Investigation of structure and particle fabrication of ferritin

Fan, Rongli

2011

Fan, R. L. (2011). Investigation of structure and particle fabrication of ferritin. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/42845>

<https://doi.org/10.32657/10356/42845>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

INVESTIGATION INTO STRUCTURE AND
PARTICLE FABRICATION OF FERRITIN

FAN RONGLI

2011

**INVESTIGATION OF STRUCTURE AND PARTICLE
FABRICATION OF FERRITIN**

FAN RONGLI

**SCHOOL OF PHYSICAL AND MATHEMATICAL
SCIENCES**

2011

INVESTIGATION OF STRUCTURE AND PARTICLE FABRICATION OF FERRITIN

FAN RONGLI

School of Physical and Mathematical Sciences

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

2011

ACKNOWLEDGEMENT

First and foremost, I would like to convey my appreciation to my supervisor, Prof. Orner for his invaluable advice and guidance. As his first Ph.D student, I learned a lot from him, especially realizing how to carry out my research projects independently through 4 years' strict training. Certainly, his positive academic spirit and life attitude have left a deep impression on me, not only in academic career but also in life.

I am grateful to Vee Vee Cheong, Shu Wen Chew, See Liang Ng for their magnificent work on my projects and the enjoyable time we have had together. I thank Profs. Julien Lescar and Anh Tuan Phan for use of their fast protein liquid chromatography systems, Ming Wei Chen for his work on protein crystallization, and the NUS School of Medicine for the use of their TEM. I would like to thank Jun Guo from NTU's School of Materials Science and Engineering for the helpful discussions regarding the diffraction data. I appreciate Prof. Hongyu Chen and Prof. Yuan Chen for DLS advice and use of their ultracentrifuge respectively. I also thank Dr Gang Chen and Li Wei for sucrose gradient centrifugation advice. I thank Ms. Ai Hua Seow for use of the CBC teaching lab equipments.

Special thanks are extended to all my lab members, Limin Yu, Yueqin Fang, Yu Zhang, Maziar Soleymani Ardejani, Tom Andrew Cornell, etc. for their helpful discussion regarding my research. My appreciation also goes to Prof. Bengang Xing and Tianhu Li for their help during my PhD study.

Last but not least, thanks to my parents and husband for their continual support, encourage and understanding these years.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	i
TABLE OF CONTENTS	ii
List of Figures	v
List of Tables	xi
ABSTRACT	xii
CHAPTER 1	
Introduction	1
1.1 Protein structure	1
1.2 Protein assembly	2
1.3 Octahedral ferritins	5
1.3.1 Tertiary and quaternary structure of octahedral ferritins	5
1.3.2 Structural comparison between mammalian ferritins and bacterioferritins	6
1.3.3 The mechanism of ferritin self-assembly	8
1.3.4 Mineralization of ferritin	9
1.4 Tetrahedral ferritin-like proteins	12
1.4.1 Structure of Dps proteins	12
1.4.2 DNA binding by Dps	13
1.4.3 The ferroxidase center	14
1.5 Icosahedral viruses	15
1.5.1 Quasi-equivalence	16
1.5.2 Cowpea chlorotic mottle virus	17
1.6 Protein cages applied to the templated mineralization of nanomaterials	18
1.6.1 Ferritins as templates to mineralize nanoparticles	18
1.6.1.1 Metal Oxides, hydroxides and oxyhydroxides	19
1.6.1.2 Metal carbonates	20
1.6.1.3 Semiconductors	21
1.6.1.4 Metals	22
1.6.1.5 General principles for the templation of nanoparticles inside ferritin	24
1.6.2 Dps proteins as templates to mineralize nanoparticles	25
1.6.3 Viruses as templates to mineralize nanoparticles	26
1.6.4 Summary	26
1.7 Research perspectives	27
1.8 References	27
CHAPTER 2	
Investigation into the structural energetic consequences of helix swapping between a ferritin and a ferritin-like protein ^[1]	35
2.1 Introduction	35
2.1.1 The ferritins – model protein cages	36
2.1.1.1 The structure of bacterioferritin	36
2.1.1.2 The structure of DNA-binding protein from starved cells	37

2.1.2	Structural comparison between BFR and DPS	38
2.1.3	Previous studies to understand ferritin self-assembly	43
2.1.4	Aims and objectives	44
2.2	Experimental	46
2.2.1	Total gene synthesis of BFR and DPS	46
2.2.1.1	Codons optimization and oligonucleotides (oligos) design	46
2.2.1.2	Assembly, amplification and extension PCR (Figure 2. 7)	47
2.2.2	Preparation of extended single stranded complementary ends and annealing to pET-32 Ek/LIC vector (Figure 2. 8).....	48
2.2.3	Colony screening by PCR and sequencing	49
2.2.4	Protein expression and extraction	50
2.2.5	Protein purification under native conditions (Figure 2. 9).....	51
2.2.6	Protein purifications under denaturing conditions (Figure 2. 9). 51	
2.2.7	Temperature dependent circular dichroism (CD) analysis	52
2.2.8	Dynamic light scattering (DLS).....	53
2.2.9	Size exclusion chromatography (SEC)	53
2.2.10	Transmission electron microscopy (TEM)	53
2.2.11	Analytical ultracentrifuge sedimentation equilibrium (SE).....	54
2.3	Results and discussion	54
2.3.1	Protein design.....	55
2.3.2	Verification of PCR products.....	56
2.3.3	Protein expression and purification	62
2.3.4	Circular dichroism-based studies.....	65
2.3.5	Transmission electron microscopy investigation.....	69
2.3.6	Dynamic light scattering and size exclusion chromatography ...	72
2.3.7	Sedimentation equilibrium calculation	76
2.4	Conclusion	77
2.5	References.....	80
CHAPTER 3	Fabrication of gold nanoparticles inside unmodified horse spleen apoferritin ^[1]	84
3.1	Introduction.....	84
3.1.1	Amino acids-induced the formation of gold nanomaterials.....	85
3.1.2	Peptide-mediated the synthesis of gold nanoparticles	87
3.1.3	Protein-assisted fabrication of gold nanoparticle.....	89
3.1.4	Aims and objectives	92
3.2	Experimental	93
3.2.1	Protein purification	93
3.2.2	Gold nanoparticles mineralization in horse spleen ferritin	93
3.2.3	Ultraviolet-Visible spectroscopy (UV)	94
3.2.4	Size exclusion chromatography (SEC)	94
3.2.5	Native polyacrylamide gel electrophoresis	95
3.2.6	Transmission electron microscopy (TEM)	95
3.2.7	High-resolution transmission electron microscopy (HRTEM), energy- dispersive X-ray spectroscopy (EDX) and selected area diffraction (SAD) analysis	95
3.3	Results and Discussion	96
3.3.1	Design of a novel two-step method to fabricate gold nanoparticles inside unmodified horse spleen ferritin	96

3.3.2	Characterization of the gold nanoparticles mineralized inside unmodified horse spleen ferritin by UV-Vis absorption.....	97
3.3.3	Characterization of gold nanoparticles inside unmodified horse spleen ferritin by SEC	98
3.3.4	Characterization of gold nanoparticles inside unmodified horse spleen ferritin by native PAGE gel	100
3.3.5	Characterization of gold nanoparticles inside unmodified horse spleen ferritin by TEM.....	100
3.3.6	Characterization of gold nanoparticles inside unmodified horse spleen ferritin by element composition analysis	101
3.3.7	Gold nanoparticles crystalline analysis by TEM and SAED	102
3.3.8	Mechanistic study to determine the necessity of the sucrose gradient ultracentrifugation step	103
3.4	Conclusion	104
3.5	References.....	105
CHAPTER 4	Conclusion	108
APPENDIXES		110
Appendix 1	Details on gene design of BFR and DPS proteins.....	110
Appendix 2	Some interactions in BFR and DPS	114
Appendix 3	Analysis of all inter-subunit interfaces in BFR.....	117
Appendix 4	Analysis of all inter-subunit interfaces in DPS.	124

