapeutic agents to form DDS. The precise chemical composition with specific and adjustable properties makes protein-based DDS promising to overcome the drawbacks of synthetic polymeric DDS, including non-uniformity and questionable biodegradability.⁵⁶

Proteins consist of repeating amino acid sequences which can be produced by genetic engineering and biological synthesis. The composition of amino acids provides the protein-based DDS with high biodegradability that is they can be degraded to peptides or amino acids through natural pathways. Encoding at gene level allows for precise control of amino acid residues, which serve as the basis to obtain accurate structures and functional features of proteins. Hydrophobicity, secondary structures, biorecognizable motifs can be modified accordingly to enhance drug conjugation and interactions with protein-based DDS.⁵⁷ Cellular synthesis of protein peptides follows the gene sequences with high fidelity, thus generating monodispersed distribution of protein products. Monodispersity of protein-based DDS is crucial in producing distinct pharmacokinetic profiles.⁵⁵ Further genetic engineering also allows

incorporation of desired functionalities in the form of inherent gene expressions on the protein polymers. Moreover, the modifiable amino acids can be chemically linked with multi-functional ligands, implying the versatility of protein-based DDS to treat different diseases.

A variety of proteins, such as bovine serum albumin (BSA), elastin, silk-like protein, gelatin, ferritin, and viral capsids, have been investigated as drug carrier. A notable protein-based formulation is Abraxane, an FDA approved commercialized injectable albumin-bound paclitaxel indicated for breast cancer therapy.^{58, 59}

Proteins that can be artificially manipulated to form different shapes for drug-loading and delivery purposes, include caged-structures, microspheres, hydrogels, films, and minirods, etc.⁵⁶

1.4.1 Non-Caged Protein as DDS

Two types of non-caged proteins used as DDS, elastin-like protein (ELP) and silk-like-protein (SLP), are represented in this thesis.

The pentapeptide sequence derived from the elastomeric domain of mammalian tropoelastin is the basic studied unit for drug delivery from ELP polymers.⁶⁰ Various ELPs form micellar structures for drug delivery usage upon temperature-sensitive coacervation.^{61, 62} Hydrophobic drug molecules can be conjugated to the hydrophobic cores of the ELP micelles. Alternatively, by changing the process and ultimate multiblock composition, ELPs can also be developed to form hydrogel, microspheres, and films as drug delivery platforms.⁶² The foreign peptides with additional functions

can be fused to the N- and C- terminus of ELP platforms to enhance drug-loading capacity or impart cell targeting ability.⁶³

Silks are naturally produced proteins with outstanding properties to be used as DDS. The recombinantly modified SLP seek to improve the special characteristics such as solubility, mechanical strength, and biodegradability⁶⁴. To date, the shapes of porous hydrogels, microparticles, tube-like structures, electrospon fibers, and thin films of self-assembled SLP have been studied as drug delivery materials⁶⁵. These reformed mechanically robust materials remain excellent in their biocompatibilities and biodegradabilities.⁶⁴ To further increase the actual delivery and transfection efficiency, cell-specific targeting groups have been introduced into the backbones of SLP.⁶⁶ For example, by incorporating repeating RGD sequence, significantly enhanced transfection efficiency of the SLP scaffold is observed.

1.4.2 Protein cage as DDS

The well-defined structures and uniform sizes of protein cages make them attractive in protein-based DDS research.^{56, 67-69} Without influencing the whole architecture of the protein cages, biological or chemical functionalities can be introduced at these interfaces to impart the protein cages with drug delivery potentials. Functional amino acids, such as cysteine and lysine, can be genetically incorporated into the protein surfaces for cross-linking of drug molecules, fluorescence labels, or imaging agents.⁷⁰⁻⁷³ Many works have proposed that the protein cages, such as virus capsids,⁷⁴⁻⁷⁷ heat shock proteins,^{73, 78} and ferritin^{68, 79, 80} can potentially be used as drug carriers for biomedical therapeutics.

Protein cages have confined inner cavities which allow for the loading of precise amount of therapeutic drugs, and prevention of random macromolecules aggregations. The uniform sizes (between 5-100 nm) of protein cages present advantages during drug spreading in the circulatory systems. Because of the high amount of growth factors, cancerous tissues have much richer blood capillaries than that of healthy tissues.⁸¹ The sizes of protein cages are much larger than the size of pores on normal blood capillaries, so that the protein cages remain in the circulatory system until they encounter cancer tissues, and rapid leakage of protein cages into healthy tissues can be avoided. On the other hand, the sizes of the protein cages can help them escape from the 'body-guard system' in liver and spleen, and facilitate the delivery to cancer sites.⁸² Furthermore, targeting ligands or phage-displayed peptides can be genetically or enzymatically modified to the exterior surface to actively target protein cage carriers to the desired cells or tissues.⁸³⁻⁸⁶ By masking the recognition surface through PEGylation or using human proteins, the effects of immune system on protein cages can be reduced.⁸⁷

In the circulatory system, the surrounding integrate shells of protein cages can protect the encapsulated drugs from undesired reactions. The assembly and disassembly of the protein cages can be controlled to load drug dynamically and to release the drugs in response to different cellular cues. Appropriate modifications made at critical sites, such as the interfaces between subunits of protein cages, will result in altered assembly and disassembly behaviours. For example, by substituting the motif located at the subunit interface with GALA peptide, the assembly of ferritin becomes pH-responsive.²⁷ The pH-responsive drug carriers can potentially be used in particular delivery pathways which will experience pH change, such as endocytosis in intracellular drug delivery.

One type of the most studied protein cages as drug carriers are VLPs. The icosahedral VLPs are stable enough to maintain their integrities even under harsh conditions, such pH values ranging from 3.5 to 9.0.⁸⁸ The amino acids on the interior and exterior surfaces of VLPs can be genetically replaced with specific functional groups, such as lysine with $-\text{NH}_2$ or cysteine with $-\text{SH}$ to be displayed for particular binding applications. Therefore, drugs or cell targeting ligands can be incorporated into the VLPs but without influencing their original structures and functions. Moreover, the rigid capsids with even spatial distribution minimize the problems associated with drug molecule aggregation.⁸⁹ By controlling the opening and closing of the pores on CCMV using pH change, substrates entry to the cage can be controlled. The size and shape of the entrapped materials are determined by the size and morphology of CCMV cage.¹⁷

1.5 Protein Cage as a Template for Nanoparticle Synthesis

Other than drug delivery purpose, the protein cages are increasingly being used as multivalent and multifunctional nanocontainers to synthesize inorganic nanoparticles. The three distinct surfaces (interior, exterior, and interface) can be used to generate nanoparticles with multiple functionalities. The extremely homogeneous size distribution of protein cages can be used to attain high degree of homogeneity of the template materials and their associated properties.

Ferritins are iron storage protein found in most living systems (animal, plant, and bacterial).³ They are composed of 24 structurally identical subunits that self-assemble into caged-like structures with inner cavities about 8 nm. The catalytic sites in size-

constrained cavities, which are composed of conserved and spatially arranged acidic residues (Glu and Asp), allow ferritins to work as reaction vessels to catalyse Fe(II) ions to form Fe(III) oxide nanoparticles.^{90, 91} The protein shells of ferritins act as passivating layers to prevent unwanted particle-particle interactions. When ferrous irons are allowed to undergo oxidation in vitro in the absence of ferritins, uncontrolled homogeneous nucleations result in mineralization and precipitation of iron oxides.

DNA-binding proteins (Dps) belong to ferritin superfamily, and composed of 12 subunits to form caged-like structures. Researchers found that Dps can also be used as size-constrained reaction vessel for iron mineralization under the non-physiological condition of elevated temperature and pH (65 °C, pH 8.5) and trace amount of oxidant (H₂O₂).⁹² Dps mineralize irons more efficiently in the presence of H₂O₂ as compared to ferritins that catalyse irons with O₂.⁹³ Therefore, Dps are thought to serve as antioxidants that protect the organism during oxidative stress.

Other than serving as iron storage protein, ferritin architectures have been utilized for the synthesis of CoPt, Pd, Ag, FePt, ZnSe, and CdS nanoparticles.^{92, 94-99} The synthesis are achieved by controlling the environmental parameters, including pH, temperature, and ionic strength. The synthesized nanoparticles had uniform narrow size distributions which were restricted by the dimension of the ferritin inner cavities. Because of possession of negatively charged interior surface, two types of cobalt oxide minerals (Co₃O₄ and Co(O)OH) have been synthesized in size-constrained vessel of Dps.⁹⁴

In addition to utilizing the inherent properties of the protein cage to synthesize iron nanoparticles in ferritins and Dps, other protein cages can be genetically modified to introduce functional peptides at defined locations of N- and C- terminus or within the

loop structure of the subunit. The most typical and widely investigated protein cages for nanoparticle synthesis are virus-like particles. After removal of negatively charged nucleic acids, VLPs present positively charged interior surfaces, which provide interfaces for inorganic crystal nucleation and growth.¹⁷ For example, triggered by complementary interactions between positively charged CCMV interior surface and anionic metals, two polyoxometalate species (Na_2WO_4 and NaVO_3) have been grown and crystallized.¹⁷

The electrostatic status of VLPs can be altered by genetic modifications. Inspired by iron mineralization capability of ferritin, which is commonly thought due to the presence of multiple glutamic acids on its interior surface, researchers have engineered other Glu-rich protein cages to impart iron mineralization capability.^{96, 100} Eight of the positively charged amino acids at the N-terminus of CCMV subunits are replaced by negatively charged Glu. The engineered CCMV is remaining caged structure while presenting negatively charged interior surface. The interactions of CCMV with cationic Fe(II) ions which promoted by the negative interior surface lead to the formations of 24 nm γ -FeOOH nanoparticles.

The exterior surface of protein cage can also be used as template for nanoparticle patterning.¹⁰¹⁻¹⁰⁴ CPMV is engineered to display sulfhydryl group on its exterior surface, which allow controlled gold nanoparticle attachments.¹⁰¹ By introducing cysteines onto its surface, CPMV can be used to attach iron nanoparticles on its exterior surface.¹⁰⁴

Both interior and exterior of protein cages can be modified as templates for metal nanoparticles. Metal-containing protein cages have potential biomedical applications. Iron nanoparticle-housing protein cages allow enhanced magnetic signals, and can

potentially be applied in imaging, such as serving as MRI contrast agents, and in cancer treatment by hyperthermia.^{92, 97}

1.6 The Model Protein Cage: E2 Protein

Protein cages have been shown to be very useful as multifunctional nanoscale platforms. There are many works focusing on investigating the encapsulation capabilities of different protein cages. However, limited works have reported the controlled release of encapsulated drug at target sites. Hence, the overall objectives of this Ph.D. work is to understand and control the self-assembly process of a model protein cage to serve as the basis for stimuli-controlled drug release application, as well as to expand the protein cage potential to other applications.

The model protein cage of this research is the E2 subunit of pyruvate dehydrogenase (PDH) multi-enzyme complex from *Bacillus stearothermophilus*. Functionally, PDH takes part in catalysing the synthesis of acetyl-CoA from pyruvate, linking glycolysis to the tricarboxylic acid cycle and the biosynthesis of fatty acids.¹⁰⁵⁻¹⁰⁷ Structurally, PDH comprises multiple copies of three subunit enzymes, which includes: pyruvate decarboxylase (E1), dihydrolipoamide acetyltransferase (E2), and dihydrolipoamide dehydrogenase (E3). The complexes are assembled around E2 subunits, to which the E1 and E3 domains are bound tightly (Figure 1.4).^{108, 109} Depending on the species, the E2 component has two different symmetries: octahedral and icosahedral. In Gram-negative bacteria, E2 consist 24 subunits to form the caged structure with octahedral 432 symmetry. In eukaryotes and Gram-positive bacteria, 60 subunits assemble into E2 component with icosahedral 532 symmetry.¹¹⁰ The detailed structure of PDH with 60-mer E2 component has been elucidated using 3-D crystal diffraction.^{108, 109, 111}

Figure 1.4 Model for active-site coupling in a hypothetical E1E2E3 complex.¹¹¹ Three E1 tetramers (purple) and three E3 dimers (yellow) are shown located in the outer protein shell above the inner icosahedron (gray) formed by 60 E2 catalytic domains. Image reproduced from Milne *et al.*; Molecular Structure of a 9-kDa Icosahedral Pyruvate Dehydrogenase Subcomplex Containing the E2 and E3 Enzymes Using Cryoelectron Microscopy.

The *B. stearothermophilus* icosahedral E2 core protein consists of 60 copies of identical polypeptide chains. Each polypeptide chain comprises of 3 mechanistic regions: an N- terminal lipoyl domain (LIP, ~ 80 residues in length, depending on the species), the peripheral subunit binding domain (PSBD, ~ 35 residues), and a C-terminal catalytic (acetyltransferase) domain (~250 residues). Each of these regions is linked to one another by highly flexible linker peptides approximately 25- 30 residues long, rich in alanine, proline and charged amino acids.^{112, 113} The main function of the LIP and PSBD regions in a typical icosahedral PDH complex is to bind E1 and E3 dimers to the E2 core, whereas the CD regions assemble to form the pentagonal dodecahedral core and is the active site of acyltransferase.^{109, 111}

Recently, the E2 core protein containing only the CD region (260 amino acids) has been engineered, while the PSBD and LIP has been removed.^{114, 115} Sixty copies of

the CD regions self-assemble into the engineered E2 core protein with a dodecahedral hollow spherical structure. The E2 protein cage has an outer diameter of 25 nm and an inner diameter of 12 nm. There are 12 openings with 5 nm each on the cage.^{108, 109}

Because of some unique properties, we are interested in the potential applications of the E2 protein as nanocapsules in drug delivery research. The E2 protein has uniform size with constrained inner cavity, which is an ideal container for drug accumulation. The 12 opening pores on the surface will facilitate drug diffusion into the cavity. The self-assembled 60 subunits of the E2 protein dodecahedral core provide precisely defined exterior and interior surface. By chemical or genetic methods, non-natural functionalities can be introduced onto the surfaces of E2 protein without influencing the correct assembly of protein cage.^{114, 116} As a result, facilitated drug adsorption or targeted drug delivery can be achieved. The E2 protein is derived from a thermophilic organism, so the association of 60 subunits is quite stable at elevated temperature.¹¹⁷ Moreover, the E2 protein has been reported to resist proteolytic degradation in serum *in vivo* and was able to penetrate cells.^{114, 116} The highly robust E2 protein cage allows for its flexible modifications.

Potential applications of the E2 protein as a nanoscale delivery system have been investigated. Dalmau *et al.* engineered a truncated codon-optimized E2 protein gene and heterologously over-expressed it in *Escherichia coli* system.¹¹⁵ The recombinant E2 protein correctly self-assembled into a dodecahedral spherical 60-mer structure. Since E2 protein composed of 60 subunits, modifying one subunit will result in 60 identical modifications on the whole protein cage. Sixty cysteines with 60 exposed thiol groups were incorporated onto the interior surface of E2 protein. The guest molecules (e.g. fluorescent dye and anticancer drug doxorubicin) were observed to interact with the thiol groups and accumulate at the inner cavity. Subsequent reports

have shown that drug molecules encapsulated in the E2 protein retains its efficacy.^{115,}

118

E2 protein cage could be internalized into cancer cells through endocytosis pathway, where the drug carrier will experience physiological pH 7.4 and lysosomal pH 5.0 in succession.¹¹⁸ Therefore, E2 protein with pH-responsive disassembly profile will be favourable for fast drug release form the inner cavity. By truncating the N-terminus of E2 protein subunits, the first pH-responsive protein cage was engineered.^{32, 33} The engineered E2 protein cage remained self-assembled into a dodecahedral structure at pH 7.4. When the pH was lowered to 5.0, the protein cage irreversibly disassembled and aggregated (Figure 1.5). However, the aggregation and denaturation of protein cage might affect drug release.

Figure 1.5 Schematic model of pH-mediated disassembly of N-terminus truncated E2 caged protein.³² At pH 7.4, E2 protein self-assembled with icosahedral symmetry, and at pH 5.0 they irreversibly disassemble and aggregate.

The exterior and interior surfaces of E2 protein are highly modifiable with non-native functionalities. By controlling the binding of PEG to the surface, the internalization rate of E2 protein cages into cancer cells could be controlled.¹¹⁹ The

drug loading capacity of E2 protein cage could also be adjusted by controlling the drug binding sites at the interior surface.¹²⁰ These properties make E2 protein promising for various drug delivery applications.

To further control and tailor the drug release profile, the understanding of the E2 protein self-assembly mechanism is of importance. By modifying the critical interactions at subunit interfaces, the association and dissociation of E2 protein can potentially be controlled in a programmed manner. Moreover, to facilitate drug delivery to different pathological cells, specific cell targeting or cell penetrating ligands can be introduced into the E2 protein cage. The remarkable stability of the caged structure allows simultaneous modifications to impart multi-functionalities on the E2 protein.

1.7 Statement of Purposes

The overall objectives of my Ph.D. studies are focused on understanding the self-assembly mechanism of E2 protein cage, as well as to explore its potential application as a template for nanoparticle synthesis for future development of dual-function carrier.

E2 protein cage has potential as protein-based DDS to deliver drugs through the endocytic pathway. The specific pH change along the pathway where the DDS travels demands protein cage with pH-responsive disassembly profiles that remains soluble. Protein denaturation has been reported to potentially cause necrosis.¹²¹ In order to design a better multi-functional DDS using E2 protein, understanding its self-assembly mechanism is essentially required.

The objectives listed in the dissertation are as follows:

(1) Chapter 3: To design a soluble pH-responsive E2 protein cage, the critical interactions at trimer-trimer interfaces were examined. Histidines were incorporated at the critical sites on the inter-trimer interface to trigger soluble and non-denatured disassembly at acidic pH.

(2) Chapter 4: To understand the self-assembly mechanism of E2 protein, we halted its self-assembly process at trimer intermediate by truncating the C-terminus. The C-terminus was proven to mediate the self-assembly using trimer structure as assembly intermediate.

(3) Chapter 5: On the basis of understanding the important role of the C-terminus described in the previous chapters, a functional pH-sensitive GALA peptide was

incorporated to substitute the last alpha-helix motif at the C-terminus. As a result, a reversible inversed pH-responsive protein cage was engineered.

(4) Chapter 6: E2 protein is a potential multi-functional protein cage. Iron-binding peptides were incorporated into the interior surface of E2 protein. The recombinant proteins were utilized as constrained nanoreactors to biomineralize irons.

The conclusion of the work and future direction of the project are described in Chapter 7.

Chapter 2

Materials and Methods

This Chapter describe the materials and methods used in the thesis.

2.1 Materials

The wild-type E2 sequence containing plasmid (pE2) is a generous gift from Prof. Szu-Wen Wang at University of California, Irvine. *E. coli* strains DH5 α (Zymo Research, Orange, CA) and BL21(DE3)C+RIL (Stratagene, La Jolla, CA) were used as host cells. The vector pET-11a was purchased from Novagen. The oligonucleotides were synthesized by 1st BASE (Singapore). Restriction enzymes (BamH I and Nde I), T4 DNA ligase, and Pfu Ultra High-Fidelity DNA polymerase, and isopropyl β -D-thiogalactopyranoside (IPTG) were obtained from Fermentas. Buffer reagent Tris and sodium phosphate were purchased from USB Corporation. Ethylene diamine tetraacetic acid (EDTA) and sodium azide were from Fluka. Cross-linking agent, glutaraldehyde, was get from Sigma Aldrich.

2.2 Gene Expression and Protein Purification

Wild-type (E2-WT) and mutant E2 proteins were produced in *E. coli* strain BL21(DE3)C+RIL adapting previously reported protocol.¹¹⁵ The *E. coli* cells were cultured in Luria-Bertani medium supplemented with 100 μ g/ml ampicillin and 50 μ g/ml chloramphenicol. Gene expression was induced by 1 mM IPTG at an optical density (OD₆₀₀) of 0.6-0.8. Cells were harvested 3 h after induction by centrifugation at 4,500 $\times$ g for 20 min and stored at -80°C. The cells were resuspended in Tris buffer

(20 mM Tris, 5 mM EDTA, and 0.02% sodium azide, pH 8.7), disrupted in a French pressure cell at 16,000 psi (Thermo Scientific), and centrifuged at 40,000×g for 1 h to remove the insoluble fraction.

To purify the protein, the supernatant of E2-WT, E2-4H (Chapter 3), E2-5H (Chapter 4), and E2-LE6 and E2-LFer (Chapter 6) from ultrasonication were heated at 72 °C for 20 min, and the native, denatured *E.coli* protein aggregates were removed by ultracentrifugation. Due to different thermostabilities, E2-ΔC9 (Chapter 4) and E2-GALA (Chapter 5) were not performed heat treatment.

The supernatant after ultracentrifugation was filtered, and then loaded onto an anion exchange chromatography column (HiPrep Q 16/10 Q FF, GE Healthcare), which had been equilibrated with Tris buffer (pH 8.7) on a liquid chromatography system (ÄKTA, GE Healthcare). The E2 protein was eluted with Tris buffer containing 1 M NaCl at elution concentration gradient set over 5 column volumes. Fractions containing the E2 protein were pooled and concentrated using ultrafiltration membrane with 100 kDa molecular weight cut-off (PBHK, Millipore) and buffer exchanged to sodium phosphate buffer, pH 7.4 (50 mM sodium phosphate, 150 mM sodium chloride, 5 mM EDTA, and 0.02% sodium azide).

To further purify the E2-ΔC9 and E2-GALA, size exclusion chromatography was performed. The concentrated E2-ΔC9 was subjected to SEC column (Superdex 200 10/300 GL, GE Healthcare), which had been equilibrated with sodium phosphate buffer. The fractions were collected and evaluated by SDS-PAGE and MALDI-TOF/TOF.

The purified protein was stored at -80°C. The concentration of E2-WT and mutant proteins were determined by using a Micro BCA Protein Assay Kit (Pierce) with bovine serum albumin (BSA) as a standard.

2.3 Molecular Mass Determination

SDS-PAGE analysis on pre-cast Bis-Tris gels (NuPAGE Novex, Invitrogen) and MALDI-TOF/TOF analysis (4800, Applied Biosystems) were adopted to determine the molecular mass of a single E2 subunit. Purified E2-WT and mutant proteins were dialyzed with deionized (DI) water and mixed with sinapinic acid as matrix prior to the MALDI-TOF/TOF analysis. The measurement was performed in positive and reflector mode.

2.4 Dynamic Light Scattering

Previous work reported that fully assembled E2 protein has representative diameter of 24-28 nm.^{32, 33, 115} Dynamic light scattering (DLS; Zetasizer nano ZS, Malvern) was performed to investigate the hydrodynamic diameters of E2 proteins. Protein (1 mg/ml) was centrifuged at 14,000 $\times g$ for 10 min before the measurement. Result was an average of three scans.

2.5 Circular Dichroism

Far-UV circular dichroism (CD) was used to evaluate the secondary structure change and to investigate the unfolding of mutant proteins. E2-WT and correctly assembled truncated mutant E2 protein gave the characteristic CD profile of an α -helix rich protein with minima at 208 and 222 nm.¹¹⁵ The crystallographic structure revealed that approximately one-third of the E2 protein secondary structure was α -helix (PDB file 1b5s). CD scans were performed on a CD spectrometer (Chirascan, Applied Photophysics). Protein samples (0.5 mg/ml) in sodium phosphate buffer were scanned from 200 nm to 260 nm at 25 °C in 1 mm path length quartz cells (Hellma Analytics). Results are an average of at least three scans. To understand the trend of secondary structure change, the secondary structure contents were determined by deconvolution algorithm of the CD spectra using CDNN program.¹²²

2.6 Transmission Electron Microscopy

The structure of mutant protein assemblies at pH 7.4 were confirmed by transmission electron microscopy (TEM). Protein samples (0.1 mg/ml) were stained for 3 min with 1.5% uranyl acetate on carbon-coated copper electron microscopy grids (Ted Pella, Inc.), and images were obtained with a JEOL JEM-1400 transmission electron microscope with working voltage of 120 KV.

2.7 Size Exclusion Chromatography

To evaluate the oligomeric state of E2 protein, size exclusion chromatography (SEC) was performed. E2 protein (100 μ l) was loaded to SEC column (Superdex 200 10/300 GL, GE Healthcare) pre-equilibrated with sodium phosphate buffer. oligomeric states of E2 proteins were analyzed from the elution volume (V_e). Protein elution profiles on SEC were monitored by measuring the absorbance at 280 nm. The standard protein curve was prepared with Blue Dextran, Ferritin (M_r 440 kDa), Aldolase (M_r 158 kDa), Conalbumin (M_r 75 kDa), and Carbonic Anhydrase (M_r 29 kDa) which were eluted at V_e = 9.2 ml, 11.8 ml, 13.6 ml, 14.8 ml, and 17.2 ml, respectively.

2.8 Protein Cross-Linking

In order to investigate the disassembly at intra-trimer or inter-trimer interfaces, cross-linking agent, glutaraldehyde, was used to create stable interactions among E2 subunits and assembled E2 protein cage (Chapter 3). Glutaraldehyde, which has a spacer arm of 5 Å,¹²³ is known as a common agent used in protein cross-linking reactions. It can specifically react with primary amine groups, such as ϵ -amino group of lysine, and form stable covalent bonds. Freshly prepared 2.3% (w/v) glutaraldehyde (50 μ l) was added to 1 ml of WT or mutant protein (0.5 mg/ml) in sodium phosphate buffer, pH 7.4. The reaction was conducted at 37 °C for 5 min, and terminated by addition of 100 μ l of 1 M Tris-HCl, pH 8.0. Cross-linked proteins were dialyzed to remove excess glutaraldehyde and to adjust the pH to 5.0.

2.9 Molecular Dynamic Simulation

To support the C-terminus mediated self-assembly (Chapter 4), simulations and analyses were performed with the GROMACS4.5.1 simulation package¹²⁴⁻¹²⁶ with the GROMOS53a6 force field.¹²⁷ The protein structure and coordinates were downloaded from the Protein Data Bank (PDB code: 1B5S). Since the downloaded structure consists of only backbones (alanine), the appropriate coordinates of the side chains for E2-WT (184-425 amino acids) and E2- Δ C9 (184-418 amino acids) were generated using Swiss-Pdb Viewer,¹²⁸ and then several steps of energy minimization were performed with position restraints applied to the backbone. Hydrogen atoms of the protein were fixed by defining an additional bond of appropriate length between the hydrogen atom and the linked atom, which allows the time step to be increased to 4 fs.¹²⁹ The SPC model was used for water. Two monomers (E2-WT or E2- Δ C9) from neighboring trimers were initially clustered together in a box of size 14 nm/side to mimic the interactions between trimers in forming the assembled 60-mer structure. The initial distance between two monomers was set close to 4 nm. These clustered proteins were solvated with ~89,000 water molecules. Counter ions of 6 Cl⁻ for E2-WT and 8 Cl⁻ for E2- Δ C9 were added to yield electroneutrality. The temperature was maintained at 310K by applying a Berendsen thermostat in an NPT ensemble.¹³⁰ A cutoff of 14 Å was used for the Lennard-Jones (LJ) potential. For the Coulomb potential, a short-range interaction with a cutoff of 10 Å and a long-range interaction with particle mesh Ewald summation (PME)¹³¹ were used. Simulations were performed for 20 ns.

2.10 Iron Mineralization and Characterizations

To test the iron mineralization within mutant protein cages (Chapter 6), freshly prepared ferrous sulfate solution in 0.1% HCl (100 mM) was dropwise added to each of the protein solutions (E2-LFer or E2-LE6, 0.2 mg/ml), and incubated for 1 h at room temperature, then followed by overnight incubation at 4 °C. E2-WT was included as a control. The iron loadings were carried out in HEPES buffer supplemented with 50 mM NaCl, pH 8.0, with loading ratio of 3000 Fe/60-mer E2 cage. Precipitation started to appear upon loading of more than 3000 irons to each E2 cage, indicating the iron-binding capacity of E2 cage was lower than 3000 irons. The binding reaction was carried out at room temperature (25 °C) for 1 h and followed by overnight incubation at 4 °C. Unbound irons were removed by buffer exchange using dialysis tubing (Sigma-Aldrich), followed by desalting column (PD-10, GE Healthcare). The amount of encapsulated iron for each mutant E2 protein was quantified using inductively coupled plasma (ICP). The formed iron nanoparticles in each E2 mutant were examined using TEM without negative staining.

2.11 Stability Evaluation of Iron Mineralization

To evaluate the stabilities of mineralized iron cores within E2-LFer and E2-LE6 against time, the contents of iron in E2-LFer or E2-LE6 at different time intervals were measured (Chapter 6). The mineralized protein samples were kept at 4 °C. Samples at day 1, day 7, and day 14 from the same batch were analysed by ICP and protein BCA kit. Before each measurement, possible unbound iron and precipitate were removed by buffer exchange using dialysis tubing, followed by desalting column.

Chapter 3

Trimer-Based Design of pH-Responsive Protein Cage

Results in Soluble Disassembled Structures

This chapter is a modified version of the previously published work.

Reprinted with permission from ‘Tao Peng, Sierin Lim; *Trimer-Based Design of pH-Responsive Protein Cage Results in Soluble Disassembled Structures*, *Biomacromolecules*, **2011**, 12(19):3131-3138.’

COPYRIGHT © 2011 AMERICAN CHEMICAL SOCIETY

3.1 Abstract

Limited studies have been done on the interactions between subunits of self-assembling protein cages. E2 protein cage from *B. stearothermophilus* was investigated in this work to impart pH-sensitive disassembly profile. Key amino acids were identified at the intra- and inter-trimer interfaces, and histidine residues were introduced to these key sites to probe for their influences on the E2 assembly. We found that both the intra-trimer and the inter-trimer modified mutant proteins have the same quaternary structures as the wild type (E2-WT) at physiological pH of 7.4. At pH 5.0, the intra-trimer modified protein maintained its spherical structure. In contrast, the inter-trimer modified protein lost its integrity as observed under the electron microscope while remained soluble and non-denatured. The identified interactions between the inter-trimers are critical in the formation of E2 protein cage. The pH-controlled disassembly of E2 protein cage in soluble and non-denatured form make it promising in nanoscale applications, especially for drug delivery and release in the endosomes.

3.2 Introduction

Nano-sized drug carriers may be internalized by cellular endocytosis. During endocytosis, the carriers encounter physiological pH of 7.4 in the extracellular environment and subsequently experience an acidic environment inside lysosome (pH 5.0).³⁹ Therefore, drug carriers with pH-sensitive characteristics are favorable for intracellular drug delivery. In order to trigger release of encapsulated macromolecules at acidic pH, protonated imidazole group at interfaces of multimeric polymers have been previously adopted.^{132, 133} The side chain of amino acid histidine (His) contains an imidazole group and has a pKa value of 6.5.¹³⁴ Protonation of the imidazole group generates repulsive interactions among multiple histidines when subjected to acidic pH. E2 protein has recently been reported to be endocytosed by breast cancer cells.¹¹⁸ E2 protein with pH-sensitive disassembly profile will facilitate drug release in this delivery pathway. Previous work showed that partial deletion of the E2 N-termini exposed His trimer clusters and triggered disassembly of the E2 protein by repulsive interactions at the intra-trimer interfaces at pH 5.0.^{32, 33} However, the disassembly resulted in denatured and insoluble protein aggregates which might lead to undesired side effects in future applications as drug carriers. In order to gain higher drug delivery efficiency and better biocompatibility, a pH-responsive E2 protein cage with non-denatured and soluble disassembly is investigated in this work.

Other protein cages, such as human ferritin light chain, could be modified at its subunit interface by incorporating a GALA peptide which resulted in a pH-responsive ferritin protein cage.²⁷ In this investigation, we focus on modifying the interactions

between subunits to design a novel pH-sensitive E2 protein cage. The E2 subunit interfaces were categorized into intra-trimer interface and inter-trimer interface. To prevent drastic disassembly to denatured aggregates, we hypothesize that introducing His residues at the critical sites on the intra-trimer interface without deleting the N-termini would disassemble the protein by disturbing the interactions within the trimers. Besides, trimers of E2 protein are tightly coupled by inter-trimer interactions. Incorporating His residues at inter-trimer interfaces may weaken the interaction at acidic pH and further disassemble the E2 protein cage.

Figure 3.1 Molecular structures highlighting identified key amino acids and relevant protein domains. (A) Quaternary structure of E2 protein viewed at the 5-fold axis of symmetry. The protein comprise 20 trimer structures, five of which are highlighted with yellow, green, red, orange, and magentas color, respectively. (B) Possible trimer structure (three subunits are shown in yellow, purple, and orange colors, respectively) and identified two native histidine residues (limon: H218 and lightblue: H222), and two interaction sites (violet: W355, and cyan: P377) that form clusters at intra-trimer interface. (C) Two trimers and the key amino acids at the inter-trimer interface (space-filled, red: I314, blue: L424, green: F315 and magentas: M425). (D) Close up view of key amino acids at intra-trimer interface. Color setting is the same as (B). (E) Close up view of (C) highlighting the interacting residues. Color setting is the same as (C).

3.3 Results and Discussion

3.3.1 Design and Construction of pH-Sensitive Mutant Proteins

The 60 identical subunits of E2 protein are held together by electrostatic and hydrophobic interactions, especially those located in the middle of the subunits. The crystallographic structure of E2 protein was visualized using PyMol¹³⁵ (Figure 3.1A). The potential key amino acids involved in subunits interactions were identified through visual inspection. Interactions at intra-trimer and inter-trimer interfaces were considered separately.

In addition to the native histidines 218 and 222 (H218 and H222) trimeric clusters at intra-trimer interfaces investigated in previous work,^{32, 33} we identified tryptophan 355 (W355) and proline 377 (P377) as potential key amino acids. The distance of W355 and P377 between three subunits was measured to be close to the Debye length, and form two trimeric clusters at the middle and the inner end of intra-trimer interface, respectively (Figure 3.1 B and D). On each subunit, three polar amino acids (G352, G353, and Q354) following the non-polar and hydrophobic W355 suggested that they might be exposed to the surrounding solvent. We hypothesized that W355 might be the first amino acid buried in the hydrophobic motif and changing the interactions between them may lead to conformational change of the motif. However, due to its buried location, W355 had limited solvent accessibility which may prevent protonation of the imidazole group. To address this issue, we also modified P377 which was likely to be solvent-exposed due to their proximity to the inner surface, thus might further act as an initial point in conformational change when these

interactions were substituted with charged or bulky residues. Similar to W355, P377 residues from three subunits also formed a trimer cluster near the inner end at intra-trimer interfaces (Figure 3.1 B and D) and was followed by two hydrophilic amino acids (E375, and K376) on each subunit.

For the inter-trimer interfaces, we identified four residues that were methionine 425 (M425), leucine 424 (L424), phenylalanine 315 (F315), and isoleucine 314 (I314). L424 and M425 were the last few amino acids at the C- terminal of the E2 subunit and located in a hydrophobic pocket between two trimers. L424 and M425 had close interactions with I314 and F315, which were located at the neighboring subunit, respectively (Figure 3.1 C and E). The distances between L424 and I314 and between M425 and F315 were about 0.3 nm which were both less than the Debye screening length of histidine.. Modifications of these potential key amino acids were expected to change the interactions between trimer structures and further influenced the stabilities of the E2 protein.

To synthesize pH-sensitive protein cage, histidine pairs were introduced sequentially at the identified key amino acids using site-directed mutagenesis. Plasmids containing mutated E2 genes, E2-H218/H222/W355H/P377H (E2-(2+2)H), and E2-I314H/L424H/F315H/M425H (E2-4H) were constructed. The mutated genes were PCR-amplified by Pfu DNA polymerase using oligonucleotides listed in table 3.1. The template used was the plasmid containing wild-type E2 gene (pE2).¹¹⁵

Table 3.1 Oligonucleotides for mutant plasmids construction. Mutation sites are in bold, genes are in uppercase, and restriction enzyme sites are underlined.

E2-(2+2)H	5'- GGT ATT GGT CGT ATA GCC GAA AAG CAT ATC GTT CGT GAC GGT GAA ATC -3' for E2-P377H 5'- CAT CGG CTC TGC AGG CAT TCA GTG GTT CAC CCC AGT TAT CAA CCA -3' for E2-W355H
E2-4H	5'- CAC GCG GAC CGT AAA CCG CAT CAT GCG CTC GCT CAG GAA ATC AAC -3' for E2-I314H/F315H 5'- GGA GAT ATA CAT ATG CTG TCT GTT CCT GGT CCC GCT - 3' (forward) 5'- tta gca gcc <u>gga tcc</u> TTA AGC TTC ATG ATG CAG CAG TTC CGG GTC GGA -3' (reverse) for E2-L424H/M425H

In order to construct the expression mutant vectors, vector pET-11a and the PCR products were digested with NdeI and BamHI restriction enzymes. Both vector and PCR products were purified with 0.8% agarose gel electrophoresis and ligated with T4 DNA ligase. The sequences of the mutated genes were confirmed by DNA sequencing service from 1st BASE (Singapore).

3.3.2 Mutant Proteins Show Correct Assemblies

In order to confirm the assemblies of both mutants at physiological pH, the E2-WT and mutants E2-(2+2)H and E2-4H were overexpressed, purified, and subsequently characterized. SDS-PAGE showed correct sizes of wild-type and mutant E2 proteins (Figure 3.2). Molecular mass of mutant subunits determined by MALDI-TOF/TOF analysis were within 0.2% of the calculated theoretical values of the corresponding protein (data not shown), indicating the proteins were correctly produced from constructed mutant sequences. The hydrodynamic diameters of mutant E2 proteins at physiological pH were measured using DLS technique (Table 3.2). The data showed

that the particle sizes of E2-(2+2)H and E2-4H were comparable to that of the E2-WT protein, indicating that the histidine substitutes at neither intra-trimer nor inter-trimer interfaces had apparent influences on the protein cage assembly at physiological pH. The assembly of the spherical hollow structure was further confirmed with TEM (Figure 3.3). The modified protein cages showed symmetry at the 2-fold, 3-fold, and 5-fold axes similar to those observed previously.¹¹⁵ The similar molecular masses and dimensions of these mutant proteins to those of the E2-WT protein indicated the correct assemblies of the mutant scaffolds. Dalmau et al. had previously demonstrated that the modified E2 protein cage with mutations in its hollow cavity had the capability to encapsulate guest drug-like molecules and antitumor drug.^{115, 118} The correct assemblies with non-native functionalities would allow further controlled release study from the E2 protein cage.

Figure 3.2 SDS-PAGE showing correctly produced and purified proteins. (1) E2-WT, (2) E2-(2+2)H, (3) E2-4H.

Figure 3.3 Electron micrographs showing the structures of wild type and mutants E2 at physiological pH 7.4. All proteins presented correctly assembled spherical structures. (A) E2-WT, (B) E2-(2+2)H, (C) E2-4H. Single protein cages are highlighted in red circles. Proteins were stained with 1.5% uranyl acetate. Scale bars are 50 nm.

Table 3.2 Hydrodynamic diameters (in nm) of WT and mutant E2 proteins in sodium phosphate buffer at pH 7.4 and 5.0.