List of Figures

- Figure 1. 1 Cartoon representation of cyclic, dihedral and cubic symmetry of oligomeric proteins. Symmetry axes are indicated as dotted lines. A circle, triangle, square, or pentamer at the end of an axis represents two-, three-, four- or five-fold symmetry respectively. The figure is taken from Janin *et al.*^[10] 4
- Figure 1. 2 (left) Three-dimensional and (right) two-dimensional views of cyclic dimer (C2), cyclic tetramer (C4), dihedral dimer (D2) and dihedral tetramer (D4). Interactions on the face-to-back interfaces are involved in C4 and D4 complexes. In C2, D2 and D4 complexes, cross interactions are made on the face-to-face and back-to-back interfaces. The figure is taken from Levy *et al.*^[11] 5
- Figure 1. 3 a) Ribbon diagram of a human light chain ferritin subunit. The A helix is colored purple, the B helix cyan, the C helix green, the D helix yellow and the E helix orange. The human light chain ferritin assembled protein cage is viewed along b) a three-fold axis and c) a four-fold axis. (Protein Data Bank ID is 2ffx and all figures are generated by using Chimera.)..... 6
- Figure 1. 4 (left) The heme is located at the two-fold symmetric, dimer interface of *E. coli* bacterioferritins. (right) Expanded view. (The Protein Data Bank ID is 1bfr and all figures are generated by using Chimera.)..... 8
- Figure 1. 5 (left) View along the three-fold axis of human H-chain ferritin (Protein Data Bank ID is 2cei) from the inside of the cage and (right) the expanded view of the iron pathway from the three-fold axis to the ferroxidase center. Residues D131 and E134 (blue labels) point into the three-fold channel and E27, Y34, E61, E62, H65, and Q141 (red labels) are involved in binding to iron atoms at the ferroxidase center. T135, H136, and Y137 (black labels) direct iron atoms to the ferroxidase center. The overall direction of the iron pathway is indicated by the orange arrow. The figure is adapted from Chasteen *et al.*^[30, 42] 10
- Figure 1. 6 a) The ferroxidase center is located entirely within a single subunit of human H-chain ferritin and in the X-ray structure, Zn^{2+} is bound to the ferroxidase center instead of Fe^{2+} . (Protein Data Bank ID is 2cei) b) Schematic depiction of the μ -oxo-diferric complex..... 11
- Figure 1. 7 NMR structure indicating positions of associated iron with increasing iron concentration. The ferroxidase site is indicated in red. As the iron concentration increased with respect to protein, sites less proximal to the ferroxidase site become occupied. At low iron concentration only the light blue sites are occupied, whereas with increasing iron concentration cyan and then dark blue sites are also occupied suggesting a movement toward the E helix. The figure is taken from Theil *et al.*^[48] 12

Figure 1. 8 a) Ribbon diagram of a <i>E. coli</i> Dps subunit. The A helix is colored purple, the B helix cyan, the BC helix green, the C helix yellow and the D helix orange. The Dps protein cage is viewed along b) Dps-like three-fold axis and c) ferritin-like three-fold axis. d) Magnified diagram of ferritin-like pores lined with aspartic acid residue. (Protein Data Bank is 1dps and all figures are generated with Chimera.)	13
Figure 1. 9 (left) The dimeric two-fold symmetry interface of <i>E. coli</i> Dps, (right) illustrating the iron binding residues at the ferroxidase center with two water molecules (orange) occupying at the iron binding sites. (Protein Data Bank ID is 1dps and all figures are generated by using Chimera.)	15
Figure 1. 10 Views of icosahedron down from two-, three-, and five-fold symmetry axes.....	15
Figure 1. 11 Selected icosahedral symmetry capsids with different T numbers. The figure is taken from Caspar <i>et al.</i> ^[66] and Zlotnick <i>et al.</i> ^[68]	16
Figure 1. 12 (left) The facet geometry for icosahedral viruses with quasi-equivalence is indicated by two coordinates (h and k). (right) They are determined by counting the number of vertices from the vertex of one pentamer to the vertex of an adjacent pentamer. See text for further explanation. The figure is taken from Zlotnick. ^[67]	17
Figure 1. 13 TEM images of nickel hydroxide fabrication in different conditions. The ferritin solution was bubbled with a) carbon dioxide (to provide carbonate ions) and b) nitrogen. The figure is taken from Yoshimura <i>et al.</i> ^[84]	20
Figure 1. 14 TEM images of CdSe nanoparticles mineralized in ferritin. The samples were stained with (left) aurothioglucose and (right) without staining. The inset is a high resolution image to demonstrate atomic ordering. The figure is taken from Yamashita <i>et al.</i> ^[88]	22
Figure 1. 15 TEM image of negatively stained ferritin with mineralized Pd nanoclusters. The scale bar is 50 nm. The figure is taken from Watanabe <i>et al.</i> ^[96]	23
Figure 2. 1 (left) Ribbon diagram representation of bacterioferritin from <i>Escherichia coli</i> (Protein Data Bank ID is 1bfr) and (right) the structure of one monomer, which consists of the A (purple), B (cyan), C (yellow), D (orange) and E (red) helices and the BC loop (green).	37
Figure 2. 2 (left) Ribbon diagram representation of DNA-binding protein from starved <i>Escherichia coli</i> (Protein Data Bank ID is 1dps) and (right) its monomer structure, which consists of the A (purple), B (cyan), C (yellow), D (orange) and BC (green) helices.	38

Figure 2. 3 Superimposition of BFR (grey) and DPS (black) monomers. The structural alignment is generated using DaliLite online program which can be found at http://www.ebi.ac.uk/Tools/dalilite/index.html	39
Figure 2. 4 Sequence comparison of the structural alignment of BFR and DPS. The grey shading indicates alpha helix regions and the identical residues are highlighted in red. The structural alignment is generated using DaliLite online program which can be found at http://www.ebi.ac.uk/Tools/dalilite/index.html	40
Figure 2. 5 Comparison of the BFR and DPS protein cages. Views of BFR along a) two-fold axis b) three-fold axis c) four-fold axis. Views of DPS along d) two-fold axis e) ferritin-like three-fold axis f) DPS-like three-fold axis.	42
Figure 2. 6 Top) Crystal structures of the DPS (PDB ID is 1dps) and BFR (PDB ID is 1bfr) protein nanocages with highlighted monomers, associated four-helix bundle monomers that make them up (not to scale) viewed down the helix bundle axis, and schematics indicating the A-D helices associated with the bundle (circles) and the additional BC and E helices (rounded rectangles) (white=DPS; black=BFR). (Bottom) Schematic design for proteins with added or deleted BC and E domains following the convention from above. This figure is taken from Fan <i>et al.</i> ^[1]	45
Figure 2. 7 Schematic of assembly, amplification and extension PCR. This figure is taken from Fan <i>et al.</i> ^[1]	48
Figure 2. 8 Schematic of the cloning strategy. This figure is taken from Fan <i>et al.</i> ^[1]	49
Figure 2. 9 Schematic of protein purification under (left) native and (right) denaturing conditions. This figure is taken from Fan <i>et al.</i> ^[1]	52
Figure 2. 10 Amino acid sequences of proteins used in this study. Grey shading indicates each of the helices (A-D in order from the N- to C-termini) in the four-helix bundle from the parent protein. A black box indicates the E helix from BFR, and a white box indicates the BC helix from DPS. This figure is taken from Fan <i>et al.</i> ^[1]	56
Figure 2. 11 DNA electrophoretic analysis of the products of gene assembly and amplification. (See Figure 2. 7 for a schematic of the gene synthesis strategy.) This figure is taken from Fan <i>et al.</i> ^[1]	58
Figure 2. 12 DNA electrophoretic analysis of the products of extension PCR with LIC primers. The bands of appropriate size (~500 bp) were excised and purified. This figure is taken from Fan <i>et al.</i> ^[1]	59
Figure 2. 13 DNA electrophoretic analysis of the products of single-colony PCR screens for vectors containing the insert genes. This figure is taken from Fan <i>et al.</i> ^[1]	61

Figure 2. 14	Expression and purification of ferritins and ferritin mutants. (from left to right, with the exception of BFR+BC-E) Lane 1: Protein ladder. Lane 2: Uninduced. Lane 3: Induced. Lane 4: Insoluble fraction. Lane 5: Soluble fraction. Lane 6: Flow through after applying soluble fraction to resin. Lane 7: Flow through from washing the resin. Lane 8: Purified protein after Ek cleavage of the affinity tags and release from the resin. Note: The bottom gel for BFR+BC-E was done after on-resin refolding. Lane 1: Ladder. Lane 2: Uninduced. Lane 3: Induced. Lane 4: Solubilized lysate. Lane 5: Flow through after applying solubilized lysate to resin. Lane 6: Flow through from washing the resin. Lane 7: Purified protein after Ek cleavage of the affinity tags and release from the resin. This figure is taken from Fan <i>et al.</i> ^[1]	64
Figure 2. 15	Temperature dependent circular dichroism (CD) spectra of the ferritins and ferritin mutants. CD spectra of the proteins at different temperatures for the BFR (left) and the DPS series (right). Spectra were taken for each protein at temperatures from 20-90 °C at 5 °C intervals. These spectra were used to determine the loss of helical structure with temperature by following the signal at 222nm and fitting the resulting data to establish a melting point. These spectra are the average of at least three replicate melts. This figure is taken from Fan <i>et al.</i> ^[1]	67
Figure 2. 16	(Top) Circular dichroism (CD) spectra of the cage proteins and mutants at 20 °C before thermal denaturation (solid lines) and at 20 °C after thermal denaturation (dotted lines) to indicate level of unfolding reversibility. (Bottom) The CD signal of the cage proteins and mutants at 222 nm during thermal denaturation. This figure is taken from Fan <i>et al.</i> ^[1]	68
Figure 2. 17	Transmission electron micrographs of ferritin and ferritin mutant negatively stained with uranyl acetate. The scale bars are 50 nm. This figure is taken from Fan <i>et al.</i> ^[1]	71
Figure 2. 18	Particle sizes of BFR and DPS mutants from the image analysis of TEM micrographs. This figure is taken from Fan <i>et al.</i> ^[1]	72
Figure 2. 19	Size exclusion chromatography calibration. Top: Chromatogram of calibration proteins. Bottom: Stokes radius (R _{st}) correlation to K _{av} as defined in the Experimental Section. Conalbumin was excluded from the Stokes radii correlation due to the fact that its radius is unknown. This figure is taken from Fan <i>et al.</i> ^[1]	74
Figure 2. 20	Dynamic light scattering (upper left) and size exclusion chromatograms (right) of cage proteins and mutants. The chromatogram peaks were correlated to elution volumes of standard proteins (Figure 2. 19) to determine Stokes diameters (bottom left). The chromatograms were normalized to 1 and the populations of particles of each diameter are indicated. This figure is taken from Fan <i>et al.</i> ^[1]	76