	E2-WT	E2-(2+2)H	E2-4H
pH 7.4	25.0 $\pm$ 0.6	28.0 $\pm$ 0.3	23.5 $\pm$ 0.8
pH 5.0	25.0 $\pm$ 0.7	27.8 $\pm$ 0.5	307 $\pm$ 219

3.3.3 The Secondary Structures of the Mutant Proteins are Altered at Physiological pH

While the quaternary structure of the intra-trimer interfaces modified E2-(2+2)H and inter-trimer interfaces modified E2-4H protein cages showed correct assembly at physiological pH, the integrity and folding of the secondary structures could not be assumed *a priori*. To investigate the secondary structure, we performed far-UV CD analysis on all purified proteins. The CD spectra of all mutant proteins gave altered

secondary structure information compared to that of E2-WT. Previous works have shown that E2-WT and N-terminal truncated E2 mutant with fully assembled structure have CD spectra with minima at both 208 and 222 nm at pH 7.4.^{32, 33, 115, 118} In this work, mutants E2-(2+2)H and E2-4H, gave similar CD spectra at physiological pH (Figure 3.4). Compared to the spectrum of E2-WT, the spectra of mutant proteins showed partial loss of 222 nm minimum while the 208 nm minimum remained unchanged. Calculated by CDNN program, the CD spectra indicated that the mutant protein cages have decreased α -helix content, but increased proportion of β -sheets and random coil (data not shown). The introduction of multiple histidines at either intra-trimer interface or inter-trimer interface led to partial unfolding of the surrounding structure and caused the decrease of α -helix content. It was reported that the quasi-equivalent interactions of all known E2 species were centered around a highly conserved anchor residue.¹⁰⁸ In this work, the secondary structure change resulting from 4 histidines in this hydrophobic pocket at inter-trimer interfaces indicated the critical roles of I314, F315, L424, and M425 in the folding and associating of the E2 subunit.

Figure 3.4 Far-UV circular dichroism showing molar ellipticity versus wavelength for E2-WT, E2-(2+2)H, and E2-4H at pH 7.4. E2-WT shows the characteristic minima of an α -helix-rich protein at 208 and 222 nm. E2-(2+2)H and E2-4H represent similar spectra with partial loss of 222 nm minimum while their 208 nm minimum remains unchanged.

3.3.4 Intra-Trimer Modified Cage Retains its Correct Assembled Structure at pH 5.0

We aimed to design a drug carrier for endocytotic pathway mediated therapeutic treatment, in which potential carriers will experience physiological pH and acidic pH in succession. Therefore, pH-dependent disassembly of E2 cage is required. Our produced E2-(2+2)H and E2-4H had correct assembled structure at pH 7.4. However, when the pH was adjusted to 5.0, their assemblies responded differently to acidic environment.

For the intra-trimer modified E2-(2+2)H containing 4 histidine clusters, the DLS results showed that its size remained unchanged when subjected to pH 5.0 (Table 3.2). The TEM images confirmed that E2-(2+2)H had integral spherical structures at both pH 7.4 and 5.0 (Figures 3.3B and 3.5B). In addition, the similar CD spectra indicated

that the secondary structure of E2-(2+2)H nearly remained unchanged when pH was lowered from 7.4 to 5.0 (Figure 3.6A). It experienced slight loss of α -helix and increase of β -sheet and random coil (calculated by CDNN program). The results may be due to the restricted solvent access to the intra-trimer interface. The N-termini of the three subunits blocked solvent access to the intrinsic histidines at residues 218 and 222 while three loop structured peptide chains prevented solvent access from the inner surface to the histidines at residues 355 and 377 (Figure 3.1D). This work further confirmed that solvent access is important in the disruption of the intra-trimer interaction as previously reported where the removal of N-terminal residues exposed the native H218 and H222 trimer clusters to surrounding solvent and resulted in pH-responsive E2 protein cage.^{32, 33}

Figure 3.5 Electron micrographs of E2-WT and mutants at pH 5.0. (A) E2-WT and (B) E2-(2+2)H show correctly assembled structures, single protein cages are highlighted in red circles. (C) E2-4H shows irregular aggregations. Inset presents the zoom-in view of an aggregation along the three-fold axis. The red circle highlighted the folded structure along three-folded axis. Proteins were stained with 1.5% uranyl acetate. Scale bars are 50 nm.

Figure 3.6 Comparisons of far-UV circular dichroism spectra of mutant E2 proteins at pH 7.4 and 5.0. Molar ellipticity versus wavelength for (A) E2-(2+2)H, and (B) E2-4H. E2-(2+2)H and E2-4H show similar spectra at both pH 7.4 and 5.0.

3.3.5 Inter-Trimer Modified Cages Show Aggregations at pH 5.0

For the inter-trimer interface modified mutant protein, E2-4H presented aggregations at pH 5.0. When pH was lowered to 5.0, the DLS results showed that the size of E2-4H increased from 23.5 ± 0.8 nm to 307 ± 219 nm (Table 3.2 and Figure 3.7). The electron micrographs indicated irregular aggregation of several E2-4H proteins (Figure 3.5C). The aggregates presented folded structure, such as the symmetry along the three-fold axis (Figure 3.5C, inset), suggesting that E2-4H still maintained part of its quaternary structures after aggregation. The interactions between trimers were influenced by protonation of histidine residues within the Debye radius at pH 5.0 and exposed amino acid residues. However, we speculated that the weakening of essential interactions at inter-trimer interfaces was not able to disassociate E2-4H into individual trimers. As a result, E2-4H might undergo partial disassembly from its inter-trimer interfaces. The exposed amino acids at inter-trimer interfaces might subsequently lead to random association between partially

disassembled E2-4H, and resulted in non-uniform aggregates. Drugs delivered in protein cage could be released at target sites by either biodegradation or disassembly of protein cage carriers. The disruption behaviors at the interface of the trimers structures might facilitate drug release from E2 protein cage in future applications.

Figure 3.7 Representative DLS scan shows a hydrodynamic diameter of E2-4H at pH 7.4 (blue) and 5.0 (red).

3.3.6 Disassembly from Inter-Trimer Interface does not Denature E2 Subunits and is Irreversible

CD can be used to evaluate denaturation and aggregation of protein.^{136, 137} In previous work, the truncation of N-terminal arm exposed histidines at the intra-trimer interface and caused aggregation of E2 protein at pH 5.0.^{32, 33} The denatured and insoluble E2-ΔN showed significant change in the CD spectrum (loss of 208 nm minimum and shift of 222 nm minimum) and presented cloudy aggregates under electron microscope. In this report, the conformational changes of E2-4H at acidic pH resulted in soluble aggregates. The secondary structure measurement supported the observation. While E2-4H lost most of its spherical quaternary structure and was present in aggregated state as observed under TEM and through DLS measurements, the CD spectra of E2-4H showed similar curves with unchanged 208 and 222 minima

at both pH-s, indicating the preserved secondary structures upon pH change (Figure 3.6B). The mutant E2-4H in this work was designed to disrupt the quaternary structure of the protein cage from the inter-trimer interfaces without denaturation. The disrupted interactions presumably only occurred at the histidine-modified inter-trimer interfaces while majority of the quaternary structure was preserved (Figure 3.5C, inset). Hence, for E2-4H, the pH-induced repulsive interactions only partially dissociated the inter-trimer interactions but maintained the folded state of the protein subunit as reflected by the unchanged CD spectra. Comparison between the disruptive effects from intra- and inter-trimer interfaces suggests that trimers might be present as a basic building block to form the fully assembled virus-like E2 protein. Some virus, such as human immunodeficiency virus (HIV),¹³⁸ Rice dwarf virus (RDV),¹³⁹ and coronavirus spike protein¹⁴⁰ adopt trimer intermediate in the formation of their capsids.

In this study, the protonated histidines at the inter-trimer interfaces triggered partial disassembly and aggregation of E2 protein. Protein fragments remained soluble during the disassembly process, which was verified by ultracentrifugation. Similar to previous work which reported that pH-triggered disassociation of N-terminal truncated E2 protein (E2-ΔN) was irreversible,³² the aggregated E2-4H remained in aggregate forms after buffer-exchange from pH 5.0 to pH 7.4 (data not shown). As the buffer pH was lowered to 5.0, the originally buried amino acids at inter-trimer interfaces were exposed and mediated the random irreversible aggregation of E2-4H.

3.3.7 Cross-Linking Verifies Non-Denatured Disassembly and Suggests Trimer as Building Block

To confirm that the observed aggregation was due to the exposed amino acid residues at the inter-trimer interface, we cross-linked the proteins using glutaraldehyde and observed them under transmission electron microscope. We verified the cross-linking between E2 subunits as well as between the protein cages using SDS-PAGE. In the presence of SDS, single subunit of ~28 kDa and higher molecular weight species were expected for uncross-linked and cross-linked E2 proteins, respectively. Figure 2.8A illustrated the molecular weights of all E2 proteins on SDS-PAGE before and after cross-linking at pH 7.4 and pH 5.0. While WT and mutant E2 proteins were present as single subunit at both pH-s, cross-linked proteins gave higher molecular weights. At both pH-s, most of the cross-linked E2-WT and E2-(2+2)H formed agglomerates that were too large to be separated on the gel. As a result, the proteins accumulated at the sample loading wells. However, a large proportion of the E2-4H were between 75 kDa and 100 kDa at both pH-s (Figure 3.8A, red rectangle) which agreed well with the molecular weight of the trimer structure of ~84 kDa. We speculated that the cross-linking further stabilized the trimer structures, while some of the interactions between trimers were lost during SDS treatment because of the modifications at the 4 critical sites (Appendix A.1). The electron micrographs indicated the correctly assembled structures of E2-WT and E2-(2+2)H at both pH 7.4 and 5.0, and E2-4H at pH 7.4 (Figure 3.8 B-F). In contrast, only some of E2-4H showed correctly assembled structures at pH 5.0 (Figure 3.8G, red circles). The observed aggregation of the majority E2-4H proteins (Figure 3.8G, red oval) - similar to that in the micrograph of uncross-linked E2-4H at pH 5.0 (Figure 3.5C, red oval) - might result from interactions between the exposed amino acids at inter-trimer

interfaces. The results confirmed that the partial disassembly and aggregation of inter-trimer modified E2-4H was a non-denaturing process. The existence of possible trimer structure might suggest its role as the building block in the assembly of the E2 protein cage.

Figure 3.8 SDS-PAGE and electron micrographs of E2-WT and mutants E2-(2+2)H and E2-4H after cross-linking. (A) SDS-PAGE showing single subunit before cross-linking and aggregate after cross-linking. Lane a and b present protein before cross-linking at pH 7.4 and 5.0, respectively. Lane c and d present protein after cross-linking at pH 7.4 and 5.0, respectively. After cross-linking, E2-WT at (B) pH 7.4 and (E) pH 5.0, E2-(2+2)H at (C) pH 7.4 and (F) pH 5.0, and E2-4H at (D) pH 7.4 show correct assembled structure under electron microscopy. E2-4H at (G) pH 5.0 show correct assembled structures and aggregates. Single protein cages are highlighted in red circles, and aggregate is highlighted in red oval. Proteins were stained with 1.5% uranyl acetate. Scale bars are 50 nm.

3.4 Conclusions

We investigated the interactions at intra- and inter-trimer interfaces of E2 protein from *B. sterothermophilus* to design a pH-sensitive protein cage. For the intra-trimer modified E2-(2+2)H, our modifications resulted in 4 histidine clusters - two of which were native - between the interfaces. The mutant protein is correctly assembled at pH 7.4, but insensitive to pH change due to restricted solvent access in the presence of N-terminal 'arms' at the outer surface and 'loops' at the inner surface. For the inter-trimer interface modified proteins, we designed E2-4H to trigger repulsive interactions. The mutant presented correctly assembled structures at pH 7.4. When the pH was adjusted to 5.0, E2-4H experienced irreversible disassociation and resulted in soluble aggregates. The observation demonstrated that the interactions between the closely coupled amino acids at inter-trimer interfaces are critical in the formation of fully assembled E2 protein. Unlike the drastic and denaturing disassembly from intra-trimer interfaces, alteration of interactions at inter-trimer interfaces led to conformational change and partial disruption of E2 quaternary structure. The secondary structures remained unchanged and the protein was soluble and non-denatured. The comparison between the effects of modifications at inter- and intra-trimer interfaces on the assembly of virus-like E2 protein may provide clues to understanding its self-assembly mechanism.

Chapter 4

Isolating a Trimer Intermediate in the Self-Assembly of E2

Protein Cage

This chapter is a modified version of the previously published work.

Reprinted with permission from ‘Tao Peng, Hwankyu Lee, and Sierin Lim; *Isolating a Trimer Intermediate in the Self-assembly of E2 Protein Cage*, Biomacromolecules, **2012**, 13(3): 699-705.’

COPYRIGHT © 2012 AMERICAN CHEMICAL SOCIETY

4.1 Abstract

Understanding the self-assembly mechanism of caged proteins provides clues to develop their potential applications in nanotechnology, such as nano-scale drug delivery system. E2 protein from *B. stearothermophilus* with virus-like caged structure has drawn much attention for their potential application as nanocapsule. To investigate its self-assembly process from subunits to spherical protein cage, we truncate the C-terminus of the E2 subunit. The redesigned protein subunit shows dynamic transition between monomer and trimer, but not the integrate 60-mer. The results indicate the role of trimer as the intermediate and building block during the self-assembly of E2 protein cage. In combination with the molecular dynamic simulation results, we conclude that the C-terminus modulate the self-assembly of E2 protein cage from trimer to 60-mer. This investigation elucidates the role of inter-subunit interactions in engineering other functionalities on other caged structure proteins.

4.2 Introduction

Previous works have demonstrated that the E2 protein cage packed with drug-like fluorescent dye or antitumor drug doxorubicin in its inner cavity can be taken into breast cancer cells *in vitro*.¹¹⁸ To control the release of molecular cargo from the protein cage in future applications, understanding the self-assembly mechanism of E2 protein cage is essentially required.

Some viruses, such as human immunodeficiency virus (HIV),¹³⁸ Rice dwarf virus (RDV),¹³⁹ and coronavirus spike protein¹⁴⁰ adopt trimer intermediates in the formation of their capsids from numbers of identical subunits. Although E2 protein cage and viral capsids have neither sequence homology nor natural function in common, the virus-like E2 protein cage are structural parallel to viral particles with dodecahedral symmetry.³² We hypothesize that the trimer intermediate forms prior to the fully assembled E2 protein cage. The virion structure of many icosahedral viruses is determined and directed by the N-terminus of the capsids subunit.^{29, 31, 141, 142} These terminal peptides are resolved in crystallographic data and shown to embrace an adjacent subunit. Deletion of these terminal peptides resulted in smaller particle size, symmetry change, or even no assembly of virus capsids.^{18, 29, 103} In contrast to the virus capsids, the formation of the assembled cage structure from E2 protein subunits may not be directed by the N-terminus. Based on the crystallographic structure, the N-terminus of E2 protein cage from *B. stearothermophilus* enfolds an adjacent subunit within a trimer cluster.¹⁰⁸ However, deletion of the N-terminus showed that the protein still correctly assembled into dodecahedral structure as wild type E2 protein

(E2-WT) at physiological pH 7.4, but denatured and aggregated at pH 5.0.^{32, 33} The absence of the N-terminus did not influence the self-assembly of E2 protein cage at pH 7.4, but exposed a histidine cluster and provided solvent access to the intratrimer interfaces. Consequently, at pH 5.0, the exposed histidine clusters at intratrimer interfaces were protonated and generated repulsive interactions that led to the dissociation of E2 protein cage. Since this pH-responsive switch resulted in irreversible denaturation, we speculated that the trimer structure might be destroyed due to the solvent access at the intratrimer interface. In a later work, the partial disruptive interactions were introduced at intertrimer interfaces and resulted in non-denatured partial disassembly and aggregation of E2 protein cage at pH 5.0.⁴² Throughout the pH-triggered disassembly, most of the secondary and quaternary structures remained unchanged. Without influencing the association within trimer structures, the E2 protein cage experienced partial disassociation from the intertrimer interfaces. Comparison between the disruptive effects from intratrimer and intertrimer interface in these two works suggests that trimer may be a basic building block in the formation of the fully assembled *B. sterothermophilus* E2 protein cage.

In other species, such as gram-negative bacteria *Azotobacter vinelandii*, the E2 component consists of 24 subunits arranged with octahedral symmetry. The crystal structure revealed the extensive interactions between 3-fold related subunits leading to a tightly associated trimer, and the interactions along the 2-fold axis leading to the assembly of the trimers into the 24-mer. The observations suggested trimer as building block for the 24-mer E2 protein of this gram-negative bacterium.^{143, 144} In contrast, the E2 component of mammals, yeast, fungi, or the gram-positive bacteria consists of 60 subunits arranged in dodecahedral symmetry.¹⁴⁵ Although the homologies in both sequence and structure between 24-mer and 60-mer E2 proteins

are distinct, the trimer of the cubic *A. vinelandii* E2 protein was assumed to be equivalent to the trimer in the dodecahedral 60-mer E2 proteins¹⁰⁸. Trimer structure is suggested to be important in the formation of these caged structures. When reassembling the guanidine hydrochloride disassociated monomers of bovine E2 protein, the assembly intermediate with similar molecular weight to that of E2 trimer was observed during sedimentation velocity analysis.^{146, 147} The preliminary evidence suggested that the formation of E2 core structure might proceed through trimer intermediate. However, in a later work, the corresponding trimer peak was not observed.¹⁴⁸

In this investigation, we attempt to isolate the trimer structures to further understand the self-assembly mechanism of E2 protein cage. Visualized from its crystallographic structure, the E2 protein forms a highly symmetrical structure at the 2-fold, 3-fold, and 5-fold axes with the trimer structure as the smallest repeating unit. The interactions between intertrimer interfaces of E2 protein cage are altered and the trimer intermediates were obtained. The redesigned E2 protein retains its secondary structures, and is present as trimer and monomer in solution. The dynamic transition between trimer and monomer suggests that the trimer is stable enough to be present as an independent structure and a building block in the formation of fully assembled E2 protein cage.

Unlike other protein cages, such as MhpD (2-hydroxypentadienoic acid hydratase)¹⁴⁹ and ferritin.¹⁵⁰ which can be produced in monomeric and dimeric subunit, E2-WT is always produced as fully assembled 60-mer structure. The extremely stable interactions of the E2 subunits render attempts to isolate the monomers, by subjecting the E2-WT to high concentration of denaturants, unsuccessful (Appendices A.1). The

E2-WT structure is reported to be heat stable up to approximately 85 °C and also in a wide range of pH.¹¹⁵

To disassociate the integrate E2 protein cage to intermediates or individual subunit, the last α -helix motif at the C-terminus was truncated by genetic modifications (Figure 4.1). In previous work, substitution of two key amino acids located at the last α -helix motif of the C-terminus, L424 and M425, with histidines affected the disassembly behavior of E2 protein at pH 5.0, indicating the critical role of the C-terminus in maintaining the protein structure.⁴² This α -helix motif at the C-terminus, which contains 9 amino acids, is located within a hydrophobic pocket between two trimer intermediates.¹⁰⁸ The truncation of the small α -helix motif is expected to result in increased solvent exposure area at the E2 protein C-terminus and disrupt the original interactions here. Hence, we anticipate that the disruptive modifications at the interfaces between the 20 groups of E2 trimers will not prevent folding and association of E2 subunits into trimer intermediates, but will prevent association between trimers, and subsequently impair the full protein cage assembly.

Figure 4.1 Molecular structures highlighting truncated C-terminal α -helix of E2 protein. (A) Overview of the E2 protein cage showing the truncated α -helix motif at C-terminus (red-colored ribbon). (B) Close-up view of two trimer intermediates. The red ribbons represent the truncated α -helix motif at C-terminus between any two trimers.

4.3 Results and Discussions

4.3.1 Design and Construction of Mutant Protein

To generate C-terminus truncated protein (E2-ΔC9; PDB 1B5S aa184-418), plasmid containing gene encoding E2-WT (pE2) was PCR-amplified using Pfu DNA polymerase (Fermentas) with oligonucleotides: 5'- gggaattc CAT ATG CTG TCT GTT CCT GGT CCC -3' (forward) and 5'- gcc GGA TTC TTA GGA CAG CAG ACG TTT GAT GTG GTT CAG – 3' (reverse). In order to construct the expression vectors, vector pET-11a, and the PCR product were digested with NdeI and BamHI restriction enzymes, purified with 0.8% agarose gel electrophoresis, and ligated with T4 DNA ligase. The sequences of the mutated genes were confirmed by DNA sequencing service from 1stBASE.