- Figure 2. 21 Sedimentation equilibrium for DPS (bottom left) and DPS+E (bottom right) at four angular velocities. The data are best fit to DPS=12-mer and DPS+E=12-mer and the residuals of these fits are presented above the data traces. This figure is taken from Fan *et al.*^[1] 77
- Figure 3. 1 TEM micrographs of gold nanoparticles prepared with different amino acids (a) lysine, (b) tryptophan, (c) arginine, and (d) tyrosine. The scale bars for the images are (a) 5 nm, (b) 50 nm, (c) 20 nm, and (d) 100 nm respectively. Selected area electron diffraction patterns of gold nanoparticles synthesized by each amino acids are presented in the insets. This figure is taken from Dong *et al.*^[11] 86
- Figure 3. 2 Scheme of synthesizing Au particles on nanotubes. (a) Immobilized histidine-rich peptide on nanotubes. (b) Immobilized Au ions on histidine-rich peptide. (c) Au nanoparticles grown on the nanotubes after reduction. (d) TEM image of Au nanoparticles grown on the nanotubes and the scale bar is 200 nm. This figure is taken from Matsui *et al.*^[17] 88
- Figure 3. 3 TEM characterization of gold nanoparticles densely fabricated on TMV capsids by TEM. (a) Low and (b) high magnification TEM images. This figure is taken from Mann *et al.*^[23] 91
- Figure 3. 4 (a) Scheme of mutated human heavy chain ferritin mineralized gold nanoparticles inside the protein cage. (b) TEM image of gold nanoparticles synthesized within the cavity of ferritin. This figure is taken from Dmochowski *et al.*^[25] 92
- Figure 3. 5 Schematic representation of the two-step reduction procedure using native horse spleen ferritin (HSAFn) to mineralize gold nanoparticles. Gold clusters (HSFn-Au clusters) are mineralized inside HSAFn by reducing AuCl_4^- trapped inside the ferritin cavity. The formation of gold nanoparticles (HSFn-AuNPs) is promoted by addition of ascorbic acid and additional AuCl_4^- . This figure is taken from Fan *et al.*^[1] 97
- Figure 3. 6 UV-Visible spectra of horse spleen apoferritin (HSAFn, black), gold clusters mineralized in horse spleen ferritin (HSFn-Au clusters, blue) and gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs, red). This figure is taken from Fan *et al.*^[1] 98
- Figure 3. 7 Size exclusion chromatography of horse spleen apoferritin (HSAFn, top), gold clusters mineralized in horse spleen ferritin (HSFn-Au clusters, middle) and gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs, bottom). Elution profiles were measured at 280 nm (black line) and 520 nm (red line). This figure is taken from Fan *et al.*^[1] 99
- Figure 3. 8 Transmission electron microscopy images of protein aggregations in the shoulder peaks (at ~10 mL elution volume) HSFn-Au cluster (left) and HSFn-AuNPs (right) after size exclusion chromatography. This figure is taken from Fan *et al.*^[1] 99

Figure 3. 9 Native gel electrophoresis of horse spleen apoferritin (HSAFn, left) and gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs, right). Gel was stained with Coomassie blue. This figure is taken from Fan <i>et al.</i> ^[1]	100
Figure 3. 10 Transmission electron micrographs of a) negatively stained horse spleen apoferritin (HSAFn) b) negatively stained gold clusters mineralized in horse spleen ferritin (HSFn-Au clusters) c) and d) negatively stained gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs). The scale bar is 50 nm for all micrographs. This figure is taken from Fan <i>et al.</i> ^[1]	101
Figure 3. 11 EDX spectrum of unstained gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs). The spectrum was obtained on a copper grid. This figure is taken from Fan <i>et al.</i> ^[1]	102
Figure 3. 12 High resolution transmission electron microscopy (HRTEM) of a) single crystal and b) polycrystal gold nanoparticles mineralized in horse spleen ferritin without negative staining. c) Selected area electron diffraction pattern (SAED) for HSFn-AuNPs and the camera length is 20 cm. This figure is taken from Fan <i>et al.</i> ^[1]	103
Figure 3. 13 TEM images of HSFn-AuNPs a) and b) before and c)-f) after sucrose gradient ultracentrifugation. This figure is taken from Fan <i>et al.</i> ^[1] ...	104

List of Tables

Table 1. 1 Key properties of different ferritins.	7
Table 1. 2 Typical ferritins and their ferroxidase activity. The table is adapted from Abdallah <i>et al.</i> ^[30]	11
Table 1. 3 List of the inorganic nanoparticles mineralized inside ferritin. Some of this data was aggregated from Yoshimura <i>et al.</i> ^[109]	25
Table 2. 1 Buried surface area ($\text{\AA}^2$) per monomer in the interfaces of BFR and DPS. The data was taken from Vijayan <i>et al.</i> ^[32-33, 40]	42
Table 2. 2 Hydrophilic interactions stabilizing the interfaces in BFR and DPS. ^[33, 41] See appendix 2 for structural depictions of these interactions.	43
Table 2. 3 Optimal codons used to translate protein sequences to gene sequences. The table is adapted from Dong <i>et al.</i> ^[46]	46
Table 2. 4 Melting temperatures of proteins used in temperature dependent CD study. ΔT_m is the change in melting temperature with respect to the parent protein. This table is taken from Fan <i>et al.</i> ^[1]	68
Table 2. 5 Summarize the properties of BFR and DPS proteins which determined by CD, TEM, DLS, and SEC analysis.	78
Table 3. 1 Theoretical and observed d-spacings for gold nanoparticles mineralized in horse spleen ferritin. This table is taken from Fan <i>et al.</i> ^[1]	103

ABSTRACT

In this thesis, we demonstrate ferritins and ferritin-like proteins are ideal model systems to investigate protein quaternary structure and to fabricate nanomaterials. Ferritins and ferritin-like proteins self-assemble into hollow, nanoscale cages with octahedral symmetry and tetrahedral symmetry respectively. Although they share little sequence homology, they have homologous tertiary structure—both fold into a four-helix bundle structure. Despite their structural similarity, these monomers assemble into nanocages with different symmetries. To understand how the information stored in protein primary and tertiary structure defines the protein-protein interactions which govern quaternary structure formation, mutants were produced by mixing key domains from ferritin and ferritin-like protein. These mutants were studied by a number of biophysical techniques to determine their folding stability, self-assembling ability, and nanostructure. We also describe our attempts to fabricate gold nanoparticles within the cavities of native ferritin cages utilizing a novel approach. The key step is to fabricate gold nanoclusters within the ferritin and the clusters can initiate the formation of particles. We characterized the resulting nanoparticles utilizing a series of techniques. We confirmed that highly monodisperse nanoparticles are formed within intact protein shells and the size of the nanoparticles correlates with the interior diameter of the cages. These gold mineralized ferritin protein cages can be further utilized in bioimaging.

CHAPTER 1 Introduction^{*}

1.1 Protein structure

Proteins are amino acid-based polyamides that fold into complex structures. The protein structure is essential in determining the function of the protein. Traditionally, protein structure is classified into four levels.^[1] The primary structure of a protein is its amino acid sequence. Although only twenty amino acids are typically utilized in the production of most proteins, this affords a large diversity of protein sequences. Secondary structure is defined as regular folding regions which permit hydrogen bonds formation in the polypeptide chains, and it is also defined by the dihedral angles favored in each residue making up a polypeptide. The clustering of the conformational preferences results in the formation of α -helices and β -strands, among other structures, which are stabilized through different hydrogen bonding patterns. The tertiary structure is the overall folding of a protein and the spatial connectivity of the secondary structural elements. The primary driving force to form most tertiary structures is the exclusion of water from hydrophobic cores. In addition, ionic interactions and disulfide bonds also provide stability and structural specificity. Quaternary structure is the arrangement of two or more folded polypeptide chains (sometimes referred to as subunits) into multimeric protein complexes.

In nature, many proteins have interacting subunits that self assemble to form quaternary structure. The first multimeric protein was identified by Svedberg in the 1920s, when he determined the molecular weight of hemoglobin by sedimentation equilibrium. The molecular weight of hemoglobin was found to be

^{*} Parts of this chapter have been published previously with the author of this thesis as the first author. *Small*, **2010**, 6, 1483.

68000 Da, suggesting that the protein is made up of four subunits.^[2-3] The protein structure of hemoglobin was not fully determined until 1960s, when Perutz first resolved its X-ray crystal structure and further confirmed Svedberg's conclusion.^[4-5] Perutz's study also demonstrated the structural changes of hemoglobin that occur upon oxygen binding. The protein-protein interactions between subunits are essential for the formation of higher-order complexes and critical to achieve the biological function of proteins. Mutations at these interfaces can result in destruction of appropriate interactions and lead to diseases. For example, a single mutation of the hydrophilic glutamic acid to hydrophobic valine on the subunit interface of hemoglobin promotes aggregation and prevents the proper transportation of oxygen, resulting in sickle-cell anemia.^[6] Therefore, establishing an understanding of the underlying principles that govern protein self-assembly and subunit-subunit interactions is essential to appreciate the fundamental biological functions and to aid in the development of drugs.

1.2 Protein assembly

Protein quaternary structure and protein assembly are stabilized by weak forces. At subunit-subunit interfaces, hydrophobic interactions generally have a large contribution to protein assembly and in combination with polar interactions provide overall stability and specificity for subunit association. Moreover, the formation of disulfide bonds between the subunits can also favor protein assembly.

The oligomerization state of assembled proteins has been determined through various methods.^[7] The molecular weights of assembled proteins and their subunits can be determined by gel electrophoresis, gel filtration chromatography, mass spectrometry, analytical centrifugation and light scattering. Moreover, X-ray

crystallography can be applied to assembled proteins to elucidate their specific quaternary structures and also to investigate their subunit-subunit interactions. With the numerous methods that have been developed, many protein quaternary structures have been resolved, which permit scientists to study the general principles that control their association/oligomerization.

Proteins which form discrete oligomers through the self-assembly of identical subunits possesses internal symmetry which plays an essential role in controlling subunit association. Many of the homomeric proteins are classified into three types of symmetry point groups: cyclic, dihedral and cubic.^[8-9]

The cyclic symmetry group (C_n) has n -fold rotational symmetry and protein subunits in the C_n group are symmetrically arranged with $360^\circ/n$ rotations to form a ring (Figure 1. 1). Therefore, in the cyclic complexes, all interfaces are “face-to-back” (Figure 1. 2).^[8, 10-11] In nature, dimeric proteins with C_2 symmetry are common and have many functions. Protein complexes with large cyclic symmetry are not frequently found due to the limited contacts between the subunits. However, a few membrane proteins have been found with large cyclic symmetry, which make the structure of the proteins more compatible to the polarity of cell membranes.

The dihedral symmetry group (D_m) contains m -fold rotational symmetry and orthogonal two-fold symmetry (Figure 1. 1).^[8, 10-11] In dihedral complexes, along with the “face-to-back” contacts between the neighboring subunits around the cyclic ring, there are cross interactions on the “face-to-face” and “back-to-back” interfaces which are related to the perpendicular two-fold axes. As a result, with the same number of subunits in cyclic oligomers, dihedral complexes possess the possibility of higher stability than the cyclic oligomers due to greater burial of

surface area.^[11] Thus, the number of large and soluble oligomeric proteins that have been discovered to have dihedral symmetry is much higher than those possessing cyclic symmetry.

The cubic symmetry group contains multiple axes of rotation which are non-orthogonal to each other. This group consists of three subclasses (Figure 1. 1) which were first proposed by Crick and Watson.^[12] The tetrahedral (T) cubic group possesses non-perpendicular two-fold and three-fold axes. The octahedral (O) and icosahedral (I) cubic groups have additional four-fold and five-fold symmetry axes respectively along with those found in the T group (Figure 1. 1).

Figure 1. 1 Cartoon representation of cyclic, dihedral and cubic symmetry of oligomeric proteins. Symmetry axes are indicated as dotted lines. A circle, triangle, square, or pentamer at the end of an axis represents two-, three-, four- or five-fold symmetry respectively. The figure is taken from Janin *et al.*^[10]

Figure 1. 2 (left) Three-dimensional and (right) two-dimensional views of cyclic dimer (C2), cyclic tetramer (C4), dihedral dimer (D2) and dihedral tetramer (D4). Interactions on the face-to-back interfaces are involved in C4 and D4 complexes. In C2, D2 and D4 complexes, cross interactions are made on the face-to-face and back-to-back interfaces. The figure is taken from Levy *et al.*^[11]

1.3 Octahedral ferritins

Ferritin was first isolated from horse spleen as an iron storage protein by Laufberger in 1937,^[13] and ferritin proteins have been widely studied since then. The crystal structure of a ferritin was first resolved in 1991 which revealed its octahedral symmetry^[14] and many other ferritin proteins have been isolated and crystallized from diverse organisms such as animals,^[15-16] plants^[17-18] and bacteria.^[19-21] Although the amino acid sequences of these ferritins share little similarity, their tertiary and quaternary structures are highly conserved.^[22]

1.3.1 Tertiary and quaternary structure of octahedral ferritins

Ferritin consists of twenty-four structurally similar or identical subunits that self-assemble into a roughly spherical protein cage with octahedral symmetry.^[23-28] The exterior diameter of the protein is ~12 nm and the interior cavity diameter is ~8 nm. The empty cavity can accommodate up to 4500 iron atoms which are stored as nanoparticles inside the ferritin.^[29-31] Each subunit is made up of a four-helix bundle (A-D helices) with a short fifth helix (the E helix)

at the C-terminus (Figure 1. 3a). Subunit interactions lead to the formation of three-fold and four-fold channels through the well-packed protein shells. The three-fold channels are hydrophilic in general and thought to serve as pathways to transport iron atoms into the cavity (Figure 1. 3b).^[32] Whereas, the channels along the four-fold axes are tight and thought to be involved in proton transport (Figure 1. 3c).^[33-34]

Figure 1. 3 a) Ribbon diagram of a human light chain ferritin subunit. The A helix is colored purple, the B helix cyan, the C helix green, the D helix yellow and the E helix orange. The human light chain ferritin assembled protein cage is viewed along b) a three-fold axis and c) a four-fold axis. (Protein Data Bank ID is 2ffx and all figures are generated by using Chimera.)

1.3.2 Structural comparison between mammalian ferritins and bacterioferritins

In this thesis, mammalian ferritins and bacterioferritins will be utilized as model systems to study ferritin biomineralization and self-assembly respectively, so these proteins will be focused on and discussed in detail. Mammalian ferritins usually consist of two types of subunits, heavy (H) and light (L) chain with molecular weight of 21 and 19 kDa respectively.^[15] Their variations in sequence do not have significant influence on their overall structure. These two forms of subunits have a large propensity to form heteromers (although some can form homomers *in vitro*) and the relative ratio of each subunit varies with biological species, organs, and possibly cell state.^[35] The H-chain subunit has a conserved

ferroxidase site which is known to catalyze the oxidation reaction of Fe^{2+} to Fe^{3+} , while the L-chain subunit does not possess this function.^[15] However, L-chain ferritin protein cages have a higher number of negatively charged residues on their interior surfaces that serve as nucleation sites to accumulate iron in the cavity.^[15] Moreover, in most of mammalian ferritins, the three-fold channels are hydrophilic in nature, while the four-fold channels are largely hydrophobic (Table 1. 1).^[33]

Unlike mammalian ferritins, most bacterioferritins consist of twenty-four identical subunits and contain up to twelve heme groups within the protein shell.^[36] The heme group is located at the two-fold symmetry interface and often interacts with a pair of methionine residues (two-fold related Met52 in *E. coli* bacterioferritin, Figure 1. 4).^[19] However, these methionine residues are not essential in the bacterioferritin family and can be replaced with threonine residues.^[37] In addition, the heme binding pocket is normally lined with hydrophobic residues and the heme group is isolated from the outside environment by two-fold symmetrically related BC loops and the hemes point two propionate functional groups into the inner cavity (Figure 1. 4). Moreover, in comparison with mammalian ferritins, the three-fold axes in bacterioferritins are less hydrophilic, while the four-fold channels are more polar and hydrophilic (Table 1. 1).^[38]

Table 1. 1 Key properties of different ferritins.

Ferritin type	H-chain ferritin	L-chain ferritin	Bacterioferritin
Source	Mammalian	Mammalian	Bacteria
Symmetry	Octahedral symmetry	Octahedral symmetry	Octahedral symmetry
Number of subunits	24 subunits	24 subunits	24 subunits
Three-fold axis	Hydrophilic	Hydrophilic	Hydrophilic
Four-fold axis	Hydrophobic	Hydrophobic	Hydrophilic

Figure 1. 4 (left) The heme is located at the two-fold symmetric, dimer interface of *E. coli* bacterioferritins. (right) Expanded view. (The Protein Data Bank ID is 1bfr and all figures are generated by using Chimera.)

1.3.3 The mechanism of ferritin self-assembly

Only a few studies have been published that address the self-assembly mechanism of ferritins. Researchers have investigated the association and reassembly of horse spleen ferritin by utilizing different methods such as spectroscopy, chromatography, cross-linking, and analytical ultracentrifugation.^[39-40] These results have elucidated that the formation of the ferritin 24-mer occurs in a sequence of well-defined steps and possibly involves assembly intermediates that are monomers, dimers, trimers, and dodecamers. The detailed pathway is indicated by the following scheme:

Where $\mathcal{M}_1$ is unfolded subunit, M_i is the intermediates with i subunits, M_{24} is completely self-assembled ferritin.^[39-40] It has not been established if this mechanism is universal for the entire family nor if it changes under different conditions.

1.3.4 Mineralization of ferritin

Ferritin is known to be an iron storage protein that oxidizes toxic Fe^{2+} ions to crystalline ferrihydrite in the protein cavity using either oxygen or hydrogen peroxide.^[24-25] The iron oxidation and mineralization chemistry involve complicated, multi-step processes and some of the steps are still unresolved. In the overall reaction, Fe^{2+} ions first enter into the ferritin cavity and bind at the ferroxidase site, where they subsequently react with oxygen to form diiron complexes. The diiron species then undergo hydrolysis and migrate to a nucleation site to induce crystal growth.^[22, 27] Details of the process will be further discussed using mammalian ferritins as examples.