In addition to the 4 histidines contained in E2-4H described in chapter 3,⁴² one more histidine was introduced at inter-trimer interfaces to evaluate the interactions involved in the associating trimers. The expression plasmid of E2 protein carrying 5 histidine mutations (E2-5H) was constructed using pE2-4H expression plasmid as a template using oligonucleotide 5'- GTT CCT GTG ATT AAA CAC GCG CAT CGT AAA CCG CAT CAT GCG CTC -3'.

4.3.2 E2-ΔC9 is Present as Both Monomer and Trimer

To examine the assembly behavior of the truncated subunit, purified E2-ΔC9 was concentrated to 4 mg/ml in sodium phosphate buffer at pH 7.4 (50 mM sodium phosphate, 150 mM sodium chloride, 5 mM EDTA, and 0.02% sodium azide), and subjected to SEC. SEC can be used to determine the multimeric state of the modified protein by indicating their molecular weight (MW) from the corresponding elution volumes. If the associations of each subunit or trimer structure were disrupted by removal of the C-terminal α -helix, the corresponding peaks of monomer or trimer would be detected on the elution profile.

Figure 4.2 shows the combined SEC profiles of E2-ΔC9 at different initial concentrations. The presence of 2 peaks at elution volume of 14.7 and 17.3 ml in all elution profiles of E2-ΔC9 correspond to species of molecular mass 81 and 27 kDa, respectively. Since the theoretical molecular mass of E2-ΔC9 subunit is 27.1 kDa, we speculate that E2-ΔC9 exists as trimer and monomer. As a control, 60-mer E2-WT with molecular mass 1687 kDa come out at the void volume (V_0) of 9.2 ml (Figure 3.2). To confirm that both eluted assemblies of E2-ΔC9 are composed of E2 subunit, SDS-PAGE and MALDI-TOF/TOF were adopted to analyze the monomeric unit of the fractions. Due to the low concentration after SEC step, trimer and monomer fractions were precipitated with 10% trichloroacetic acid (TCA) before SDS-PAGE. Figure 4.3 indicates that the trimer and monomer fractions have similar molecular weights to that of E2-WT on SDS-PAGE gel. Molecular masses of both fractions determined by MALDI-TOF/TOF were within 0.3% of calculated theoretical value of E2-ΔC9 subunit, implying both the fractions were recombinant E2 protein (data not shown). The DLS result shows that the hydrodynamic diameter of E2-ΔC9 trimer is

8.01 ± 0.21 nm (Figure 4.4). A single peak was observed on the DLS spectrum due to the limitation of the DLS technique in distinguishing the small proportion of monomer and the small size difference between trimer and monomer.

Figure 4.2 SEC profiles of E2-ΔC9 at different concentrations. E2-WT elutes at the void volume of the column (9.2 ml). The inset presenting E2-ΔC9 at 0.13 mg/ml has two elution peaks at 14.7 and 17.3 ml.

Figure 4.3 SDS-PAGE analysis of (1) E2-WT, (2) trimer fraction and (3) monomer fraction of E2-ΔC9. The trimer and monomer fractions were collected and precipitated with 10% TCA before loading onto the gel.

Figure 4.4 Representative DLS scan shows the hydrodynamic diameter of about 8 nm and 25 nm for trimer (blue) and wild-type 60-mer (red) structures of E2 protein, respectively.

To further confirm the composition and folding of the mutant E2- Δ C9, we used CD to evaluate the secondary structures of each fraction from SEC, with E2-WT as a control. CD is an excellent method to evaluate the secondary structure, folding, and binding properties of proteins¹⁵¹. Since α -helical secondary structure requires precise interactions and conformational geometries, the unfolding of α -helix is always detected in CD measurement.^{152, 153} The crystallographic structure revealed that approximately one-third of the E2-WT secondary structure was α -helix. In this investigation, both trimer and monomer fractions give similar spectra as the E2-WT (Figure 4.5). The presence of two minima at 208 and 222 nm, which is characteristic of α -helix rich protein,¹¹⁵ indicates correct folding of the monomers and trimers. The secondary structure content calculation using CDNN program¹²² implies a slight loss of α -helix structure from the trimer and monomer compared to that of E2-WT (data not shown). The result is consistent with the removal the last α -helix motif at the C-terminus in the experimental design.

Figure 4.5 CD spectra of E2-WT and fractions of E2- Δ C9 from SEC. Both trimer and monomer fractions show characteristic spectra of α -helix rich protein with minima at 208 and 222 nm similar to that of the E2-WT.

The E2- Δ C9 is present as monomers and trimers, but cannot form larger assemblies in solution. The truncation of the α -helix motif completely disrupted the association between trimers. Previous work showed that the modifications on this motif together with other surrounding amino acids changed the recognition features at intertrimer interface and led to larger aggregations from random assembly in acidic environment.⁴² The results implied that the integrity of the C-terminus was essentially required to modulate the assembly of spherical 60-mer. In this work, the association between trimers is eliminated by the absence of the C-terminus, while the intratrimer interactions is preserved, as the subunits of E2- Δ C9 can still assemble into trimers. This is consistent with a work by Dalmau et al. that described the truncation of N-terminus on E2 protein would not affect the formation of the 60-mer at pH 7.4³². We speculate that unlike the formation of some viruses capsids which is guided by subunit

N-terminus,^{29, 31, 141, 142} the assembly process from trimer to 60-mer of the virus-like E2 protein cage is guided by the C-terminus. Upon intracellular production, the individual E2 subunits are likely to self-assemble and form the intermediates - trimers - first, and subsequently the C-termini take part in modulating the trimers to form the fully assembled 60-mer. In this work, the truncation of the C-terminus halts the self-assembly of the recombinant E2 protein at trimer intermediate state.

4.3.3 E2-ΔC9 Shows Dynamic Transitions Between Monomer and Trimer

To assess the dynamic effect on the trimer intermediate formation in solution, different concentrations of E2-ΔC9 preparation were subjected to SEC column. The SEC profiles on Figure 4.2 imply that the formation of trimer is protein concentration dependent. The dominant existence indicates that upon disruption of the intertrimer interactions, trimer is the more stable state than monomer. Although the trimers presented as majority in solution at all concentration range from 0.13 to 4 mg/ml, the proportion of trimer and monomer are different in the various concentrations. The integrated signals of SEC profiles in Figure 4.2 were analyzed and the relation between trimer percentage and concentration of E2-ΔC9 was obtained (Figure 4.6A). More than 90% of the protein is present as trimer when the concentration is equal or higher than 0.25 mg/ml. The transition between trimer and monomer is a dynamic process. As the protein concentration decreases, the proportion of monomer increases. However, as indicated by our results, when the concentration of E2-ΔC9 is higher than 0.25 mg/ml, the trend of the multimeric transition reaches a plateau. We speculate that the missing of C-terminus halts the self-assembly process which results

in the dominant presence of the trimer. The trimer percentage is drastically dropped to about 65% when the protein concentration is 0.13 mg/ml. The inset on Figure 4.2 indicates the comparable trimer and monomer peaks at protein concentration 0.13 mg/ml. The results support the hypothesis that trimer is an intermediate in the formation of fully assembled E2 protein cage.

Figure 4.6 Relations between trimer percentages and concentrations of (A) E2-ΔC9 and (B) Redistributed trimer fraction based on the integrated signals of SEC profiles in Figure 3.2 and Figure 3.7, respectively.

To further confirm that the trimer was formed from monomers after gene expression, we investigated the transition between trimers and monomers. SEC fractions containing trimer were pooled and concentrated. After overnight incubation in sodium phosphate buffer at 4 °C, different concentrations of trimer fractions (0.5, 0.25, and 0.13 mg/ml) were reloaded onto the SEC column. The combined profiles in Figure 3.7 indicate the redistribution of monomer and trimer. The trimer percentages were also obtained by analyzing the integrated signals (Figure 4.6B). We also observed the concentration-dependent dynamic transitions between trimer and monomer, which showed that the monomers were readily present as a result of trimer dissociation. As reflected in Figure 4.6B, the redistribution of collected trimer

fractions shows comparable trimer percentages at the experimental concentrations to that of E2- Δ C9 in Figure 4.6A. The similar distributions of trimer in these two cases suggest that the concentration-dependent transitions of E2 subunits follow a specific constant. In contrast to E2-WT which is present only as 60-mer in a wide range of ionic concentration and pH¹¹⁵, E2- Δ C9 is present as a mixture of trimer and monomer. We speculate that the intertrimer interactions anchor the assembly process of the E2 protein cage, while the association and disassociation between monomers and trimers are reversible depending on the different protein concentrations. The non-existence of other sizes of multi-mers during the transition further confirmed the role of trimer as intermediate in the self-assembly of E2 protein cage.

Figure 4.7 SEC profiles of collected trimer fractions at different concentrations. Trimer-containing fractions were pooled, concentrated, and then reloaded to SEC column at different concentrations.

4.3.4 Molecular Dynamics Simulations Support the Importance of Interactions between Trimers

To understand the intertrimer interactions at atomic level and investigate the mechanism of C-terminus modulated assembly of E2 protein cage, molecular dynamics simulations of E2-WT and E2- Δ C9 were performed with explicit water.

Figure 4.8 shows snapshots at the beginning (0 ns; top) and the end of simulations (20 ns; bottom) of E2-WT (Figure 3.8A) and E2- Δ C9 (Figure 4.8B). The monomers of E2-WT form close coupled conformation throughout the whole simulation process, whereas the E2- Δ C9 monomers drift apart by the end of the simulation. These results indicate that the C-terminal- α -helix-guided intertrimer interaction is essential in associating the trimers and forming the 60-mer structure, supporting the experimental observations of the assembled 60-mer structure for E2-WT but not for E2- Δ C9. To further predict the effect of the C-terminus in the unfolding of E2 protein cage, the secondary structure changes of E2-WT and E2- Δ C9 were monitored during the simulation. For both E2-WT and E2- Δ C9, the α -helices and β -sheets retain most of their structures throughout the simulation time frame of 20 ns suggesting that the unfolding of the E2-WT and E2- Δ C9 structure is unlikely to occur (Figure 4.9). The results are in accordance with the CD results.

Figure 4.8 Snapshots at the beginning (0 ns, top) and the end (20 ns, bottom) of simulations of (A) E2-WT, (B) E2- Δ C9, and (C) E2-5H. Two E2 subunits are represented as blue and yellow ribbons, respectively. C-terminal α -helices of E2-WT are highlighted as RED ribbons. The images were created with visual molecular dynamics (VMD). (D) Molecular structure highlighting the identified key amino acids of E2-5H. Other than the 4 key amino mentioned in previous work,⁴² the newly identified D310s are represented as green spheres.

Figure 4.9 Secondary structures of E2-WT (top) and E2- Δ C9 (bottom) as a function of time during the simulation process.

4.3.5 C-Terminus Mediates the Self-Assembly from Trimer to 60-mer

To understand the molecular mechanism of the C-terminus modulated intertrimer interactions, the interactions on the last α -helix motif were analyzed. In our previous work, the histidine introduction on 4 key amino acids at intertrimer interfaces (E2-4H), two of which were located at the C-terminus, did not change the self-assembly behavior of E2 protein at physiological pH.⁴² To enhance the influence, another key amino acid (Asp-310) which is assumed to take part in mediating interactions at intertrimer interface was also modified to histidine in addition to the original 4 histidines resulting in 5-histidine containing E2 protein (E2-5H). Simulation was performed on the E2-5H to evaluate the roles of these key amino acids on self-assembly between trimers at physiological pH. Figure 4.8C illustrates that E2-5H represents the same assembly behavior to that of E2-WT, as the two monomers are closely coupled at intertrimer interface throughout the simulation process. The associations between trimers are not affected by those 5 modified histidines as supported by an experimental result.

Detailed interacting amino acids for E2-5H are illustrated on Figure 4.8D. The electron micrograph indicated the correct assembled spherical structures of E2-WT and E2-5H (Figure 4.10 A and B) at physiological pH. The modified protein cages showed symmetry at the 2-fold, 3-fold, and 5-fold axes similar to those observed previously.¹¹⁵ Compared to E2-WT, the CD spectrum of E2-5H gave similar profile to E2-4H showing partial loss of 222 nm minimum while the 208 nm minimum remained unchanged (Figure 4.10 C). The CD spectra indicated that E2-5H has decreased α -helix content, but increased proportion of β -sheets and random coil. The

introduction of multiple histidines at the inter-trimer interface led to partial unfolding of the surrounding structure and caused the decrease of α -helix content. However, the self-assembly to 60-mer E2 protein cage was not affected by the inter-trimer 5 histidines introduced at those particular amino acids.

Figure 4.10 Transmission electron micrographs of (A) E2-WT and (B) E2-5H showing correct assembled structures at pH 7.4. Proteins were stained with 1.5% uranyl acetate. Scale bars are 50 nm. (C) Comparisons of far-UV circular dichroism spectra of E2-WT and E2-5H at pH 7.4.

To assess the critical sites in maintaining the association between trimers, the interactions on key amino acids are simulated. Figure 4.11 shows that anionic Asp419 (D419) and Glu421 (E421) from one trimer subunit strongly interact with cationic Lys233 (K233), Arg238 (R238), and Lys240 (K240) on the neighboring trimer subunit. In particular, R238 and E421 are observed to have relatively stronger interaction with each other. Both 233-240 and 419-425 residues form α -helix structures. These results indicate that D419 and E421-induced interhelical charge interactions play an important role in increasing the stability of the intertrimer interaction, which is essential for the formation of fully assembled E2 protein cage.

Figure 4.11 Snapshot of the interactions between two E2-WT monomers at intertrimer interface. The monomers are represented as blue and yellow ribbons, respectively. The interacting residues are represented as colored bonds. C, N, O, and H atoms are represented as light blue, dark blue, red, and white colors.

Based on the mediator role of the C-terminus, future modifications can be introduced at the intertrimer to control the self-assembly behavior under different pH or ionic strength. Furthermore, the accessibilities of different amino acid residues or interfaces, provides a platform onto which additional functionalities can be incorporated without influencing the self-assembly of the protein cage.

4.4 Conclusions

Truncation of α -helix at the C-terminus of *B. stearothermophilus* E2 protein allows us to isolate the trimer structure. The truncation of α -helix completely disrupted the interactions between trimers but not the formation of trimers from the monomers. As a result, the E2- Δ C9 is present as both monomer and trimer in solution. The dynamic transition between monomer and trimer and the dominant presence of the trimer suggest its critical role as an intermediate in the formation of fully assembled E2 protein cage. The non-existence of 60-mer indicates the key role of the C-terminus in modulating the trimers to form the 60-mer *B. stearothermophilus* E2 protein. Identification of the trimer intermediate may allow future triggers to be designed and engineered onto the C-terminus at the trimer interfaces for controlled release applications. The study of self-assembly mechanism of the 60-mer dodecahedral E2 protein cage provides groundwork to investigate the recognitions and interactions of other multi-subunit proteins. Understanding their self-assembly mechanism will benefit the design of macrostructure/functionalities incorporation on the protein cage in future applications.

Chapter 5

Design of Reversible Inversed pH-Responsive E2 Protein

Cage

5.1 Abstract

The self-assembled caged-like E2 protein from pyruvate dehydrogenase has potential in a wide range of nanotechnology applications. The self-assembly can be designed to respond to pH change in two directions: with increasing or with decreasing pH. The artificial GALA peptide experiences helix-to-coil transition inducible by pH change. By incorporating a GALA peptide at the C-terminus of E2 protein cage, we report the first engineered caged-like protein with reversible pH-responsive capability. The redesigned E2 protein dissociates to trimers at pH 7.0 and further dissociates to monomers at pH 5.0. However, when pH is lowered to 4.0, the monomers self-assembled into both caged-like 60-mer and irregular assemblies. The assembly and disassembly processes of the protein cage are reversible. This special pH trigger will broaden the potential applications of caged-like proteins as molecular switches.

5.2 Introduction

Modification of the interactions at the subunit interfaces imparts non-native self-assembly behavior on the caged-like proteins (CLP) including controlled self-assembly capabilities in response to particular environmental cues. The controllable self-assembly behavior can be utilized in various applications, such as controlled release drug delivery system. For example, pH trigger has been used in designing pH-inducible protein molecular switches.^{27, 33, 42} In previous works, the subunit interfaces of E2 and ferritin protein cages have been genetically modified to impart pH-responsive self-assembly behaviors^{27, 33, 42} that is distinct from the native pH-inert CLPs. These modified CLPs are characterized to be full assembly at high pH (hi-pH assembly; hiA) and disassembly at low pH (lo-pH disassembly; loD). Interestingly, some naturally occurring CLPs present an alternative pH-responsive self-assembling behavior. For example, the capsids of cowpea chlorotic mottle virus (CCMV) and Norwalk virus disassemble at high pH, while remain stable and assembled dodecahedral core at acidic pH.¹⁵⁴⁻¹⁵⁶ Despite previous attempts, there has been no report on any engineered CLP with inversed pH-sensitive self-assembly, that is disassembled at high pH (hi-pH disassembly; hiD) while remain assembled at low pH (lo-pH assembly; loA). Furthermore, engineering a reversible self-assembly property is a challenge. Many efforts to reassemble E2 protein from disassembled subunits have been unsuccessful.^{32, 33, 42} Controlling the self-assembly of protein cages will set groundwork to better understand the interactions between CLP subunits and broaden their potential applications under specific conditions. In this work, we aim to engineer the E2 caged-like protein with reversible controlled hiD/loA self-assembly features.

E2 protein provides highly amenable subunit interfaces. Previous works reported that modifying the N- and C- terminus of E2 subunit resulted in hiA/loD pH-responsive CLPs, which maintained the assembled structures at pH 7.4 but disassembled at pH 5.0.^{33, 42}

E2 protein has always been produced as fully assembled 60-subunit structure. Attempts to denature/renature E2 protein into its individual subunits in vitro have been unsuccessful suggesting that the in vivo assembly upon translation is extremely stable (Appendix A.1). Our previous observations suggest that the E2 protein assembles through the formation of trimer intermediates.¹⁵⁷ We speculate that the N-termini of three polypeptides come together upon translation and subsequently twenty of the trimers form the fully assembled E2 protein through interaction of the C-termini. Therefore, trimer structure is the building block to form fully assembled 60-mer E2 protein.^{42, 157} C-terminus is located at the trimer-trimer interface to mediate self-assembly.¹⁵⁷ To impart the alternative pH-responsive manner to the E2 protein, we plan to incorporate a GALA peptide at the C-terminus. GALA peptide refers to artificial synthetic amino acid repeats of Glu-Ala-Leu-Ala, of which the length can be varied according to different functions.¹⁵⁸ The structure of GALA peptide is pH-responsive. The formation of extended random coil at neutral or basic pH and folded α -helix at acidic pH has been shown to be a reversible conversion process.^{158, 159} Choi et al has incorporated the GALA peptides to the terminus of ferritin CLP to impart pH-responsive capability.²⁷

We notice the last motif at the C-terminus of E2-WT subunit is α -helix. The α -helix terminus is essentially required in associating and maintaining the assembled 60-mer structure of E2 protein, as the truncation of this α -helix results in the absence of 60-mer.¹⁵⁷ Since E2-WT maintains the 60-mer assembled structure in a wide range of

pH,³³ the original α -helix C-terminus should be resistant to pH change. In this work, we use a pH-responsive GALA peptide to substitute the original C-terminal α -helix of E2 protein (Figure 5.1). In order to maintain the spatial structure with constrained length of C-terminus at inter-trimer interfaces, as well as introduce functional GALA peptide with helix-to-coil transitions, EAALAEALA was incorporated to replace the original ELLLMEA (amino acids 421- 427) at the C-terminus to obtain recombinant E2 protein (E2-GALA). As triggered by pH change, the self-assembly of E2-GALA is controlled by the coil-to-helix transition of the incorporated GALA peptide.

Figure 5.1 Molecular structures highlighting the substitution of C-terminal α -helix in E2 protein with GALA peptide. (A) Overview of the 60-mer E2 protein cage showing the C-terminal α -helix in blue. (B) Close-up view of the target α -helix highlighted in dash purple oval at trimer-trimer interface. (C) Substitution the C-terminal α -helix with GALA peptide. GALA peptide presents reversible α -helix-to-coil transition.