The hydrophilic three-fold axes have been implicated as the most likely pathways for the diffusion of Fe^{2+} to the ferroxidase centers. Electrostatic calculations using human H-chain ferritin have revealed that an electrostatic gradient along the three-fold axis could direct cations toward the inner cavity.^[33] Moreover, metal inhibition assays indicate that some metal ions (Zn^{2+} and Tb^{3+}) can bind the three-fold axes blocking Fe^{2+} access to the ferroxidase centers.^[41] These results reinforce the proposal that the three-fold axes are the main entrance for Fe^{2+} . In a recent study, Chasteen and coworkers have utilized stopped-flow spectroscopy to illustrate that the residues Thr135, His136, Thr137 might be involved in directing Fe^{2+} entry to the ferroxidase sites through the three-fold axes (Figure 1. 5).^[30, 42]

Figure 1. 5 (left) View along the three-fold axis of human H-chain ferritin (Protein Data Bank ID is 2cei) from the inside of the cage and (right) the expanded view of the iron pathway from the three-fold axis to the ferroxidase center. Residues D131 and E134 (blue labels) point into the three-fold channel and E27, Y34, E61, E62, H65, and Q141 (red labels) are involved in binding to iron atoms at the ferroxidase center. T135, H136, and Y137 (black labels) direct iron atoms to the ferroxidase center. The overall direction of the iron pathway is indicated by the orange arrow. The figure is adapted from Chasteen *et al.*^[30, 42]

After Fe^{2+} ions enter the cavity through the three-fold axes, the iron atoms are transferred to the ferroxidase centers possibly via conserved polar residues. The ferroxidase centers are reactive sites to catalyze iron oxidation and present only in the H-chain like subunits. Along with mammalian H-chain subunits, many other ferritin subunits such as bullfrog middle (M) chain ferritin subunits and *E. coli* bacterioferritin possess ferroxidase activity (Table 1. 2).^[32, 43] In human H-chain ferritin, the enzymatic site involves seven residues including Glu27, Tyr34, Glu61, Glu62, His65, Glu107, and Gln141 (Figure 1. 6a).^[44] Two Fe^{2+} ions react with oxygen at the ferroxidase site to form a transient μ -peroxo-diferric intermediate that subsequently decays within seconds to a μ -oxo-diferric complex (Figure 1. 6b) with concurrent release of H_2O_2 which conceivably can be used as an oxidant to oxidize more Fe^{2+} ions.^[44-45]

Table 1. 2 Typical ferritins and their ferroxidase activity. The table is adapted from Abdallah *et al.*^[30]

Ferritin	Subunit composition	Residues at ferroxidase center	Ferroxidase activity
Human H-chain ferritin ^[44-46]	24 H-chain subunits	E27, Y34, E61, E62, H65, E107, Q141	Yes
Human L-chain ferritin ^[15]	24 L-chain subunits	No	No
Bullfrog M-chain ferritin ^[47-48]	24 H-chain like subunits	E23, E58, H61, E103, Q137, D140	Yes
<i>Escherichia coli</i> bacterioferritin ^[19, 49]	24 H-chain like subunits	E18, E51, H54, E94, E127, H130	Yes

Figure 1. 6 a) The ferroxidase center is located entirely within a single subunit of human H-chain ferritin and in the X-ray structure, Zn^{2+} is bound to the ferroxidase center instead of Fe^{2+} . (Protein Data Bank ID is 2cei) b) Schematic depiction of the μ -oxo-diferric complex.

The resulting diferric complex disassociates from the ferroxidase center and iron slowly migrates to a nucleation site where it will begin the mineralization process. The pathway of iron translocation remains unresolved, although it is actively being investigated by Theil and coworkers using NMR.^[48] They have demonstrated that in frog M-chain ferritin the ferric clusters creep along each subunit from the ferroxidase center to the E helix when the concentration of iron is increased (Figure 1. 7). It is suggested that the E helices then guide the iron clusters deeper into the inner cavity to serve as nucleation sites initiating the mineralization of more iron.

Figure 1. 7 NMR structure indicating positions of associated iron with increasing iron concentration. The ferroxidase site is indicated in red. As the iron concentration increased with respect to protein, sites less proximal to the ferroxidase site become occupied. At low iron concentration only the light blue sites are occupied, whereas with increasing iron concentration cyan and then dark blue sites are also occupied suggesting a movement toward the E helix. The figure is taken from Theil *et al.*^[48]

1.4 Tetrahedral ferritin-like proteins

DNA-binding protein from starved cells (Dps) is a ferritin-like protein which was first isolated from *E. coli* culture under starvation conditions in 1992 and is found with tetrahedral symmetry.^[50-51] Afterward, many members of the Dps family have been discovered in bacteria^[52-55] and archaea.^[56-58]

1.4.1 Structure of Dps proteins

The tertiary and quaternary structures of the Dps proteins are well conserved.^[59] Each subunit of Dps folds into a four-helix bundle (A-D helices) with an additional fifth helix (the BC helix) located in the middle of the loop which connects the B and C helices (Figure 1. 8a).^[51] Twelve identical subunits are assembled to form a quasi-spherical, hollow protein cage with an interior diameter of ~4.5 nm and an exterior diameter of ~9 nm (Figure 1. 8b).^[51] The Dps protein cage adopts tetrahedral symmetry with two types of three-fold axes.^[51] One type resembles the packing manner of the three-fold axes in ferritin (Figure 1. 8c).

These ferritin-like three-fold axes in Dps connect its internal cavity with the external environment, and are lined with negatively charged residues to facilitate the entry of cations (Figure 1. 8d). The other type of three-fold axes is constructed mostly by the C-termini of the subunits and is referred to as the Dps-like three-fold axis (Figure 1. 8b).

Figure 1. 8 a) Ribbon diagram of a *E. coli* Dps subunit. The A helix is colored purple, the B helix cyan, the BC helix green, the C helix yellow and the D helix orange. The Dps protein cage is viewed along b) Dps-like three-fold axis and c) ferritin-like three-fold axis. d) Magnified diagram of ferritin-like pores lined with aspartic acid residue. (Protein Data Bank is 1dps and all figures are generated with Chimera.)

1.4.2 DNA binding by Dps

It has been known for many years that Dps proteins can protect DNA from damage^[59] and some Dps proteins can bind to DNA without apparent sequence specificity.^[50, 60] The DNA binding region in *E. coli* Dps is highly dynamic and involves the solvent-exposed N-terminus. It contains three lysine residues which facilitate electrostatic interactions with negatively charged DNA. Large and stable

complexes can be formed between wild type Dps proteins and DNA plasmids at neutral pH which are resistant to DNase.^[50] Successive removal of the N-terminal lysine residues resulted in the gradual loss of DNA binding and condensation ability.^[60] Moreover, bacterial mutants lacking Dps proteins showed significant changes in protein expression during starved conditions.^[50] Therefore, Dps proteins seem to play an important cellular role in regulating gene expression level and protecting DNA from damage.^[50, 60]

1.4.3 The ferroxidase center

It is conceivable that Dps also protects DNA from oxidative damage through its ferroxidase activity. Dps proteins have a strong preference to utilize hydrogen peroxide to oxidize iron and subsequently sequester up to 500 iron atoms in the protein cavity thereby protecting DNA from oxidative damage.^[59] During this process, iron atoms first coordinate with the ferroxidase centers in Dps and the iron ions are rapidly oxidized by hydrogen peroxide and mineralize in the negatively charged inner cavity (Note: the overall reaction is shown in eq 1.1).^[61] In Dps proteins, the ferroxidase center is located at the dimeric two-fold symmetry interface and the residues involved in iron binding are highly conserved (Figure 1. 9).^[62-63] Most of the ferroxidase sites in the Dps family are composed of two histidine residues (H51 and H63 in *E. coli* Dps) from one subunit and two carboxylate side chain-containing residues (D78 and E82 in *E. coli* Dps) on the two-fold symmetry-related subunit (Figure 1. 9).^[63-64] It is interesting to note that all members of the Dps family have ferroxidase centers, while, only some can bind DNA.^[65]

Figure 1. 9 (left) The dimeric two-fold symmetry interface of *E. coli* Dps, (right) illustrating the iron binding residues at the ferroxidase center with two water molecules (orange) occupying at the iron binding sites. (Protein Data Bank ID is 1dps and all figures are generated by using Chimera.)

1.5 Icosahedral viruses

Icosahedral symmetry for small viruses was proposed, with little structural data, by Crick and Waston in 1956.^[12] It was thought that this symmetry is preferred due to the fact that it is one of the most efficient ways of assembling viruses with identical subunits.^[12] A geometric icosahedron with 532-point symmetry is comprised of twelve vertices, twenty equilateral triangular facets, and thirty edges. Five-fold axes pass through vertices to the center of the icosahedron; three-fold and two-fold axes run through the center of the facets and the edges respectively (Figure 1. 10). To form an icosahedral virus, at least sixty identical subunits are required as three subunits for one facet of an icosahedron. This is referred to as a T1 virus (see below and Figure 1. 12).

Figure 1. 10 Views of icosahedron down from two-, three-, and five-fold symmetry axes.