5.3 Results and Discussions

5.3.1 Construction of GALA Incorporated E2 Protein

Our previous work has reported that Asp419 and Glu421 on the C-terminal α -helix are required to recognize the neighboring trimer intermediates during self-assembly.¹⁵⁷ Therefore, these two amino acids should be preserved in the newly designed E2 protein with self-assembly capability. In order to maintain the spatial structure with constrained length of C-terminus at inter-trimer interfaces, as well as introduce functional GALA peptide with helix-to-coil transitions, EAALAEALA was incorporated to replace the original ELLLMEA (amino acids 421- 427) at the C-terminus to obtain recombinant E2 protein (E2-GALA). pET-11a plasmid containing wild-type E2 gene (E2-WT) was used as template to generate mutant E2 gene containing GALA sequence. Primer pairs 5'- gggaattc cat ATG CTG TCT G TTC CTG GTC CC -3' (forward) and 5'- gcggatcc TTA AGC CAG AGC TTC AGC CAG CGC CGC TTC CGG GTC GGA CAG CAG ACG T -3' (reverse) were used to PCR-amplify the E2-GALA gene (in capital letters), where the underlined bases represent the NdeI and BamHI restriction enzyme cutting sites. The amplified E2-GALA PCR product was purified by 0.8% DNA electrophoresis and then inserted into vector pET-11a using T4 DNA ligase. The sequence of the mutant gene was confirmed by DNA sequencing service from 1st Base (Singapore).

5.3.2 The Incorporation of GALA Affects E2 Assembly at pH 7.0

The mutant E2-GALA gene was constructed and overexpressed. The mutant protein was purified and characterized. Similar to wild-type E2 protein (E2-WT), E2-GALA is expressed in soluble fraction and has a high expression level (data not shown). Molecular mass determined by MALDI-TOF/TOF analysis was within 0.2% of the calculated theoretical value of the E2-GALA subunit (data not shown), indicating the correct expression of E2-GALA from the designed gene sequence.

E2-WT remains self-assembled 60-mer structure through a wide range of pH from 4.0 to 9.0.^{33, 42} In this work, the C-terminal α -helix substituted E2-GALA is present at distinct self-assembled manners. At pH 7.0, E2-GALA shows a hydrodynamic diameter of 8.05 ± 0.19 nm, which is comparable to the size of E2 trimer intermediate (8.01 ± 0.21 nm) measured in previous work.¹⁵⁷ SEC profile in Figure 5.2 also shows that the elution volume of E2-GALA at pH 7.0 is 13.8 ml, which corresponds to the protein species with molecular weight around 85 kDa. Since the theoretical molecular weight of trimeric structure of E2-GALA is 84.5 kDa, together with the E2 composition verified by SDS-PAGE (Figure 5.3), we deduce that E2-GALA is present as trimeric structure at pH 7.0. The original C-terminal α -helix plays critical role in associating the self-assembly of E2-WT protein from trimer intermediates into integrate 60-mer.¹⁵⁷ Removal of the α -helix region completely disrupts the inter-trimer interactions, and results in the presence of trimer intermediates.¹⁵⁷ At pH 7.0, the GALA peptide extends to random coil rather than α -helix. The extended form of GALA peptide at the C-terminus is no longer able to maintain the inter-trimer interactions which are required in associating E2 trimer structures. As a result, similar

to the effects of C-terminal truncation on assembly of E2 protein,¹⁵⁷ E2-GALA is present as trimer structures in solution at pH 7.0.

Figure 5.2 SEC profiles showing the oligomeric states of E2-GALA at different pH-s. E2-GALA elutes at 13.8, 16.1, and 8.7 ml at pH 7.0, 5.0, and 4.0, respectively.

Figure 5.3 SDS-PAGE verifies the E2 protein compositions of E2-GALA at different pH-s. E2-WT is used as a control.

In another work which reported similar substitution with GALA peptide at the C-terminal α -helix, the ferritin protein cage presented different self-assembly behavior from our E2-GALA. Choi *et al.* reported that human ferritin light chain still maintained assembled caged structure after incorporation of GALA at the C-terminus at physiological pH,²⁷ implying that the original C-terminus in ferritin was not essential in maintaining the association of its caged structure. The extended random coil of GALA peptide did not occupy and jeopardize the spatial structures at subunit interfaces. In contrast, the α -helical C-terminus guides the association of E2 protein from trimer to 60-mer. Hence, at pH 7.0, the essential interactions between trimers of E2 protein were eliminated by the random coil GALA peptide at the C-terminus.

To evaluate the folding and unfolding of trimeric E2-GALA at different pH-s, far-UV CD was performed. Approximately one third of the E2 protein secondary structures form α -helix as examined from its crystallographic structure (PDB file 1b5s). The α -helix rich E2-WT gives a characteristic CD profile with two minima at 208 and 222 nm.^{42, 115, 157} Figure 5.4 shows the α -helix rich CD profile of E2-GALA at pH 7.0 with 208 and 222 nm minima, indicating the correct folding of the trimer structure. The slight change of 208 nm minimum compared to E2-WT implies slight loss of α -helix content in E2-GALA (calculated with CDNN program; data not shown). The observation is in accordance to the design that GALA at C-terminus is not present as α -helix at pH 7.0. The evidence proves that the incorporation of GALA peptide at C-terminus only breaks the association among trimers rather than influencing the folding and formation of trimer intermediate.

Figure 5.4 CD spectra of E2-GALA at different pH-s with E2-WT as the control. E2-GALA shows characteristic profiles of α -helix rich protein with minima at 208 and 222 nm at all pH-s. At pH 4.0, the protein presents slight loss of α -helix.

5.3.3 pH-Responsive Self-Assembly of E2-GALA

To evaluate the pH-responsive self-assembly after GALA peptide incorporation, the E2-GALA was dialyzed and incubated in sodium phosphate buffer at pH 5.0 and 4.0. The oligomeric states at different pH-s were examined by SEC. Figure 5.2 shows that E2-GALA elutes out at 16.1ml at pH 5.0, which corresponds to molecular mass of 28 kDa. Since the theoretical molecular mass of the E2-GALA subunit is 28.1 kDa, together with the SDS-PAGE verifying the presence of E2 subunit (Figure 5.3), we conclude that E2-GALA is present as monomer at pH 5.0. Our previous work demonstrated the importance of C-terminal α -helix in the self-assembly of E2 protein.¹⁵⁷ At pH 5.0, the GALA peptide is supposed to form α -helix at the C-terminus of E2-GALA and induces further assembly of the trimer intermediate to the fully assembled 60-mer. However, the self-assembly of E2-GALA into 60-mer was

not observed, which is expected to be mediated by the C-terminal α -helix. Unexpectedly, the trimer structure, which is present at pH 7.0, further dissociates into monomers at pH 5.0. It has reported that the original histidine clusters at intra-trimer interfaces are able to break the trimer structure of E2 protein due to the repulsive interactions from their protonated imidazole groups at acidic pH.³³ Although the embracing N-terminus at the exterior surface and the loop structure at the interior surface prevent solvent access to the intra-trimer interfaces,⁴² the dynamic transitions between trimer and monomer structures are also reported to exist for E2 protein.¹⁵⁷ As a result, the protonation of histidines resulting from access of acidic buffer is seemed to be the most possible cause for the observed dissociation of trimeric structures into monomeric E2-GALA at pH 5.0. Interestingly, when E2 protein presents as fully assembled 60-mer, the intra-trimer interface is inaccessible to acidic solvent.^{33, 42} In this work, the mechanism of conformational change which leads to the different solvent access to intra-trimer interfaces between 60-mer and trimer E2 proteins has not been investigated. To further confirm the hypothesis that self-assembly triggered by C-terminal α -helix is required to initiate solvent access protection at intra-trimer interfaces, an experiment involving C-terminal truncated E2 protein (described in Chapter 4) can be performed. Incubation and characterization of E2- Δ C9 incubated at pH 7.0, 5.0, and 4.0 as performed on E2-GALA will provide more insights into the solvent accessibility profile and its impact on the oligomeric states.

Figure 5.4 shows nearly overlapping CD spectra for E2-GALA at pH 7.0 and 5.0, implying that there is no apparent secondary structure change when pH is changed from 7.0 to 5.0 (calculated with CDNN program; data not shown). The pH change results only in complete dissociation of E2 trimer structure, while the folding of E2-GALA is not influenced.

5.3.4 E2-GALA Self-assembles at pH 4.0

To further evaluate the self-assembly of E2-GALA in response to pH change, the pH value of incubation buffer was lowered to 4.0. The GALA peptide we used (EAALAEALA) in this work is shorter than that in previous work (LAEALAEHLAEALAE),²⁷ so the pH was decreased to enhance the formation of α -helix. At pH 4.0, most of E2-GALA elute at 8.7 ml, which is the void volume (V_0) of SEC column used in this work (Figure 5.2). Small portion of the protein elutes at 16.2 ml, indicating the presence of monomeric state. Interestingly, the SEC profile indicates that the majority of E2-GALA forms large assembly with molecular weight larger than 600 kDa which is the separation limit of the column. DLS analysis shows that the large assembly has a hydrodynamic diameter of 24.32 ± 0.43 nm (Figure 5.5), which is comparable to the size of correctly assembled 60-mer E2-WT.

Figure 5.5 Representative DLS scans show the hydrodynamic diameters of E2-GALA at different pH-s. E2-GALA has diameter of about 8 nm at pH 7.0 (blue) and pH 5.0 (red), and about 25 nm at pH 4.0 (black).

TEM was performed to examine the morphology of the assembled E2-GALA structures. The electron micrograph in Figure 5.6 presents the small diameters of trimer (Figure 5.6A) and monomer (Figure 5.6B) structures at pH 7.0 and 5.0, respectively. At pH 4.0, some E2-GALA form spherical hollow structures (Red circles in Figure 5.6C). The dimensions and the symmetries of the caged structure at the two-, three, and five-fold axis are comparable to those previously observed of correctly assembled wild-type and mutant E2 proteins.^{33, 42, 115} The observation indicates that self-assembly to form 60-mer spherical structures is occurring for E2-GALA at pH 4.0. However, we still observe some irregularly assembled E2-GALA structures (blue ovals in Figure 5.6C). CD scan shows molar ellipticity minima at 208 and 222 nm, indicating the rich content of α -helix secondary structure of E2-GALA at pH 4.0 (Figure 5.4). The slightly stronger 208 minimum compared to the spectra at pH 7.0 and 5.0 indicates a slight loss of α -helix. Although GALA peptide is supposed to form stronger α -helix at pH 4.0, the reassembly of E2 subunits results in both spherical 60-mers and irregular assemblies. Therefore, the loss of α -helix may result from conformational changes during the formation of the irregular assemblies.

Figure 5.6 Electron micrographs of E2-GALA at different pH-s: (A) pH 7.0 show trimer structures (red arrows); (B) pH 5.0 show monomer structures (red arrows); (C) pH 4.0 show correct assembled 60-mer (red circles) and irregular assemblies (blue ovals). Proteins were stained with 1.5% uranyl acetate. Scale bars are 50 nm.

When the pH is lowered to 4.0, the GALA peptide at C-terminus is speculated to continuously fold to form an α -helix, which is spatially similar to the original C-terminal α -helix in E2-WT. The formation of this appropriate α -helix is a sign to trigger self-assembly of E2 protein. The monomer, which is formed at pH 5.0, is associated into trimer intermediate, followed by the formation of spherical 60-mers or irregular assemblies. At this instance, due to the protection from N-terminal arms at the exterior surface and amino acids loops at the interior surfaces, the intra-trimer interfaces is no longer accessible to the acidic solvent upon formation of the 60-mer. Since the only monomer, instead of trimer, is present at pH 5.0, we speculate that self-assembly triggered by the C-terminal α -helix is required to initiate this solvent access protection. However, although both GALA peptide and original E2 C-terminus form similar α -helices, the sequence difference between them still cause conformational change during subunit associations, and results in imperfect self-assembly of E2-GALA at pH 4.0. The undesired interactions may happen among GALA peptide and

other motifs on any E2 subunits, resulting in the disruption of self-assembly process or association of several assembly intermediates. Hence, the irregularly shaped E2-GALA assemblies are present. The apparent loss of alpha-helix (Figure 5.4) may result from conformational changes during the formation of the irregular assemblies.

5.3.5 Reversible pH-responsive self-assembly

To evaluate the reversibility of this pH-triggered self-assembly of E2-GALA, the ability to dissociate from 60-mer to trimer was investigated. The protein sample at pH 4.0 was neutralized to pH 7.0 by buffer-exchange. The neutralized sample was subjected to SEC using the same setting as previous experiments. The elution volume of 13.9 ml indicates E2-GALA dissociate back into trimers at pH 7.0 (Figure 5.7A). DLS result suggests that the size of neutralized E2-GALA decreases to 7.99 ± 0.16 nm, which is comparable to the size of trimer structure. CD spectra show the profile with 208 and 222 nm minima of re-formed trimer structure, while the 208 nm minimum is recovered compared to the E2-GALA at pH 4.0 (Figure 5.7B). As the pH increases to 7.0, random coil forms at the C-terminal GALA peptide, and eliminate the association interactions between trimer structures. As a result, 60-mer and irregular assemblies can no longer be maintained. We observe the oligomeric state of trimer. The results tell us the self-assembly of E2-GALA controlled by pH change is reversible.

Figure 5.7 Reversibility analysis of E2-GALA in response to pH change. (A) SEC profiles comparing the elution volumes at pH 4.0 and pH 7.0 which is buffer-exchanged from pH 4.0. (B) CD spectra comparing the secondary structure change, indicating the recovery of 208 minimum at pH 7.0 compared to pH 4.0.

5.4 Conclusions

In summary, we design a non-native, reversible pH-responsive E2 protein cage from *B. stearothermophilus* with inversed self-assembly behavior. A GALA peptide is incorporated at the C-terminus of E2 protein to replace the original α -helix, which guides the self-assembly of E2-WT from trimers to 60-mer. The pH-inducible coil-to-helix feature of GALA peptide is used to control the self-assembly of the engineered protein. While the native E2 protein cage remain fully assembled 60-mer structures at a wide pH range from 4.0 to 9.0, the GALA incorporated E2 protein presents pH-responsive self-assembly behaviors. At neutral pH, the extended random coil of GALA leads to dissociation of protein cage into trimer structures. At pH 5.0, the trimer further dissociates into monomers. When the pH is lowered to 4.0, the formation of GALA α -helix triggers the self-assembly process and we observe the formation of fully assembled 60-mer. The partial irregular assemblies may result from the sequence difference between the GALA and the original C-terminus. The most notable behavior of the E2-GALA is the reversible self-assembly and disassembly simply by changing the pH. The engineered E2 protein cage has potential application as pH-inducible molecular switch. The work also provide clues to understand and control the self-assembly of other biomacromolecules.

Chapter 6

Designing Non-native Iron-Binding Site on E2 Protein Cage

6.1 Abstract

In biomineralization process, supramolecular organic structure is often used as a template for inorganic nanomaterial synthesis. The E2 protein cage of pyruvate dehydrogenase from *B. stearothermophilus* has been functionalized with non-native iron-mineralization capability by incorporating two types of iron-binding peptides. The non-native peptides introduced at the interior surface do not affect the self-assembly of E2 protein. In contrast to the wild-type, the mutant E2 proteins can serve as size- and shape- constrained reactors for the synthesis of iron nanoparticles. Electrostatic interactions between anionic amino acids and cationic iron molecules drive the formation of iron oxide within the mutant E2 protein cages. The work expands the investigations on nanomaterial synthesis using inherent host-guest properties of protein cage.

6.2 Introduction

The synthesis of inorganic nanoparticles using protein cage templates is currently of interest in material chemistry and bionanotechnology. These caged-like proteins are formed by self-assembly of a number of subunits into defined uniform spherical hollow structures. In most cases, the nanoparticles are formed naturally or synthetically within the protein cages, such as ferritins and virus capsids,^{17, 92, 160-95, 98, 161, 162}, either on the exterior or interior surfaces. The well-defined exterior surfaces can be used as templates for controlled nanoparticle attachments while the interior cavity can work as shape- and size- constrained reactors for nanoparticles synthesis.

Ferritins are naturally existing iron storage proteins found in most living organisms.^{3, 4} They convert ferrous iron to ferric complexes that mineralize in their internal cavities. The constrained sizes of the internal cavities result in the formation of iron nanoparticles with uniform narrow size distributions.^{11, 90, 92, 150} Researchers have found that several amino acids, which are highly conserved among all species, form the dinuclear ferroxidase sites and catalyze the iron nucleation reactions in ferritins.³ Besides ferritin, other protein cages have been engineered for nanomaterial synthesis. For example, some virus-like particles (VLP) were engineered with anionic interior surfaces. The electrostatic interactions between anionic VLPs and cationic inorganic ions triggered the formation of nanoparticles within the protein cages.¹⁷ In other cases, chemical interactions between nanoparticles and thiol or ϵ -amino groups on VLPs are adopted to form confined pattern of inorganic materials.^{88, 101}

The hollow caged structure (Figure 6.1A) with superior stability suggests the possibility of using E2 protein as size-constrained reaction vessels for inorganic nanoparticle synthesis. The interior surface of E2 protein could be modified to incorporate different functionalities. Moreover, since the E2 protein comprises of 60 identical subunits, one modification made on the subunit will result in the same 60 modifications on the protein cage. In previous works, substitution of an amino acid located at the internal cavity with cysteins allows encapsulation of fluorescent dye and drug molecules through interactions with the thiol groups.^{115, 118} Thus far, there is no report on the use E2 protein as a template for inorganic nanoparticle synthesis.

Figure 6.1 PyMol representation of E2 protein cage for iron mineralization. (A) Self-assembled E2 protein showing the pores on the surface. (B) Single E2 subunit with RDGE loop highlighted in red. (C) Overview of half E2 protein cage and 30 iron-binding sites on the interior surface highlighted in red.

In this work, we engineer a ferritin-like catalytic domain in the inner cavity of the E2 protein to impart a non-native iron mineralization capability. The functional iron binding peptides are incorporated into the appropriate sites on the interior surface of E2 protein, while the caged structure is maintained. Iron molecules will diffuse into

the cavity through the 12 openings on the protein surface. The expected size of the mineral core is 12 nm which corresponds to the size of the inner cavity.

6.3 Results and Discussions

6.3.1 Design and Construction of Ferritin-like Catalytic Domain in E2 Protein

To design functional iron-binding peptides, the catalytic sites of ferritins from previous works were investigated. The mineralization functions of ferritins are believed to be controlled by relatively conserved amino acids. By visualizing the crystallographic structure, the diiron catalytic center of frog M-ferritin was reported to form from an amino acid motif containing Gln137, Glu23, His61, Glu58, Asp140, and Glu103 (QEHEDE)^{163, 164}. Other works found that the high negative charge density of amino acid clusters, such as Glu, at the interior surface of VLPs or synthetic ferritin could also work as active sites for synthetic mineral nucleation^{17, 96, 160}.

Visualization of the crystal structure of E2 protein (Protein data bank file: 1b5s) using PyMol¹³⁵ revealed a floppy RDGE loop (amino acid 380-383) structure on each subunit located at the interior surface (Figure 6.1B and 6.1C). The non-formation of any secondary structure suggested that the RDGE loop might not be essential in maintaining the assembled E2 protein cage. Therefore, it was the ideal site to introduce the iron-binding peptide. Frog M-ferritin-mimicking iron-binding peptide (QEHEDE) or negative charged Glu peptides (EEEEEE) was incorporated to substitute the original RDGE loop structure and resulted in two mutant proteins: E2-LFer or E2-LE6. Since 60 iron-binding peptides would be present on the fully assembled 60-mer protein, the recombinant E2 proteins were expected to be functionalized with iron-mineralization capabilities.

Mutant plasmids were generated using site-directed mutagenesis. Oligonucleotides 5'- GAA AAG CCG ATC GTT **CAG GAA CAT GAA GAT GAA** ATC GTT GCT GCT CC -3' for E2-LFer or 5'- GAA AAG CCG ATC GTT **GAA GAA GAA GAA GAA GAA** ATC GTT GCT GCT CC -3' for E2-LE6 was used to PCR-amplify the mutant plasmid upon pET-11a plasmid containing wild-type E2 gene.