1.5.1 Quasi-equivalence

After icosahedral symmetry had been proposed, scientists found that many spherical viruses consist of more than sixty subunits. Therefore, in 1961, Caspar and Klug proposed the concept of quasi-equivalence to describe how to build up larger viruses with icosahedral symmetry and multiples of sixty subunits.^[66] In their theory, the T (triangulation) number is defined as the number of subunits contained in the repeating asymmetric unit and three asymmetric units make up of a triangular icosahedral facet (Figure 1. 11). Therefore each virus is comprised of $60T$ subunits ($T \times 3 \times 20$ facets = $60T$ subunits). Moreover, the T number is also defined by the equation: $T = h^2 + hk + k^2$, where h and k are coordinates on two axes which intersect with a 60° relationship. The coordinates can be determined by counting the number of vertices to walk over along the axes from the vertex of one pentamer to the vertex of an adjacent pentamer (Figure 1. 12).^[67-69]

Figure 1. 11 Selected icosahedral symmetry capsids with different T numbers. The figure is taken from Caspar *et al*^[66] and Zlotnick *et al*.^[68]

Figure 1. 12 (left) The facet geometry for icosahedral viruses with quasi-equivalence is indicated by two coordinates (h and k). (right) They are determined by counting the number of vertices from the vertex of one pentamer to the vertex of an adjacent pentamer. See text for further explanation. The figure is taken from Zlotnick.^[67]

1.5.2 Cowpea chlorotic mottle virus

Cowpea chlorotic mottle virus (CCMV) is a classic icosahedral plant virus and it contains three unique, single stranded RNA molecules inside the protein shell.^[70-71] The highly negatively charged RNA is found to interact with the highly positively charged interior capsid surface. CCMV self-assembles from 180 protein subunits ($T=3$) into an icosahedral architecture. The exterior diameter is ~ 28 nm and the interior cavity is ~ 24 nm in dimension.^[70-71] It has been reported that CCMV is sensitive to pH and metal ions. It undergoes swelling by increasing the dimension up to 10% and allows molecules to access the interior cavity. Investigation into the self-assembly of CCMV indicates that the interactions between the nucleic acid and the virus capsid, in combination with the subunit-subunit hydrophobic interactions are the main driving forces to form viruses.^[72] In addition, this virus has been used extensively in the templation of nanomaterials.

1.6 Protein cages applied to the templated mineralization of nanomaterials

Nanomaterials have attracted extensive attention as a result of their emerging properties and their potential for a multitude of applications such as electronics, catalysis, sensors, and medical diagnosis.^[73-75] With the hope of discovering novel properties for future applications, various methods have been developed to synthesize inorganic nanomaterials. The use of biological systems, inspired by naturally evolved processes, is an emerging trend in their fabrication, and it has been reported that material size, shape, and morphology can be controlled by interactions between biomolecules and inorganic materials.^[76] Proteins and peptides, due to their ready availability and large structural and functional diversity, have high utility in the manipulation of materials. Moreover, biotemplate-directed syntheses have the potential to be more ‘green’ than traditional methods due to the required mild reaction conditions such as lower temperature, near neutral pH, and the fact that they often employ aqueous reaction solutions. Proteins that assemble into nanocage structures create robust, size-constrained nanocontainers which have been often utilized as templates to produce many types of nanomaterials.^[76] The size of protein cages varies from several nanometers to hundreds of nanometers and shape from spherical to nanotube-like structures, where these characteristics are important to determine the properties of the materials.

1.6.1 Ferritins as templates to mineralize nanoparticles

Ferritins can store iron as ferrihydrite nanoparticles inside their cavities and they exhibit remarkable stability withstanding high temperature (up to 70 °C) and pH extremes (many are stable from pH 2 to 10) without significant disruption to their quaternary structure. These properties have inspired work to demonstrate the

feasibility of removing the mineralized cores and utilizing the empty cages (i.e., apoferritin) as size-constrained reaction vessels to synthesize different types of nanoparticles including metal oxides,^[77-80] hydroxides or oxyhydroxides,^[81-85] carbonates^[86], semiconductors,^[87-92] and metals.^[93-106] Many of these nanoparticles possess a narrow size distribution arising from growth restriction within the cage whose uniformity is a result of the precision of protein self-assembly. Moreover, it is thought that the protein cages could enhance the solubility and chemical stability of the particles.

1.6.1.1 Metal Oxides, hydroxides and oxyhydroxides

Non-native metal oxides and hydroxides were the first unnatural particles to be mineralized in apoferritin. Mann and coworkers showed that manganese oxide nanoparticles could be fabricated within ferritin cavities.^[77] Subsequently, Douglas and coworkers mineralized cobalt oxide and cobalt hydroxide inside horse spleen ferritin. Aliquots of Co^{2+} and H_2O_2 were added into the non-buffered ferritin solution which was dynamically titrated (using an autotitrator) to pH 8.5.^[83] In these two reports, the success of the mineralization mainly relied upon complementary electrostatic interactions between the negatively charged interior surface of ferritin and oppositely charged metal ions resulting in a higher local concentration of metal ions in the cavity than in solution and promotes the nanoparticle formation inside the cage.

Subsequently Yoshimura and coworkers reported that nickel hydroxide nanoparticles can be successfully synthesized in horse spleen ferritin.^[84] They found that nickel ions rapidly hydrolyze in solution and precipitate along with the proteins. To decrease the rate of bulk precipitation, ammonium ions were added to the solution to serve as ligands for the metal ions. However, the cores still could

not form in the cavity of ferritin in the presence of ammonium ions. Subsequently, they demonstrated that by precisely maintaining the pH of the buffer and adding carbonate ions, nickel ions were accumulated and fabricated in the ferritin cavity with high efficiency (Figure 1. 13). Using similar conditions, chromium hydroxide nanoparticles could also be synthesized.

Figure 1. 13 TEM images of nickel hydroxide fabrication in different conditions. The ferritin solution was bubbled with a) carbon dioxide (to provide carbonate ions) and b) nitrogen. The figure is taken from Yoshimura *et al.*^[84]

1.6.1.2 Metal carbonates

Nanoparticles made up of alkaline earth metal carbonates (such as CaCO_3 , SrCO_3 and BaCO_3) have been also prepared in apoferritin using a novel approach.^[86] A challenge of the mineralization is that cations from group IIA have low binding affinities to the interior of apoferritin.^[107] As a result, there is no preference to mineralize nanoparticles within the ferritin cage. To suppress metal carbonate mineralization on the outside the ferritin, polyelectrolytes such as poly(methacrylic acid) or sodium polyphosphate were added to the reaction solution to inhibit bulk crystallization under the assumption that these polymers are too large to diffuse into the protein cage therefore selectively promoting mineralization only inside the protein.

1.6.1.3 Semiconductors

Ferritins have been utilized as size-constrained reaction vessels to mineralize various nano-scaled semiconductors such as quantum dots (QDs) since they have great potential in electronics and imaging. Generally chemical methods to produce QDs require high temperature and organic solvents. Moreover, the surfactants used to stabilize QDs restrict their water solubility and limit their biomedical applications. To overcome these problems and to control their size, it was thought that protein cages could be used.

Apo ferritins were first used to synthesize QDs by Mann and coworkers.^[87] Previously they had shown that Cd^{2+} ions could diffuse through the channels into the protein cage and bind to the inner cavity with high stoichiometry.^[107] To fabricate CdS QDs, aliquots of Cd^{2+} and HS^- were added into the ferritin solution. During the first addition, CdS clusters formed and nucleated on the protein inner surface. With subsequent additions, the clusters functioned as nucleation centers for the mineralization of the nanoparticles.^[87]

In a subsequent report, Yamashita and coworkers developed a slow chemical reaction strategy to fabricate CdSe inside ferritin.^[88] They noticed that the reaction between Cd^{2+} and Se^{2-} ions in aqueous solution was too fast to control and a lot of bulk precipitations were formed. Therefore, to slow down the reaction rate, ammonium ions were added to stabilize the Cd^{2+} ions in the form of $\text{Cd}(\text{NH}_3)_4^{2+}$. To slowly release Se^{2-} into the solution, the degradation of selenourea was employed. As a result, bulk precipitation was suppressed and CdSe nanoparticles were formed in the apoferritin (Figure 1. 14). By applying a similar strategy, Yamashita and coworkers also synthesized ZnSe nanoparticles in ferritin.^[89]

Figure 1. 14 TEM images of CdSe nanoparticles mineralized in ferritin. The samples were stained with (left) aurothioglucose and (right) without staining. The inset is a high resolution image to demonstrate atomic ordering. The figure is taken from Yamashita *et al.*^[88]

Guo and coworkers have also successfully synthesized CdSe nanoparticles inside ferritin. They used EDTA as the stabilization reagent instead of ammonium ions.^[92] It is thought that EDTA/Cd²⁺ chelates can not penetrate into the cavity due to their large size. Therefore, free Cd²⁺ ions accumulate in the inner cavity and react with Se²⁻ to form the nanoparticles.^[92]

In a recent report, PdS QDs have been synthesized in the ferritin cage by Thomas and coworkers.^[91] They mixed lead acetate trihydrate (Pb(AcO)₂·3H₂O), sodium sulphide (Na₂S), thiol capping agents with ferritin solution to prepare the nanoparticles. The resulting PdS nanoparticles had stable fluorescent emission in the near-infrared wavelength region and the protein cage rendered PdS QDs with water solubility. In addition, near-infrared QDs have great potential *in vivo* for deep-tissue imaging.