6.3.2 E2 Proteins Assembled Correctly upon Incorporation of Iron-Binding Peptides

To evaluate the correct assemblies of mutant proteins, purified E2-LFer and E2-LE6 were characterized by DLS and TEM. E2-LFer and E2-LE6 self-assemble into approximately 26.86 ± 0.64 and 26.13 ± 0.34 nm diameter particles as indicated by DLS (Figure 6.2A). The diameters are comparable to the correctly assembled wild-type E2 protein (E2-WT) with a diameter of 25~28 nm.^{42, 115} TEM results also confirm the correctly assembled spherical structures for both mutant proteins with diameters of approximately 25 nm (Figure 6.2B and 6.2C). The results imply that both mutant proteins correctly self-assemble into spherical structures after substituting the original RDGE loops with 6 functional amino acids. The RDGE loop at the interior surface is amenable while maintaining the correctly assembled structure of E2 protein. The presence of monodisperse caged structure is a prerequisite for the templated size-constrained synthesis of nanoparticles. The presence of 60 iron-binding peptides on the interior surfaces of both E2-LFer and E2-LE6 will act as reactive sites.

Figure 6.2 Correct assemblies of E2-LFer and E2-LE6 before iron-loading indicated by DLS and TEM. (A) Overlaid DLS profiles showing the sizes of both mutant proteins are comparable to the size of E2-WT. Electron micrographs of (B) E2-LFer and (C) E2-LE6 indicating the correctly assembled caged structures. Proteins were stained with 1.5% uranyl acetate. Scale bars are 50 nm.

6.3.3 Iron Mineralization does not affect the Protein Structures

The purified E2-LFer and E2-LE6 were loaded with ferrous irons in HEPES buffer. In the presence of either E2-LFer or E2-LE6, the reaction proceeds to form homogenous yellow-coloured solution. In contrast, just like the solution without any protein, adding ferrous iron to E2-WT solution results in dark yellow color followed by precipitations (Figure 6.3). The results suggest that E2-WT lacks the ability to mineralize irons in its constrained inner architecture. Therefore, the unconstrained iron oxidation was observed in the presence of E2-WT, leading to the formation of

rusty precipitate. The lack of precipitation in the E2-LFer and E2-LE6 reactions, together with the clear yellow color of the solutions, are consistent with the fact that the protein-cage dependent oxidation of ferrous irons has occurred in a spatially selective manner to form soluble iron oxide.

Figure 6.3 Colors of E2 protein solutions before and after iron-loading. E2-WT presents precipitations as that of HEPES buffer due to uncontrolled iron oxidation; E2-LFer and E2-LE6 show clear yellow color solutions due to protein-dependent iron mineralization.

To evaluate the effect of iron mineralization on protein assembly, DLS was performed. After mineralization, the mean diameters as determined using DLS are 27.13 ± 0.69 and 26.88 ± 0.73 nm for E2-LFer and E2-LE6, respectively. There are no noticeable changes of protein sizes compared to the empty mutant E2 proteins. The DLS graphs also show agreements between the distribution profiles (Figure 6.4).

Nanoparticles with other sizes are not observed, indicating the formation of iron nanoparticles are closely coupled to the protein cages. The morphologies and integrities of iron-mineralized mutant proteins were also confirmed by TEM. Upon negative staining with uranyl acetate, both E2-LFer and E2-LE6 present intact caged structures (Figure 6.5A and 6.5C) which are comparable to their correctly self-assembled spherical structures before iron-loading in Figure 6.2B and 6.2C. The accumulation and mineralization of iron using mutant E2 proteins as templates have no influence on the correct self-assemblies of E2-LFer and E2-LE6.

Figure 6.4 Comparisons of DLS profiles for (A) E2-LFer and (B) E2-LE6 before and after iron-loading; Correlograms are shown in the right. The overlaid size distributions indicating the unchanged protein sizes upon iron mineralization for both mutant E2 proteins.

Figure 6.5 Electron micrographs of iron-mineralized mutant E2 protein cages. Negative-stain samples of (A) E2-LFer and (C) E2-LFer indicating correctly assembled caged protein structures. Unstained samples of (B) E2-LFer and (D) E2-LE6 indicating discrete iron nanoparticles with sizes ranging from 3 to 6 nm.

6.3.4 Iron Mineralization within E2 Protein Cages

The iron oxide nanoparticles are imaged by TEM using unstained mineralized protein samples (Figure 6.5B and 6.5D). The iron forms mineralized nanoparticles within mutant protein cages with sizes ranging from 3 to 6 nm. The observations demonstrate the size-controlled mineralization within the mutant E2 protein cages. However, the sizes of formed nanoparticles are much smaller than the size of E2 inner

cavity of 12 nm. In ferritin, the oxidized iron core with the protein cage has diameter comparable to the size of ferritin's inner cavity of 8 nm.¹⁵⁰ We speculate that the 12 opening pores with 5 nm each on the E2 protein cage limit the net interactions among those functionalized iron-binding sites on the interior surface. E2 protein assembles from trimer intermediate as building block,¹⁵⁷ while the pore is surrounded by 5 trimer structures seen from 5-fold axis (Figure 6.1C). The irons are oxidized at the iron-binding peptides, but could not be further accumulated to larger nanoparticles, which may be due to the interference from the big pores (Appendix A.2). E2 protein is self-assembled using trimer structures as building block, and each pore on the surface is surrounded by 5 trimer structures. Hereby, after substitution of RDGE loops, we get the distribution of iron-binding peptides on the interior surface shown in Figure A.2.2. The distance between iron-binding loops within the trimer is around 2 nm, while the distance between neighbouring trimers is around 7 nm. Because of the existence of the large pore, we speculate the connection of iron mineralization from more than 3 trimer structures is affected. As a result, the iron nanoparticles formed within E2-LFer and E2-LE6 cages have diameters around 3-6 nm, which are located in the range of 2-7 nm.

Size exclusion chromatography was conducted to analyze the composite nature of the mineralized E2-LFer and E2-LE6. The UV-vis spectrums reveal that the mineralized proteins present a broad absorption band centered at around 350 nm other than the 280 nm peak due to the formation of the iron core (Appendix A.2). Hence, elution of each mineralized proteins from SEC was monitored at both 280 nm and 350 nm which correspond to the absorbance of the protein cage and the iron core,

respectively. Both profiles of E2-LFer and E2-LE6 show a co-elution of the protein cage and the iron core (Figure 6.6), indicating the protein-mineral composite.

Figure 6.6 SEC profiles indicating co-elutions of protein cages and mineralized iron cores: (A) E2-LFer, and (B) E2-LE6.

6.3.5 Mutant E2 Proteins Show High Iron-Binding Capacities

To determine the iron mineralization capacities for E2-LFer and E2-LE6, the mean iron content in each protein cage was quantified using ICP. At loading ratio higher than 3000 iron/protein cage for both mutant E2 proteins, we started to observe precipitation of proteins and ferric irons. Therefore, we hypothesized that up to 3000 iron could be loaded into E2-LFer and E2-LE6. After removal of unbound irons the amount of iron loading was determined to be 2700 iron/E2-LFer cage and 2600 iron/E2-LE6 cage. Frog M-ferritin-mimicking E2-LFer and 6-Glu incorporated E2-LE6 show no distinct differences on iron mineralization capacities. Ferritin is reported to be able to mineralize 5000 iron in its 8 nm inner cavity.^{11, 150} However, compared to ferritin, we observed much less iron mineralized within E2-LFer and E2-LE6 which have larger inner cavity of 12 nm. The ferritin outstanding mineralization capability

requires not only the particular conserved amino acids, but also the need for these amino acids to form specific spatial motif.^{90, 164, 165} The replacement of the RDGE loop with M-ferritin-mimicking iron-binding peptides might form a floppy structure instead of any fixed spatial structures. As a result, E2-LFer cannot present mineralization capacity as high as the ferritin. Previous researches have shown that the relationship between protein cage and mineralization was based primarily on complementary electrostatic interactions.^{17, 96, 166} The regions of high charge density on the interior surfaces of protein cages facilitated synthetic mineral nucleation. The negatively-charged QEHEDE on E2-LFer and EEEEEEE on E2-LE6 act as active sites for iron oxidative mineralization and present similar reactive capabilities.

To evaluate the stabilities of mineralized iron cores within E2-LFer and E2-LE6, the contents of iron in each protein cage at day 1, day 7, and day 14 from the same batch were analysed by ICP and protein BCA kit. We observe that the iron content in each protein cage remain relatively constant through day 14 for both E2-LFer and E2-LE6. However, we started to observe the presence of yellow cloudy precipitations from day 16 onwards for both proteins. As a comparison, the empty mutant proteins at the same condition and the same protein concentrations are present as clear solutions until day 16. The iron mineralization within the protein cage influences the stability of the latter for long-term storage. For both mutants E2-LFer and E2-LE6, the mineralized irons together with the protein cage remain stable in solution for up to 14 days at 4 °C.

More detailed characterizations on the magnetic property, oxidation state, and the crystallinity of the iron mineral core can be investigated in the future.

6.4 Conclusion

Two functional iron-binding peptides (Frog M-ferritin-mimicking and 6-Glutamic acids) have been incorporated into the interior surface of E2 protein without influencing its self-assembled caged structure. Ferritin-mimicking peptide incorporated into E2 protein does not perform the same function as ferritin does, which may result from the inability to form particular spatial structures in E2 protein. Iron molecules diffuse into the inner cavities of both mutant E2 proteins through the pores on the surfaces. Complementary electrostatic interactions trigger the mineralization of cationic irons at anionic iron-binding peptides and lead to the formation of iron nanoparticles in the inner cavity of E2 proteins. The results are confirmed by DLS, SEC, and negative-stain TEM. The inner cavity of E2 protein acts as size- and shape- constrained vessels for iron oxidation to form nanoparticles with sizes of 3-6 nm. However, the existence of the pores on the surface affects the accumulation of iron oxide to form large nanoparticles. Our investigation suggests the possibility to use E2 protein cage for inorganic nanomaterial synthesis by introducing non-native functionalities.

Chapter 7

Conclusions and Future Directions

7.1 Conclusions

The long-term objective of our research on E2 protein is to design a multi-functional nanoscaffold, which can be used as a novel nanocapsule for drug delivery and as nanoreactor for nanoparticle synthesis for possible application as a MRI contrast agent. Based on the previously established E2 core protein model of pyruvate dehydrogenase from *B. stearrowthermophilus*, we have investigated its self-assembly mechanism and explored its potential functionalities in this dissertation.

The established data in previous works support the function of E2 protein as drug delivery vehicle. To facilitate future applications, the controlled release of drugs from E2 protein is another important aspect to be considered. It is desirable for the fully assembled E2 protein cage to open or close in response to particular environmental cues, such as pH change. E2 protein provides well-defined surfaces (exterior, interior, and subunit interface) which results from precise self-assembly from 60 subunits. Modifications made on these amenable surfaces provide feasible solutions to influence the self-assembly of E2 protein.

Our results demonstrate that the integrated 60-mer E2 protein is self-assembled from 20 trimers intermediate, which is mediated by the protein C-terminus. The isolation of trimer structures by C-terminus truncation indicates the amenable features of inter-trimer interface where the C-termini of two intermediate structures interact (described in Chapter 4).

Based on understanding the amenable inter-trimer interactions, histidine pairs are introduced at the inter-trimer interfaces to engineer E2 protein with a pH-responsive

disassembly profile (described in Chapter 3). The engineered protein maintains fully assembled structure at pH 7.4, but dissociates and aggregates to soluble larger assemblies at pH 5.0. This characteristic will facilitate drug delivery through cell endocytosis, which is proven to effectively take in E2 protein cages.

Since the C-terminus mediate the self-assembly of E2 protein, we further manage to control the self-assembly of E2 protein by controlling the conformational change of C-terminus (described in Chapter 5). The secondary structure of artificial GALA peptide is pH-inducible, which presents reversible helix-to-coil transitions when pH is adjusted between neutral and acidic. Replacement of the original rigid C-terminal α -helix with GALA peptide results in reversible pH-responsive E2 protein cage with inverted self-assembly behavior. The protein is present as trimer structure at neutral pH, but self-assembled at pH 4.0. Moreover, the disassembly and assembly process is reversible.

Other than the work on understanding the self-assembly mechanism, we are also exploring the introduction of more functionalities to the caged structure of E2 protein. For example, we use the defined inner cavity of E2 protein for iron nanoparticle synthesis (described in Chapter 6). Iron-binding peptides are incorporated at the interior surface of E2 protein without influencing the fully assembled caged structure. The mutant proteins serve as size- and shape- constrained nanoreactors for iron mineralization through complementary electrostatic interactions.

Our work suggests the potential of using E2 protein as a powerful platform for biomedical applications. The investigation on subunit-subunit interactions also set groundwork to understand the interactions among other macromolecules.

7.2 Future Directions

The results presented in this dissertation support the E2 protein as promising multi-functional scaffold. However, some improvements and future investigations need to be done.

We have engineered soluble pH-responsive disassembly of E2 protein cage (described in Chapter 3). To confirm similar behavior in the cellular environment, *in vitro* studies can be performed to test its drug release in the endocytosis pathway. Drug binding groups can be incorporated onto the interior surface to make the multi-functional E2 protein capable of drug loading. We anticipate seeing the rapid release of drugs within cells after their internalizations.

By introducing iron-binding peptides at the interior surface of E2 protein, iron molecules accumulate and form nanoparticles within the inner cavity (described in Chapter 6). However, to obtain nanoparticles with more uniform size distribution, the iron-binding reactions can be improved by modifying some of the parameters. For example, pH value can be increased to enhance the negative charges on protein inner surface while remaining assembled structure, thus to enhance more effective iron binding reactions. Since complementary electrostatic interactions are the main driving forces for the synthesis of inorganic materials, we hypothesize that iron is not the only potential metal ion for biomineralization within our mutant E2 protein. Other cationic metal ions, such as gadolinium, manganese can be introduced as guest molecules that may interact with the anionic ‘iron-binding peptide’. The feasibility of loading various

types of inorganic metals provides possibilities of using E2 protein as contrast agent in biomedical applications.

The exterior surface of E2 protein where its N-terminus located can be modified to impart different functionalities. The N-terminus (47 amino acids) is not required in maintaining the fully assembled structures of E2 protein. Therefore, other functional ligands can be incorporated onto the exterior surface by replacing the original N-terminus. The advantage of E2 protein may be that it can tolerate the substitution at exterior surface with long amino acid peptide. For example, a 33-amino-acid long IgG-binding peptide from protein A can be engineered at the N-terminus, resulting in a construct capable of multiple antibody attachments for antibody-mediated targeting of the E2 protein cage for drug delivery usage. Specific metal-binding peptide can also be located at the N-terminus to facilitate nanoparticle patterning on E2 protein exterior surface.

References

1. Caspar, D. L. D. Citation Classic - Physical Principles in the Construction of Regular Viruses. *Current Contents/Life Sciences* **1984**, 15-15.
2. Lidmar, J.; Mirny, L.; Nelson, D. R. Virus shapes and buckling transitions in spherical shells. *Physical Review E* **2003**, 68.
3. Theil, E. C. Ferritin: Structure, Gene-Regulation, and Cellular Function in Animals, Plants, and Microorganisms. *Annual Review of Biochemistry* **1987**, 56, 289-315.
4. Chasteen, N. D. Ferritin. Uptake, storage, and release of iron. *Metal Ions in Biological Systems, Vol 35* **1998**, 35, 479-514.
5. Walter, S.; Buchner, J. Molecular chaperones - Cellular machines for protein folding. *Angewandte Chemie-International Edition* **2002**, 41, 1098-1113.
6. Uchida, M.; Klem, M. T.; Allen, M.; Suci, P.; Flenniken, M.; Gillitzer, E.; Varpness, Z.; Liepold, L. O.; Young, M.; Douglas, T. Biological containers: Protein cages as multifunctional nanoplateforms. *Advanced Materials* **2007**, 19, 1025-1042.
7. Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M. A.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G.; Thomas, C. D.; Shaw, W. V.; Harrison, P. M. Solving the Structure of Human H-Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature* **1991**, 349, 541-544.
8. Hempstead, P. D.; Yewdall, S. J.; Fernie, A. R.; Lawson, D. M.; Artymiuk, P. J.; Rice, D. W.; Ford, G. C.; Harrison, P. M. Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *Journal of Molecular Biology* **1997**, 268, 424-448.

9. Wardrop, A. J.; Wicks, R. E.; Entsch, B. Occurrence and expression of members of the ferritin gene family in cowpeas. *Biochemical Journal* **1999**, 337, 523-530.
10. Macedo, S.; Romao, C. V.; Mitchell, E.; Matias, P. M.; Liu, M. Y.; Xavier, A. V.; LeGall, J.; Teixeira, M.; Lindley, P.; Carrondo, M. A. The nature of the di-iron site in the bacterioferritin from *Desulfovibrio desulfuricans*. *Nature Structural Biology* **2003**, 10, 285-290.
11. Uchida, M.; Kang, S.; Reichhardt, C.; Harlen, K.; Douglas, T. The ferritin superfamily: Supramolecular templates for materials synthesis. *Biochimica Et Biophysica Acta-General Subjects* **2010**, 1800, 834-845.
12. Zhang, Y.; Orner, B. P. Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. *International Journal of Molecular Sciences* **2011**, 12, 5406-5421.
13. Zhu, Y. X.; Carragher, B.; Kriegman, D. J.; Milligan, R. A.; Potter, C. S. Automated identification of filaments in cryoelectron microscopy images. *Journal of Structural Biology* **2001**, 135, 302-312.
14. Douglas, T.; Young, M. Viruses: Making friends with old foes. *Science* **2006**, 312, 873-875.
15. Singh, P.; Prasuhn, D.; Yeh, R. M.; Destito, G.; Rae, C. S.; Osborn, K.; Finn, M. G.; Manchester, M. Bio-distribution, toxicity and pathology of cowpea mosaic virus nanoparticles in vivo. *Journal of Controlled Release* **2007**, 120, 41-50.
16. Manchester, M.; Singh, P. Virus-based nanoparticles (VNPs): Platform technologies for diagnostic imaging. *Advanced Drug Delivery Reviews* **2006**, 58, 1505-1522.
17. Douglas, T.; Young, M. Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **1998**, 393, 152-155.

18. Gillitzer, E.; Suci, P.; Young, M.; Douglas, T. Controlled ligand display on a symmetrical protein-cage architecture through mixed assembly. *Small* **2006**, *2*, 962-966.
19. Singh, P.; Destito, G.; Schneemann, A.; Manchester, M. Canine parvovirus-like particles, a novel nanomaterial for tumor targeting. *Journal of Nanobiotechnology* **2006**, *4*, 2.
20. Hooker, J. M.; Kovacs, E. W.; Francis, M. B. Interior surface modification of bacteriophage MS2. *Journal of the American Chemical Society* **2004**, *126*, 3718-3719.
21. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223-230.
22. van den Berg, B.; Wain, R.; Dobson, C. M.; Ellis, R. J. Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. *Embo Journal* **2000**, *19*, 3870-3875.
23. Deutscher, M. P., Precipitation technique. In *Guide to Protein Purification*, Vol. 182, pp 285-300.
24. Chandler, D.; Hagan, M. F. Dynamic pathways for viral capsid assembly. *Biophysical Journal* **2006**, *91*, 42-54.
25. Fan, R. L.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P. A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009**, *48*, 5623-5630.
26. Luzzago, A.; Cesareni, G. Isolation of point mutations that affect the folding of the H chain of human ferritin in E.coli. *The EMBO Journal* **1989**, *8*, 569-576.
27. Choi, S. H.; Choi, K.; Kwon, I. C.; Ahn, H. J. The incorporation of GALA peptide into a protein cage for an acid-inducible molecular switch. *Biomaterials* **2010**, *31*, 5191-5198.

28. Levi, S.; Luzzago, A.; Cesareni, G.; Cozzi, A.; Franceschinelli, F.; Albertini, A.; Arosio, P. Mechanism of Ferritin Iron Uptake - Activity of the H-Chain and Deletion Mapping of the Ferro-Oxidase Site - a Study of Iron Uptake and Ferro-Oxidase Activity of Human-Liver, Recombinant H-Chain Ferritins, and of 2 H-Chain Deletion Mutants. *Journal of Biological Chemistry* **1988**, 263, 18086-18092.
29. Tang, J. H.; Johnson, J. M.; Dryden, K. A.; Young, M. J.; Zlotnick, A.; Johnson, J. E. The role of subunit hinges and molecular "switches" in the control of viral capsid polymorphism. *Journal of Structural Biology* **2006**, 154, 59-67.
30. Larson, S. B.; Lucas, R. W.; McPherson, A. Crystallographic structure of the T=1 particle of brome mosaic virus. *Journal of Molecular Biology* **2005**, 346, 815-831.
31. Hsu, C.; Singh, P.; Ochoa, W.; Manayani, D. J.; Manchester, M.; Schneemann, A.; Reddy, V. S. Characterization of polymorphism displayed by the coat protein mutants of tomato bushy stunt virus. *Virology* **2006**, 349, 222-229.
32. Dalmau, M.; Lim, S. R.; Wang, S. W. Design of a pH-Dependent Molecular Switch in a Caged Protein Platform. *Nano Letters* **2009**, 9, 160-166.
33. Dalmau, M.; Lim, S.; Wang, S. W. pH-Triggered Disassembly in a Caged Protein Complex. *Biomacromolecules* **2009**, 10, 3199-3206.
34. Kim, S.; Kim, J. H.; Jeon, O.; Kwon, I. C.; Park, K. Engineered polymers for advanced drug delivery. *European Journal of Pharmaceutics and Biopharmaceutics* **2009**, 71, 420-430.
35. Singh, P.; Gonzalez, M. J.; Manchester, M. Viruses and their uses in nanotechnology. *Drug Development Research* **2006**, 67, 23-41.

36. Sahoo, S. K.; Labhasetwar, V. Nanotech approaches to delivery and imaging drug. *Drug Discovery Today* **2003**, 8, 1112-1120.
37. Pannier, A. K.; Shea, L. D. Controlled release systems for DNA delivery. *Molecular therapy : the journal of the American Society of Gene Therapy* **2004**, 10, 19-26.
38. Langer, R. New Methods of Drug Delivery. *Science* **1990**, 249, 1527-1533.
39. Christie, R. J.; Grainger, D. W. Design strategies to improve soluble macromolecular delivery constructs. *Advanced Drug Delivery Reviews* **2003**, 55, 421-437.
40. Hockel, M.; Vaupel, P. Tumor hypoxia: Definitions and current clinical, biologic, and molecular aspects. *Journal of the National Cancer Institute* **2001**, 93, 266-276.
41. Hu, Y.; Litwin, T.; Nagaraja, A. R.; Kwong, B.; Katz, J.; Watson, N.; Irvine, D. J. Cytosolic delivery of membrane-impermeable molecules in dendritic cells using pH-Responsive core-shell nanoparticles. *Nano Letters* **2007**, 7, 3056-3064.
42. Peng, T.; Lim, S. Trimer-Based Design of pH-Responsive Protein Cage Results in Soluble Disassembled Structures. *Biomacromolecules* **2011**, 12, 3131-3138.
43. Kataoka, K.; Kwon, G. S.; Yokoyama, M.; Okano, T.; Sakurai, Y. Block-Copolymer Micelles as Vehicles for Drug Delivery. *Journal of Controlled Release* **1993**, 24, 119-132.
44. Gaucher, G.; Dufresne, M. H.; Sant, V. P.; Kang, N.; Maysinger, D.; Leroux, J. C. Block copolymer micelles: preparation, characterization and application in drug delivery. *Journal of Controlled Release* **2005**, 109, 169-188.
45. Schaeffer, H. E.; Krohn, D. L. Liposomes in Topical Drug Delivery. *Investigative Ophthalmology & Visual Science* **1982**, 22, 220-227.

46. Foldvari, M.; Gesztes, A.; Mezei, M. Dermal Drug Delivery by Liposome Encapsulation - Clinical and Electron-Microscopic Studies. *Journal of Microencapsulation* **1990**, 7, 479-489.
47. Park, J. W. Liposome-based drug delivery in breast cancer treatment. *Breast Cancer Research* **2002**, 4, 93-97.
48. Gillies, E. R.; Frechet, J. M. J. Dendrimers and dendritic polymers in drug delivery. *Drug Discovery Today* **2005**, 10, 35-43.
49. De Jesus, O. L. P.; Ihre, H. R.; Gagne, L.; Frechet, J. M. J.; Szoka, F. C. Polyester dendritic systems for drug delivery applications: In vitro and in vivo evaluation. *Bioconjugate Chemistry* **2002**, 13, 453-461.
50. Patri, A. K.; Majoros, I. J.; Baker, J. R. Dendritic polymer macromolecular carriers for drug delivery. *Current Opinion in Chemical Biology* **2002**, 6, 466-471.
51. Herrero-Vanrell, R.; Refojo, M. F. Biodegradable microspheres for vitreoretinal drug delivery. *Advanced Drug Delivery Reviews* **2001**, 52, 5-16.
52. Soppimath, K. S.; Aminabhavi, T. M.; Kulkarni, A. R.; Rudzinski, W. E. Biodegradable polymeric nanoparticles as drug delivery devices. *Journal of Controlled Release* **2001**, 70, 1-20.
53. Qiu, Y.; Park, K. Environment-sensitive hydrogels for drug delivery. *Advanced Drug Delivery Reviews* **2001**, 53, 321-339.
54. Gupta, P.; Vermani, K.; Garg, S. Hydrogels: from controlled release to pH-responsive drug delivery. *Drug Discovery Today* **2002**, 7, 569-579.
55. Frandsen, J. L.; Ghandehari, H. Recombinant protein-based polymers for advanced drug delivery. *Chemical Society Reviews* **2012**, 41, 2696-2706.

56. Lin, Y. H.; MaHam, A.; Tang, Z. W.; Wu, H.; Wang, J. Protein-Based Nanomedicine Platforms for Drug Delivery. *Small* **2009**, 5, 1706-1721.
57. Haider, M.; Megeed, Z.; Ghandehari, H. Genetically engineered polymers: status and prospects for controlled release. *Journal of Controlled Release* **2004**, 95, 1-26.
58. Paclitaxel Albumin-stabilized Nanoparticle Formulation.
<http://www.cancer.gov/cancertopics/druginfo/nanoparticlepaclitaxel>
59. Kratz, F. Albumin as a drug carrier: Design of prodrugs, drug conjugates and nanoparticles. *Journal of Controlled Release* **2008**, 132, 171-183.
60. Urry, D. W.; Haynes, B.; Zhang, H.; Harris, R. D.; Prasad, K. U. Mechanochemical Coupling in Synthetic Elastomeric Polypeptides by Charge Modulation of an Inverse Temperature Transition. *Biophysical Journal* **1988**, 53, A453-A453.
61. Urry, D. W.; Harris, R. D.; Prasad, K. U. Chemical-Potential Driven Contraction and Relaxation by Ionic-Strength Modulation of an Inverse Temperature Transition. *Journal of the American Chemical Society* **1988**, 110, 3303-3305.
62. Kim, W.; Chaikof, E. L. Recombinant elastin-mimetic biomaterials: Emerging applications in medicine. *Advanced Drug Delivery Reviews* **2010**, 62, 1468-1478.
63. Meyer, D. E.; Chilkoti, A. Quantification of the effects of chain length and concentration on the thermal behavior of elastin-like polypeptides. *Biomacromolecules* **2004**, 5, 846-851.

64. D.L. Kaplan, C. M. M., S. Arcidiacono, S. Fossey, K. Senecal, W. Muller, Silk. In *Protein based materials*, K. McGrath, D. L. K., Ed. Birkhauser: Boston, 1998; pp 103-131.
65. Numata, K.; Kaplan, D. L. Silk-based delivery systems of bioactive molecules. *Advanced Drug Delivery Reviews* **2010**, 62, 1497-1508.
66. Numata, K.; Kaplan, D. L. Silk-Based Gene Carriers with Cell Membrane Destabilizing Peptides. *Biomacromolecules* **2010**, 11, 3189-3195.
67. Flenniken, M. L.; Liepold, L. O.; Crowley, B. E.; Willits, D. A.; Young, M. J.; Douglas, T. Selective attachment and release of a chemotherapeutic agent from the interior of a protein cage architecture. *Chemical Communications* **2005**, 447-449.
68. Uchida, M.; Flenniken, M. L.; Allen, M.; Willits, D. A.; Crowley, B. E.; Brumfield, S.; Willis, A. F.; Jackiw, L.; Jutila, M.; Young, M. J.; Douglas, T. Targeting of cancer cells with ferrimagnetic ferritin cage nanoparticles. *Journal of the American Chemical Society* **2006**, 128, 16626-16633.
69. Ren, Y.; Wong, S. M.; Lim, L. Y. Folic acid-conjugated protein cages of a plant virus: A novel delivery platform for doxorubicin. *Bioconjugate Chemistry* **2007**, 18, 836-843.
70. Ng, B. C.; Yu, M.; Gopal, A.; Rome, L. H.; Monbouquette, H. G.; Tolbert, S. H. Encapsulation of Semiconducting Polymers in Vault Protein Cages. *Nano Letters* **2008**, 8, 3503-3509.
71. Soh, N. Selective chemical labeling of proteins with small fluorescent molecules based on metal-chelation methodology. *Sensors* **2008**, 8, 1004-1024.

72. Gillitzer, E.; Willits, D.; Young, M.; Douglas, T. Chemical modification of a viral cage for multivalent presentation. *Chemical Communications* **2002**, 2390-2391.
73. Flenniken, M. L.; Willits, D. A.; Brumfield, S.; Young, M. J.; Douglas, T. The Small Heat Shock Protein Cage from *Methanococcus jannaschii* Is a Versatile Nanoscale Platform for Genetic and Chemical Modification. *Nano Letters* **2003**, 3, 1573-1576.
74. Beames, B.; Lanford, R. E. Insertions within the Hepatitis-B Virus Capsid Protein Influence Capsid Formation and Rna Encapsidation. *Journal of Virology* **1995**, 69, 6833-6838.
75. Braun, H.; Boller, K.; Lower, J.; Bertling, W. M.; Zimmer, A. Oligonucleotide and plasmid DNA packaging into polyoma VP I virus-like particles expressed in *Escherichia coli* (vol 29, pg 31, 1999). *Biotechnology and Applied Biochemistry* **1999**, 29, 269-269.
76. Taylor, K. M.; Lin, T. W.; Porta, C.; Mosser, A. G.; Giesing, H. A.; Lomonosoff, G. P.; Johnson, J. E. Influence of three-dimensional structure on the immunogenicity of a peptide expressed on the surface of a plant virus. *Journal of Molecular Recognition* **2000**, 13, 71-82.
77. DS, P. A Viral Platform for Chemical Modification and Multivalent Display. *Journal of Nanobiotechnology* **2003**, 1, 5.
78. Flenniken, M. L.; Willits, D. A.; Harmsen, A. L.; Liepold, L. O.; Harmsen, A. G.; Young, M. J.; Douglas, T. Melanoma and lymphocyte cell-specific targeting incorporated into a heat shock protein cage architecture. *Chemistry & Biology* **2006**, 13, 161-170.

79. Meldrum, F. C.; Wade, V. J.; Nimmo, D. L.; Heywood, B. R.; Mann, S. Synthesis of Inorganic Nanophase Materials in Supramolecular Protein Cages. *Nature* **1991**, 349, 684-687.
80. Ma-Ham, A.; Wu, H.; Wang, J.; Kang, X.; Zhang, Y.; Lin, Y. Apoferritin-based nanomedicine platform for drug delivery: equilibrium binding study of daunomycin with DNA. *Journal of Materials Chemistry* **2011**, 21, 8700-8708.
81. Papetti, M.; Herman, I. M. Mechanisms of normal and tumor-derived angiogenesis. *American Journal of Physiology-Cell Physiology* **2002**, 282, C947-C970.
82. Reichen, J. The role of the sinusoidal endothelium in liver function. *News in Physiological Sciences* **1999**, 14, 117-121.
83. Kickhoefer, V. A.; Han, M.; Raval-Fernandes, S.; Poderycki, M. J.; Moniz, R. J.; Vaccari, D.; Silvestry, M.; Stewart, P. L.; Kelly, K. A.; Rome, L. H. Targeting Vault Nanoparticles to Specific Cell Surface Receptors. *Acs Nano* **2009**, 3, 27-36.
84. Brown, K. C. Hitting the target with bifunctional phage. *Chemistry & Biology* **2004**, 11, 1033-1035.
85. Ruoslahti, E. Targeting tumor vasculature with homing peptides from phage display. *Seminars in Cancer Biology* **2000**, 10, 435-442.
86. Ruoslahti, E. Specialization of tumour vasculature. *Nat Rev Cancer* **2002**, 2, 83-90.
87. Raja, K. S.; Wang, Q.; Gonzalez, M. J.; Manchester, M.; Johnson, J. E.; Finn, M. G. Hybrid virus-polymer materials. 1. Synthesis and properties of PEG-decorated cowpea mosaic virus. *Biomacromolecules* **2003**, 4, 472-476.

88. Wang, Q.; Lin, T.; Tang, L.; Johnson, J. E.; Finn, M. G. Icosahedral Virus Particles as Addressable Nanoscale Building Blocks. *Angewandte Chemie International Edition* **2002**, 41, 459-462.
89. Khor, I. W.; Lin, T. W.; Langedijk, J. P. M.; Johnson, J. E.; Manchester, M. Novel strategy for inhibiting viral entry by use of a cellular receptor-plant virus chimera. *Journal of Virology* **2002**, 76, 4412-4419.
90. Chasteen, N. D.; Harrison, P. M. Mineralization in ferritin: An efficient means of iron storage. *Journal of Structural Biology* **1999**, 126, 182-194.
91. Treffry, A.; Zhao, Z. W.; Quail, M. A.; Guest, J. R.; Harrison, P. M. How the presence of three iron binding sites affects the iron storage function of the ferritin (EcFtnA) of Escherichia coli. *Febs Letters* **1998**, 432, 213-218.
92. Allen, M.; Willits, D.; Mosolf, J.; Young, M.; Douglas, T. Protein Cage Constrained Synthesis of Ferrimagnetic Iron Oxide Nanoparticles. *Advanced Materials* **2002**, 14, 1562-1565.
93. Wiedenheft, B.; Mosolf, J.; Willits, D.; Yeager, M.; Dryden, K. A.; Young, M.; Douglas, T. An archaeal antioxidant: Characterization of a Dps-like protein from Sulfolobus solfataricus. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, 102, 10551-10556.
94. Allen, M.; Willits, D.; Young, M.; Douglas, T. Constrained Synthesis of Cobalt Oxide Nanomaterials in the 12-Subunit Protein Cage from Listeria innocua. *Inorganic Chemistry* **2003**, 42, 6300-6305.
95. Douglas, T.; Stark, V. T. Nanophase cobalt oxyhydroxide mineral synthesized within the protein cage of ferritin. *Inorganic Chemistry* **2000**, 39, 1828-1830.

96. Douglas, T.; Strable, E.; Willits, D.; Aitouchen, A.; Libera, M.; Young, M. Protein engineering of a viral cage for constrained nanomaterials synthesis. *Advanced Materials* **2002**, 14, 415-+.
97. Klem, M. T.; Willits, D.; Solis, D. J.; Belcher, A. M.; Young, M.; Douglas, T. Bio-inspired synthesis of protein-encapsulated CoPt nanoparticles. *Advanced Functional Materials* **2005**, 15, 1489-1494.
98. Klem, M. T.; Mosolf, J.; Young, M.; Douglas, T. Photochemical mineralization of europium, titanium, and iron oxyhydroxide nanoparticles in the ferritin protein cage. *Inorganic Chemistry* **2008**, 47, 2237-2239.
99. Varpness, Z.; Peters, J. W.; Young, M.; Douglas, T. Biomimetic synthesis of a H-2 catalyst using a protein cage architecture. *Nano Letters* **2005**, 5, 2306-2309.
100. Brumfield, S.; Willits, D.; Tang, L.; Johnson, J. E.; Douglas, T.; Young, M. Heterologous expression of the modified coat protein of Cowpea chlorotic mottle bromovirus results in the assembly of protein cages with altered architectures and function. *Journal of General Virology* **2004**, 85, 1049-1053.
101. Blum, A. S.; Soto, C. M.; Wilson, C. D.; Cole, J. D.; Kim, M.; Gnade, B.; Chatterji, A.; Ochoa, W. F.; Lin, T. W.; Johnson, J. E.; Ratna, B. R. Cowpea mosaic virus as a scaffold for 3-D patterning of gold nanoparticles. *Nano Letters* **2004**, 4, 867-870.
102. Anderson, E. A.; Isaacman, S.; Peabody, D. S.; Wang, E. Y.; Canary, J. W.; Kirshenbaum, K. Viral nanoparticles donning a paramagnetic coat: Conjugation of MRI contrast agents to the MS2 capsid. *Nano Letters* **2006**, 6, 1160-1164.
103. Klem, M. T.; Willits, D.; Young, M.; Douglas, T. 2-D array formation of genetically engineered viral cages on Au surfaces and imaging by atomic force microscopy. *Journal of the American Chemical Society* **2003**, 125, 10806-10807.

104. Martinez-Morales, A. A.; Portney, N. G.; Zhang, Y.; Destito, G.; Budak, G.; Ozbay, E.; Manchester, M.; Ozkan, C. S.; Ozkan, M. Synthesis and Characterization of Iron Oxide Derivatized Mutant Cowpea Mosaic Virus Hybrid Nanoparticles. *Advanced Materials* **2008**, 20, 4816-+.
105. Reed, L. J. Multienzyme complexes. *Accounts of Chemical Research* **1974**, 7, 40-46.
106. Perham, R. N. Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complex - a paradigm in the design of a multifunctional protein. *Biochemistry* **1991**, 30, 8501-8512.
107. Lessard, I. A. D.; Domingo, G. J.; Borges, A.; Perham, R. N. Expression of genes encoding the E2 and E3 components of the *Bacillus stearothermophilus* pyruvate dehydrogenase complex and the stoichiometry of subunit interaction in assembly in vitro. *European Journal of Biochemistry* **1998**, 258, 491-501.
108. Izard, T.; Aevarsson, A.; Allen, M. D.; Westphal, A. H.; Perham, R. N.; de Kok, A.; Hol, W. G. J. Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, 96, 1240-1245.
109. Milne, J. L. S.; Shi, D.; Rosenthal, P. B.; Sunshine, J. S.; Domingo, G. J.; Wu, X. W.; Brooks, B. R.; Perham, R. N.; Henderson, R.; Subramaniam, S. Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. *Embo Journal* **2002**, 21, 5587-5598.
110. Oliver, R. M.; Reed, L. J., *Electron Microscopy of Proteins*. Harris, J. R.: Academic, London, 1982; Vol. 2.