1.6.1.4 Metals

Watanabe and coworkers have successfully synthesized Pd nanoclusters inside apoferritin. They mixed K₂PdCl₄ with a solution of apoferritin at pH 8.5 and Pd²⁺ ions were subsequently reduced by NaBH₄ to yield Pd clusters. TEM images indicated that the resulting Pd clusters were monodisperse, spherical particles with a diameter of 2.0±0.3 nm (Figure 1. 15). The channels on the ferritin selectively

permitted the diffusion of small olefin substrates to enter into the ferritin cavity for subsequent hydrogenation.

Figure 1. 15 TEM image of negatively stained ferritin with mineralized Pd nanoclusters. The scale bar is 50 nm. The figure is taken from Watanabe *et al.*^[96]

It has been found that silver ions could not be specifically mineralized inside the native ferritin with respect to the bulk solution and two different strategies have been designed to overcome this problem. In the first strategy, a phage display-selected peptide has been utilized. Naik and coworkers reported that the dodecapeptide AG4 (NPSSLFRYLPSD) could reduce silver ions to metallic silver nanoparticles at neutral pH.^[108] The ferritin was endowed with silver binding affinity by fusing the AG4 peptide to the C-terminus which faces the interior of the cavity and does not disturb the protein self-assembly. The purified recombinant apoferritin was incubated with AgNO₃ at 37 °C for several hours and silver nanocrystals were mineralized within the cavity.^[95] In the second strategy, Dominguez-Vera and coworkers incubated silver ions with ferritin solution at pH 8 for 24 h. The silver ions that were not incorporated into the ferritin were removed by dialysis or chromatography. Hence, silver ions were only present within the cavity and they were quickly reduced by NaBH₄ to form nanocrystals.^[101]

1.6.1.5 General principles for the templation of nanoparticles inside ferritin

In conclusion, quite a few different nanoparticles have been mineralized in ferritin using multiple strategies (Table 1. 3). Many of the strategies have capitalized on natural electrostatic interactions or specific binding between metal ions and the interior surface of native ferritins^[77, 83] to increase the local concentration and thus facilitate the formation of nanoparticles. It has been assumed due to the anionic nature of the ferritin cavity, and the direction of the electrostatic gradient in the ion-entry channels at the three-fold axes (see above), that only cations would be successful with this strategy. Other strategies respond to the fact that some metal ions have a natural affinity for the ferritin exterior, or no affinity preference to either the interior or the exterior, resulting in substantial mineralization on the outside of the ferritin. In one remedy to this problem, ion-bound ferritins are first subjected to dialysis or chromatography before subsequent reduction inside the cavity.^[105] Alternatively, ammonium ions, EDTA, or polyelectrolytes have been included in the reaction solution to retard or prevent mineralization on the outside of the cage.^[84, 86, 88-89, 92] Another strategy has been to genetically or chemically modify the proteins to endow the cavity with an enhanced ion binding affinity or the ability to promote particle formation.^[95, 102]

Table 1. 3 List of the inorganic nanoparticles mineralized inside ferritin. Some of this data was aggregated from Yoshimura *et al.*^[109]

Materials	Type of ferritin	Ion source	reference
Metal oxides			
Mn ₃ O ₄	Horse spleen ferritin	Mn ²⁺	[77]
Co ₃ O ₄	Horse spleen ferritin	Co ²⁺	[78]
	Recombinant horse spleen light chain ferritin		[79]
In ₂ O ₃	Recombinant horse liver light chain ferritin	In ³⁺	[80]
Metal hydroxide and oxyhydroxides			
Mn(O)OH	Horse spleen ferritin	Mn ²⁺	[81-82]
Co(O)OH	Horse spleen ferritin	Co ²⁺	[83]
Eu(O)OH	Horse spleen ferritin	Eu ²⁺	[85]
Ni(OH) ₃	Horse spleen ferritin	Ni ²⁺	[84]
	Recombinant horse spleen light chain ferritin		
Cr(OH) ₃	Horse spleen apoferritin	Cr ³⁺	[84]
	Recombinant horse spleen light chain ferritin		
Metal Carbonates			
CaCO ₃ , SrCO ₃ , BaCO ₃	Horse spleen ferritin	Ca ²⁺ , Sr ²⁺ , Ba ²⁺ , CO ₃ ²⁻	[86]
Semiconductors			
CdS	Horse spleen ferritin	Cd ²⁺ , HS ⁻	[87]
CdSe	Horse spleen ferritin	Cd ²⁺ , Se ²⁻	[88, 90, 92]
ZnSe	Recombinant horse liver light chain ferritin	Zn ²⁺ , Se ²⁻	[89]
PbS	Horse spleen ferritin	Pb ²⁺ , HS ⁻	[91]
Metals			
Pd	Horse spleen ferritin	PdCl ₄ ²⁻	[96, 100]
Ag	Recombinant human light chain ferritin	Ag ⁺	[95]
	Mutated human heavy chain ferritin		[102]
	Horse spleen ferritin		[105]
	<i>Pyrococcus furiosus</i> ferritin		[106]
Cu	Horse spleen ferritin	Cu ²⁺	[97, 103-105]
Ni	Horse spleen ferritin	Ni ²⁺	[99, 105]
Co	Horse spleen ferritin	Co ²⁺	[99, 105]
Au	Mutated human heavy chain ferritin	AuCl ₄ ⁻	[102]
CoPt	Horse spleen ferritin	Co ²⁺ , PdCl ₄ ²⁻	[93]

1.6.2 Dps proteins as templates to mineralize nanoparticles

As described above, Dps proteins adopt cage-like structures and the empty cages can be used for nanoparticle mineralization in analogy to the ferritins. The inner cavity of Dps is ~4.5 nm in diameter and it can accommodate up to 500 iron atoms. Moreover, Dps proteins have comparable stability to ferritins. Therefore, Dps protein cages are considered as ideal reaction vessels for nanoparticle mineralization that are complementary to ferritin. Different nanoparticles including Co(O)OH,^[110] Co₃O₄,^[110-111] and CdS^[112] have been fabricated within Dps protein

cavities using similar techniques and philosophies described above for ferritin. The resulting nanoparticles are ~4 nm in diameter which is much smaller than those mineralized in ferritin resulting in different properties.

1.6.3 Viruses as templates to mineralize nanoparticles

Spherical viruses are complicated, highly organized biological supramolecules (see above) and their size varies from several to hundred nanometers in dimension. Therefore it is expected that the size of nanoparticles mineralized within these viruses would reflect this diversity. Many early studies of biotemplated-material synthesis utilize cowpea chlorotic mottle virus (CCMV) as reaction vessels. In Douglas and coworkers' pioneering work, empty CCMV viral capsids were used as templates, and the positively charged interior surface nucleated the mineralization of anionic polyoxometalate salts (tungstates $\text{H}_2\text{W}_{12}\text{O}_{42}^{10-}$ and vanadates $\text{V}_{10}\text{O}_{28}^{6-}$).^[113] In subsequent work, the interior surface charge of the virus was switched by mutating the positively charged amino acids to negatively charged glutamic acids. The electrostatically altered viral protein cages self-assembled into cages resembling the wild type CCMV. This mutated virus was employed to promote the nucleation of Fe_3O_4 and Co_2O_3 inside the highly positively charged cavity.^[114]

1.6.4 Summary

The utilization of protein cages as biotemplates for nanomaterials synthesis has many advantages. In nature, many protein cages have different shapes, sizes and associated properties and the same protein molecules have identical structure. The properties of nanomaterials are highly sensitive to size and shape and being able to control their dimensions precisely, consistently, and rationally. In addition, many protein cages can be easily expressed and purified in large scale. They can

be chemically and genetically modified to assist in mineralization, solubilization properties or to afford bioconjugation. It could be easily envisioned that the resulting materials could be used for catalytic reactions, cell specific targeting probes, and MRI contrast agents among other applications.

1.7 Research perspectives

In this chapter, the structure, self-assembly and biotemplate synthesis of protein cages have been discussed in detail. The aims of this thesis were to gain insight into the fundamentals of bacterioferritin octahedral and Dps tetrahedral assembly, and to develop a novel method to template inorganic nanomaterials using ferritin.

1.8 References

1. Campbell, N. A., and Reece, J. B. (2002) *Biology*, 6th ed., Benjamin Cummings, San Francisco.
2. Svedberg, T., and Fahraeus, R. (1926) A new method for the determination of the molecular weight of the proteins, *Journal of the American Chemical Society* 48, 430-438.
3. Svedberg, T., and Nichols, J. B. (1927) The application of the oil turbine type of ultracentrifuge to the study of the stability region of carbon monoxide-hemoglobin, *Journal of the American Chemical Society* 49, 2920-2934.
4. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960) Structure of haemoglobin: A three-dimensional fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis, *Nature* 185, 416-422.
5. Perutz, M. F. (1960) Structure of hemoglobin, *Brookhaven symposia in biology* 13, 165-183.
6. Pauling, L., Itano, H. A., Singer, S. J., and Wells, I. C. (1949) Sick cell anemia, a molecular disease, *Science* 110, 543-548.
7. Laue, T. M., and Rhodes, D. G. (1990) Determination of size, molecular-weight, and presence of subunits, *Methods in Enzymology* 182, 566-587.
8. Goodsell, D. S., and Olson, A. J. (2000) Structural symmetry and protein function, *Annual Review of Biophysics and Biomolecular Structure* 29, 105-153.
9. Levy, E. D. (2007) PiQSi: Protein quaternary structure investigation, *Structure* 15, 1364-1367.