111. Milne, J. L. S.; Wu, X. W.; Borgnia, M. J.; Lengyel, J. S.; Brooks, B. R.; Shi, D.; Perham, R. N.; Subramaniam, S. Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy. *Journal of Biological Chemistry* **2006**, 281, 4364-4370.
112. Henderson, C. E.; Perham, R. N.; Finch, J. T. Structure and symmetry of *B. stearothermophilus* pyruvate dehydrogenase multienzyme complex and implications for eucaryote evolution. *Cell* **1979**, 17, 85-93.
113. Domingo, G. J.; Chauhan, H. J.; Lessard, I. A. D.; Fuller, C.; Perham, R. N. Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus*. *European Journal of Biochemistry* **1999**, 266, 1136-1146.
114. Domingo, G. J.; Orru, S.; Perham, R. N. Multiple display of peptides and proteins on a macromolecular scaffold derived from a multienzyme complex. *Journal of Molecular Biology* **2001**, 305, 259-267.
115. Dalmau, M.; Lim, S.; Chen, H. C.; Ruiz, C.; Wang, S. W. Thermostability and Molecular Encapsulation Within an Engineered Caged Protein Scaffold. *Biotechnology and Bioengineering* **2008**, 101, 654-664.
116. Domingo, G. J.; Caivano, A.; Sartorius, R.; Barba, P.; Backstrom, M.; Piatier-Tonneau, D.; Guardiola, J.; De Berardinis, P.; Perham, R. N. Induction of specific T-helper and cytolytic responses to epitopes displayed on a virus-like protein scaffold derived from the pyruvate dehydrogenase multienzyme complex. *Vaccine* **2003**, 21, 1502-1509.
117. Jung, H. I.; Cooper, A.; Perham, R. N. Identification of key amino acid residues in the assembly of enzymes into the pyruvate dehydrogenase complex of

- Bacillus stearothermophilus: A kinetic and thermodynamic analysis. *Biochemistry* **2002**, 41, 10446-10453.
118. Ren, D.; Kratz, F.; Wang, S.-W. Protein Nanocapsules Containing Doxorubicin as a pH-Responsive Delivery System. *Small* **2011**, 7, 1051-1060.
 119. Molino, N. M.; Bilotkach, K.; Fraser, D. A.; Ren, D. M.; Wang, S. W. Complement Activation and Cell Uptake Responses Toward Polymer-Functionalized Protein Nanocapsules. *Biomacromolecules* **2012**, 13, 974-981.
 120. Ren, D.; Dalmau, M.; Randall, A.; Shindel, M. M.; Baldi, P.; Wang, S.-W. Biomimetic Design of Protein Nanomaterials for Hydrophobic Molecular Transport. *Advanced Functional Materials* **2012**, n/a-n/a.
 121. Cell Injury and Cell Death. In *Cells, Tissues, and Disease: Principles of General Pathology*, 2nd ed.; Guido Majno; Joris, I., Eds. Oxford University Press: New York, 2004; pp 186-245.
 122. Bohm, G.; Muhr, R.; Jaenicke, R. Quantitative-Analysis of Protein Far Uv Circular-Dichroism Spectra by Neural Networks. *Protein Engineering* **1992**, 5, 191-195.
 123. Krohn, N. M.; Yanagisawa, S.; Grasser, K. D. Specificity of the stimulatory interaction between chromosomal HMGB proteins and the transcription factor Dof2 and its negative regulation by protein kinase CK2-mediated phosphorylation. *Journal of Biological Chemistry* **2002**, 277, 32438-32444.
 124. Berendsen, H. J. C.; Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, 26, 1701-1718.

125. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, 4, 435-447.
126. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001**, 7, 306-317.
127. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **2004**, 25, 1656-1676.
128. Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *ELECTROPHORESIS* **1997**, 18, 2714-2723.
129. Berendsen, H. J. C.; Feenstra, K. A.; Hess, B. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **1999**, 20, 786-798.
130. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **1984**, 81, 3684-3690.
131. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, 103, 8577-8593.
132. Park, J. S.; Han, T. H.; Lee, K. Y.; Han, S. S.; Hwang, J. J.; Moon, D. H.; Kim, S. Y.; Cho, Y. W. N-acetyl histidine-conjugated glycol chitosan self-assembled nanoparticles for intracytoplasmic delivery of drugs: Endocytosis, exocytosis and drug release. *Journal of Controlled Release* **2006**, 115, 37-45.

133. Putnam, D.; Gentry, C. A.; Pack, D. W.; Langer, R. Polymer-based gene delivery with low cytotoxicity by a unique balance of side-chain termini. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, 98, 1200-1205.
134. Stryer, L., In *Biochemistry*, 3rd ed.; Freeman and Co: New York, 1988.
135. DeLano, W., The PyMol molecular graphic system. In DeLano Scientific: Palo Alto, 2002.
136. Benjwal, S.; Verma, S.; Rohm, K. H.; Gursky, O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Science* **2006**, 15, 635-639.
137. Salvatella, X.; Calloni, G.; Lendel, C.; Campioni, S.; Giannini, S.; Gliozzi, A.; Relini, A.; Vendruscolo, M.; Dobson, C. M.; Chiti, F. Structure and dynamics of a partially folded protein are decoupled from its mechanism of aggregation. *Journal of the American Chemical Society* **2008**, 130, 13040-13050.
138. Hill, C. P.; Worthylake, D.; Bancroft, D. P.; Christensen, A. M.; Sundquist, W. I. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, 93, 3099-3104.
139. Iwasaki, K.; Miyazaki, N.; Hammar, L.; Zhu, Y.; Omura, T.; Wu, B.; Sjoborg, F.; Yonekura, K.; Murata, K.; Namba, K.; Caspar, D. L.; Fujiyoshi, Y.; Cheng, R. H. Pleomorphic Configuration of the Trimeric Capsid Proteins of Rice dwarf virus that Allows Formation of Both the Outer Capsid and Tubular Crystals. *Journal of Molecular Biology* **2008**, 383, 252-265.

140. Delmas, B.; Laude, H. Assembly of Coronavirus Spike Protein into Trimers and Its Role in Epitope Expression. *Journal of Virology* **1990**, 64, 5367-5375.
141. Belyi, V. A.; Muthukumar, M. Electrostatic Origin of the Genome Packing in Viruses. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, 103, 17174-17178.
142. Calhoun, S. L.; Speir, J. A.; Rao, A. L. N. In vivo particle polymorphism results from deletion of a N-terminal peptide molecular switch in brome mosaic virus capsid protein. *Virology* **2007**, 364, 407-421.
143. Mattevi, A.; Obmolova, G.; Kalk, K. H.; Westphal, A. H.; de Kok, A.; Hol, W. G. J. Refined Crystal Structure of the Catalytic Domain of Dihydrolipoyl Transacetylase (E2p) from *Azotobacter vinelandii* at 2.6 Å Resolution. *Journal of Molecular Biology* **1993**, 230, 1183-1199.
144. Mattevi, A.; Obmolova, G.; Schulze, E.; Kalk, K. H.; Westphal, A. H.; Dekok, A.; Hol, W. G. J. Atomic-Structure of the Cubic Core of the Pyruvate-Dehydrogenase Multienzyme Complex. *Science* **1992**, 255, 1544-1550.
145. Reed, L. J.; Hackert, M. L. Structure-Function-Relationships in Dihydrolipoamide Acyltransferases. *Journal of Biological Chemistry* **1990**, 265, 8971-8974.
146. Behal, R. H.; Debuysere, M. S.; Demeler, B.; Hansen, J. C.; Olson, M. S. Pyruvate-Dehydrogenase Multienzyme Complex - Characterization of Assembly Intermediates by Sedimentation-Velocity Analysis. *Journal of Biological Chemistry* **1994**, 269, 31372-31377.
147. Wynn, R. M.; Davie, J. R.; Zhi, W.; Cox, R. P.; Chuang, D. T. In-Vitro Reconstitution of the 24-Meric E2 Inner-Core of Bovine Mitochondrial

- Branched-Chain Alpha-Keto Acid Dehydrogenase Complex - Requirement for Chaperonins Groel and Groes. *Biochemistry* **1994**, 33, 8962-8968.
148. Allen, M. D.; Perham, R. N. The catalytic domain of dihydrolipoyl acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus* - Expression, purification and reversible denaturation. *Febs Letters* **1997**, 413, 339-343.
 149. Montgomery, M. G.; Coker, A. R.; Taylor, I. A.; Wood, S. P. Assembly of a 20-nm Protein Cage by *Escherichia coli* 2-Hydroxypentadienoic Acid Hydratase. *Journal of Molecular Biology* **2010**, 396, 1379-1391.
 150. Sana, B.; Johnson, E.; Sheah, K.; Poh, C. L.; Lim, S. Iron-based ferritin nanocore as a contrast agent. *Biointerphases* **2010**, 5, Fa48-52.
 151. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **2006**, 1, 2876-2890.
 152. Berova, N.; Nakanishi, K.; Woody, R. W., *Circular Dichroism: Principles and Applications* 2nd ed.; Wiley-VCH: New York, 2000.
 153. Creighton, T. E., In *Proteins: Structures and Molecular Properties*, 2nd ed.; Freeman and Co.: New York, 1993.
 154. Bancroft, J. B.; Hills, G. J.; Markham, R. A study of the self-assembly process in a small spherical virus formation of organized structures from protein subunits in vitro. *Virology* **1967**, 31, 354-379.
 155. Johnson, J. E.; Speir, J. A. Quasi-equivalent viruses: A paradigm for protein assemblies. *Journal of Molecular Biology* **1997**, 269, 665-675.
 156. Ausar, S. F.; Foubert, T. R.; Hudson, M. H.; Vedvick, T. S.; Middaugh, C. R. Conformational stability and disassembly of Norwalk virus-like particles -

- Effect of pH and temperature. *Journal of Biological Chemistry* **2006**, 281, 19478-19488.
157. Peng, T.; Lee, H.; Lim, S. Isolating a Trimer Intermediate in the Self-Assembly of E2 Protein Cage. *Biomacromolecules* **2012**, 13, 699-705.
 158. Subbarao, N. K.; Parente, R. A.; Szoka, F. C.; Nadasdi, L.; Pongracz, K. pH-Dependent Bilayer Destabilization by an Amphipathic Peptide. *Biochemistry* **1987**, 26, 2964-2972.
 159. Li, W. J.; Nicol, F.; Szoka, F. C. GALA: a designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery. *Advanced Drug Delivery Reviews* **2004**, 56, 967-985.
 160. Douglas, T.; Young, M. Virus particles as templates for materials synthesis. *Advanced Materials* **1999**, 11, 679-681.
 161. Klem, M. T.; Young, M.; Douglas, T. Biomimetic synthesis of beta-TiO(2) inside a viral capsid. *Journal of Materials Chemistry* **2008**, 18, 3821-3823.
 162. de la Escosura, A.; Verwegen, M.; Sikkema, F. D.; Comellas-Aragones, M.; Kirilyuk, A.; Rasing, T.; Nolte, R. J. M.; Cornelissen, J. J. L. M. Viral capsids as templates for the production of monodisperse Prussian blue nanoparticles. *Chemical Communications* **2008**, 1542-1544.
 163. Ha, Y.; Shi, D. S.; Small, G. W.; Theil, E. C.; Allewell, N. M. Crystal structure of bullfrog M ferritin at 2.8 angstrom resolution: analysis of subunit interactions and the binuclear metal center. *Journal of Biological Inorganic Chemistry* **1999**, 4, 243-256.
 164. Tosha, T.; Hasan, M. R.; Theil, E. C. The ferritin Fe₂ site at the diiron catalytic center controls the reaction with O(2) in the rapid mineralization pathway.

Proceedings of the National Academy of Sciences of the United States of America **2008**, 105, 18182-18187.

165. Masuda, T.; Goto, F.; Yoshihara, T.; Mikami, B. Crystal Structure of Plant Ferritin Reveals a Novel Metal Binding Site That Functions as a Transit Site for Metal Transfer in Ferritin. *Journal of Biological Chemistry* **2010**, 285, 4049-4059.
166. Douglas, T.; Dickson, D. P. E.; Betteridge, S.; Charnock, J.; Garner, C. D.; Mann, S. Synthesis and Structure of an Iron(III) Sulfide-Ferritin Bioinorganic Nanocomposite. *Science* **1995**, 269, 54-57.

Appendices

A.1 Cross-linking of E2 Protein

Glutaraldehyde, which has a spacer arm of 5 Å,¹²³ is known as a common agent used in protein cross-linking reactions (Figure A.1.1A). It can specifically react with primary amine groups, such as ε-amino group of lysine, and form stable covalent bonds.

Figure A.1.1 Crystallographic structure of trimers and highlighted lysine residues. (A) The formula of glutaraldehyde, with the spacer arm of 5 Å. (B) Surface representation of possible trimer structure is shown in grey. All lysine residues on different subunit are highlighted in red, magentas, and blue, respectively. The possible waving arms are represented in yellow. (C) Two trimer structures showing the positions of lysines. The color setting is the same as (B).

Figure A.1.1B shows 21 lysine residues on each subunit that were widely distributed on the polypeptide. Some of the residues were exposed to the exterior and interior surfaces which facilitated the cross-linking by glutaraldehyde. Previous work suggested that the N-terminal arm was not essential to maintain the correct assembled structure of E2 protein.^{32, 33} The crystallographic structure also indicated that the N-terminus are anchored to neighboring subunit and do not form any secondary structures. The clues suggested that the N-terminal arms may be waving on the surface of trimer structures. Subsequently, the distances between the lysines on these unfixed N-terminus and the lysines on neighboring subunits may be within the spacer arm of glutaraldehyde, thus facilitate the cross-linking. As a result, the trimer structures were stabilized by both inter-subunit interactions and numbers of cross-linkers. On the other hand, there were few numbers of lysines that were available to form inter-trimer cross-linking (Figure A.1.1C). Hence, the trimers were coupled by both inter-trimer interactions and few cross-linkers.

Although E2-4H can present as correctly assembled structure at physiological pH, the inter-trimer interactions were weakened by incorporation of histidines. During SDS-PAGE, cross-linkers were no longer able to maintain the association of trimers while some of the inter-trimer interactions were lost. However, three subunits from the trimer structure were still cross-linked. Thus, we observed possible trimer bands on the SDS-PAGE. In contrast, the structures of E2-WT and E2-(2+2)H were tightly held together by inter-subunit interactions and cross-linkers. Both of the proteins presented agglomerate on the SDS-PAGE.

A.2 Denaturant Effects on E2 Protein Self-Assembly

To obtain dissociated E2 subunits for investigation of self-assembly mechanism, denaturant treatment on E2 protein was adopted. We use two types of denaturants on our experiments: urea and guanidine hydrochloride (GuHCl).

The dissociation status of E2 protein would be reflected by the change of its hydrodynamic diameters. About 400 µg wild-type E2 protein (E2-WT) was incubated for 1 h with one kind of denaturant dissolved in sodium phosphate buffer with the final measurement volume of 1 ml, and final pH of 7.4. For each sample measurement, the hydrodynamic diameters of the E2 assemblies were determined by dynamic light scattering (DLS, Zetasizer Nano ZS, Malvern) upon denaturant treatment.

Figure A.2.1A indicated that GuHCl had no effect on E2 protein size when its concentration was lower than 2.5 M. When the concentration of GuHCl was increased gradually from 2.5 M to 4.5 M, the protein size decreased abruptly and reached around 13 nm which might be the size of either individual subunit or stable intermediate composed of multiple subunits. Compared to GuHCl denaturation, E2 protein presented distinct response to urea (Figure A.2.1B). The hydrodynamic diameters of E2 protein increased slowly until the concentration of urea was 6.5 M.

Figure A.2.1 The effect of denaturants on hydrodynamic diameters of E2-WT. (A) GuHCl; (B) urea.

To further characterize the disassociation of E2-WT with GuHCl, size exclusion chromatography (SEC) was performed (Figure A.2.2). In 1 M GuHCl, E2-WT presented the same elution profile as E2-WT with no GuHCl. When the GuHCl concentration was higher than 2 M, there was a small peak indicating smaller size of protein on E2-WT profile. However, the elution volume of that smaller peak was different at different GuHCl concentration (2, 3, 4 M), indicating different sizes of these smaller proteins. The results were comparable to the DLS data, showing different protein sizes at different GuHCl concentration. Majority of the E2 proteins still eluted out at the same volume (Void volume of the SEC column) at all GuHCl concentrations. Since the separation limit of the SEC column was 600 kDa, which is much smaller than the size of 60-mer E2 protein (1687 kDa), we could not tell the precise disassociation states under each GuHCl concentration. However, the results still told us that the complete dissociation of E2 protein could not be achieved using denaturant treatments.

Figure A.2.2 SEC profiles of E2-WT at different concentrations of GuHCl.

A.3 Iron Mineralization Supporting Characterizations

To determine the wavelength at which iron mineral has maximum absorbance, UV-vis scan was performed on E2-LFer and E2-LE6. After iron-loading, proteins present a broad absorption band centered at around 350 nm other than the 280 nm peak due to the formation of the iron core (Figure A.3.1). Therefore, 350 nm was set on SEC program to detect the iron composition within the protein cage.

Figure A.3.1 UV-vis absorbance scans of E2-LFer and E2-LE6 solutions. Other than 280 nm peak for protein, iron-treated E2-LFer and E2-LFer show a broad absorption band centered at around 350 nm.

The accumulation of iron minerals in iron-binding E2 protein is affected by the 12 pores on the surface (Figure A.3.2).

Figure A.3.2 PyMol representing the locations of RDGE loops around the pore on E2 protein. RDGE loops are highlighted in red. The distance is linked with yellow lines and highlighted in green numbers. The unit is angstrom.

A.4 Site-Directed Mutagenesis Protocol

Procedures:

1. Identify site for mutation.
2. Design appropriate primer pair.
3. PCR to amplify the mutant gene using Pfu ultra-high-fidelity DNA polymerase (Fermentas).
4. PCR clean-up using PCR purification kit (Qiagen).
5. Digest PCR product with Dpn I (Fermentas) to remove methylated template DNA.
6. Transformation of PCR product into host *E.coli* strain DH5 α .
7. Miniprep the mutant plasmid from Agar plates for then construct.
8. Sequence mutant sequence (1st base, Singapre).

PCR Reaction

1. PCR mix (50 μ l):

DNA template (pE2-WT)	1 μ l (~50 ng)
Primer mix (from 1 st base)	2 μ l (~150 ng)
dNTP mix (Fermentas)	1 μ l
10X reaction buffer (Fermentas)	5 μ l
DI water	40 μ l
Pfu DNA polymerase	1 μ l

2. PCR thermo-cycling:

95 °C	1 min	
95 °C	30 sec	} 18 cycles
55 °C	1 min	
72 °C	8 min	
72 °C	5 min	
4 °C	Hold	

A.5 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Reducing sample loading buffer:

Mix the Laemmli sample buffer (Bio-rad) with 2-merkapto ethanol (Sigma-Aldrich) at 19:1 ratio.

Sample preparation:

Mix the protein sample with reducing buffer at 1:1 ration, then heat at 90 °C for ~10 min.

Sample Running:

10 ul of each prepared sample is loaded onto each well on the gel. The gel is run at 100 V for 10 min, then 175 V for 25 min. Then gel is stained with Bio-sage Coomassie stain to get visible protein bands.

Gel recipes: (Prepared in fume hood)

Resolving Gel (10%): (5 ml preparation)

Distilled water	2.45 ml
40% Acrylamide-bis solution	1.25 ml
1.5 M Tris (pH 8.8)	1.25 ml
10% SDS solution	50 µl
10% (w/v) APS (ammonium persulfate)	25 µl
TEMED (tetramethylethylenediamine)	5 µl

Stacking Gel (4%): (5 ml preparation)

Distilled water	3.2 ml
40% Acrylamide-bis solution	0.4 ml
0.5 M Tris (pH 6.8)	1.25 ml
10% SDS solution	50 µl
10% (w/v) APS (ammonium persulfate)	25 µl
TEMED (tetramethylethylenediamine)	5 µl

Note: Add APS & TEMED just before casting the gel

A.6 Negative staining of TEM samples:

1. Make the protein solution of 0.1 – 0.2 mg/ml
2. Prepare a dilute solution (1 - 5%) of uranyl acetate in water
3. Place a drop (~50 μ l) of protein solution on parafilm and float the carbon coated copper grid (for TEM) on the drop (dark side facing the drop)
4. After 3 min, take out the grid from the sample
5. Remove excess sample by soaking with blotting paper and air-dry for about 5 min
6. Place a drop (~100 μ l) of uranyl acetate solution on the parafilm and float the TEM grid on the drop (sample side facing the drop).
7. After 3 min, take the grid from the drop
8. Remove excess uranyl acetate by soaking with blotting paper and air-dry for about 30 min
9. Put the grid in desiccator until observation in TEM
10. Keep excess uranyl acetate solution for reuse
11. Dispose the blotting paper, parafilm and contaminated pipette tips in designated container.