10. Janin, J., Bahadur, R. P., and Chakrabarti, P. (2008) Protein-protein interaction and quaternary structure, *Quarterly Reviews of Biophysics* 41, 133-180.
11. Levy, E. D., Erba, E. B., Robinson, C. V., and Teichmann, S. A. (2008) Assembly reflects evolution of protein complexes, *Nature* 453, 1262-U1266.
12. Crick, F. H. C., and Watson, J. D. (1956) Structure of small viruses, *Nature* 177, 473-475.
13. Laufberger, V. (1937) Sur la cristallisation de la ferritine, *Bull. Soc. Chim. Biol* 19, 1575-1582.
14. 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., and Harrison, P. M. (1991) Solving the structure of human H-ferritin by genetically engineering intermolecular crystal contacts, *Nature* 349, 541-544.
15. Hempstead, P. D., Yewdall, S. J., Fernie, A. R., Lawson, D. M., Artymiuk, P. J., Rice, D. W., Ford, G. C., and Harrison, P. M. (1997) Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution, *J. Mol. Biol.* 268, 424-448.
16. Granier, T., Gallois, B., d'Estaintot, A. D., Dautant, A., Chevelier, J. M., Mellado, J. M., Beaumont, C., Santambrogio, P., Arosio, P., and Precigoux, G. (2001) Structure of mouse L-chain ferritin at 1.6 angstrom resolution, *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 57, 1491-1497.
17. Vanwuytswinkel, O., Savino, G., and Briat, J. F. (1995) Purification and characterization of recombinant pea-seed ferritins expressed in *Escherichia coli* - influence of N-terminus deletions on protein solubility and core formation in-vitro, *Biochemical Journal* 305, 253-261.
18. Wardrop, A. J., Wicks, R. E., and Entsch, B. (1999) Occurrence and expression of members of the ferritin gene family in cowpeas, *Biochemical Journal* 337, 523-530.
19. Dautant, A., Meyer, J. B., Yariv, J., Precigoux, G., Sweet, R. M., Kalb, A. J., and Frolow, F. (1998) Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from *E. coli*, *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 54, 16-24.
20. Cobessi, D., Huang, L. S., Ban, M., Pon, N. G., Daldal, F., and Berry, E. A. (2002) The 2.6 angstrom resolution structure of *Rhodobacter capsulatus* bacterioferritin with metal-free dinuclear site and heme iron in a crystallographic 'special position', *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 58, 29-38.
21. Macedo, S., Romao, C. V., Mitchell, E., Matias, P. M., Liu, M. Y., Xavier, A. V., LeGall, J., Teixeira, M., Lindley, P., and Carrondo, M. A. (2003) The nature of the di-iron site in the bacterioferritin from *Desulfovibrio desulfuricans*, *Nature Structural Biology* 10, 285-290.
22. Andrews, S. C., Arosio, P., Bottke, W., Briat, J. F., Vondarl, M., Harrison, P. M., Laulhere, J. P., Levi, S., Lobreaux, S., and Yewdall, S. J. (1992) Structure, function, and evolution of ferritins, *J. Inorg. Biochem.* 47, 161-174.
23. Arosio, P., Ingrassia, R., and Cavadini, P. (2009) Ferritins: A family of molecules for iron storage, antioxidation and more, *Biochimica Et Biophysica Acta-General Subjects* 1790, 589-599.

24. Crichton, R. R. (1971) Ferritin - Structure, synthesis and function, *N. Engl. J. Med.* 284, 1413-1422.
25. Clegg, G. A., Fitton, J. E., Harrison, P. M., and Treffry, A. (1980) Ferritin - Molecular-structure and iron-storage mechanisms, *Progress in Biophysics & Molecular Biology* 36, 53-86.
26. Theil, E. C. (1983) Ferritin - Structure, function, and regulation, *Advances in Inorganic Biochemistry* 5, 1-38.
27. Theil, E. C. (1990) The ferritin family of iron storage proteins, *Advances in Enzymology and Related Areas of Molecular Biology* 63, 421-449.
28. Theil, E. C. (1987) Ferritin - Structure, gene-regulation, and cellular function in animals, plants, and microorganisms, *Annual Review of Biochemistry* 56, 289-315.
29. Andrews, S. C. (2010) The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor, *Biochimica Et Biophysica Acta-General Subjects* 1800, 691-705.
30. Bou-Abdallah, F. (2010) The iron redox and hydrolysis chemistry of the ferritins, *Biochimica et Biophysica Acta (BBA) - General Subjects* 1800, 719-731.
31. May, C. A., Grady, J. K., Laue, T. M., Poli, M., Arosio, P., and Chasteen, N. D. (2010) The sedimentation properties of ferritins. New insights and analysis of methods of nanoparticle preparation, *Biochimica Et Biophysica Acta-General Subjects* 1800, 858-870.
32. Harrison, P. M., and Arosio, P. (1996) Ferritins: Molecular properties, iron storage function and cellular regulation, *Biochimica Et Biophysica Acta-Bioenergetics* 1275, 161-203.
33. Douglas, T., and Ripoll, D. R. (1998) Calculated electrostatic gradients in recombinant human H-chain ferritin, *Protein Science* 7, 1083-1091.
34. Theil, E. C., Takagi, H., Small, G. W., He, L., Tipton, A. R., and Danger, D. (2000) The ferritin iron entry and exit problem, *Inorganica Chimica Acta* 297, 242-251.
35. Grossman, M. J., Hinton, S. M., Minakberner, V., Slaughter, C., and Stiefel, E. I. (1992) Unification of the ferritin family of proteins, *Proc. Natl. Acad. Sci. U. S. A.* 89, 2419-2423.
36. Carrondo, M. A. (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint, *Embo Journal* 22, 1959-1968.
37. Laulhere, J. P., Laboure, A. M., Vanwuytswinkel, O., Gagnon, J., and Briat, J. F. (1992) Purification, characterization and function of bacterioferritin from the cyanobacterium *Synechocystis* PCC-6803, *Biochemical Journal* 281, 785-793.
38. Frolov, F., Kalb, A. J., and Yariv, J. (1993) Location of heme in bacterioferritin of *Escherichia coli*, *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 49, 597-600.
39. Gerl, M., and Jaenicke, R. (1987) Mechanism of the self-assembly of apoferritin from horse spleen - Cross-linking and spectroscopic analysis, *European Biophysics Journal with Biophysics Letters* 15, 103-109.
40. Gerl, M., Jaenicke, R., Smith, J. M. A., and Harrison, P. M. (1988) Self-assembly of apoferritin from horse spleen after reversible chemical modification with 2,3-dimethylmaleic anhydride, *Biochemistry* 27, 4089-4096.

41. Bou-Abdallah, F., Arosio, P., Levi, S., Janus-Chandler, C., and Chasteen, N. D. (2003) Defining metal ion inhibitor interactions with recombinant human H- and L-chain ferritins and site-directed variants: an isothermal titration calorimetry study, *Journal of Biological Inorganic Chemistry* 8, 489-497.
42. Bou-Abdallah, F., Zhao, G., Biasiotto, G., Poli, M., Arosio, P., and Chasteen, N. D. (2008) Facilitated diffusion of iron(II) and dioxygen substrates into human H-chain ferritin. A fluorescence and absorbance study employing the ferroxidase center substitution Y34W, *Journal of the American Chemical Society* 130, 17801-17811.
43. Chasteen, N. D. (1998) Ferritin. Uptake, storage, and release of iron, in *Metal Ions in Biological Systems, Vol 35*, pp 479-514.
44. Yang, X. K., Chen-Barrett, Y., Arosio, P., and Chasteen, N. D. (1998) Reaction paths of iron oxidation and hydrolysis in horse spleen and recombinant human ferritins, *Biochemistry* 37, 9743-9750.
45. Bou-Abdallah, F., Papaefthymiou, G. C., Scheswohl, D. M., Stanga, S. D., Arosio, P., and Chasteen, N. D. (2002) μ -1,2-peroxobridged di-iron(III) dimer formation in human H-chain ferritin, *Biochemical Journal* 364, 57-63.
46. Toussaint, L., Bertrand, L., Hue, L., Crichton, R. R., and Declercq, J. P. (2007) High-resolution X-ray structures of human apoferritin H-chain mutants correlated with their activity and metal-binding sites, *J. Mol. Biol.* 365, 440-452.
47. Tosha, T., Hasan, M. R., and Theil, E. C. (2008) The ferritin Fe₂ site at the diiron catalytic center controls the reaction with O₂ in the rapid mineralization pathway, *Proc. Natl. Acad. Sci. U. S. A.* 105, 18182-18187.
48. Turano, P., Lalli, D., Felli, I. C., Theil, E. C., and Bertini, I. (2010) NMR reveals pathway for ferric mineral precursors to the central cavity of ferritin, *Proc. Natl. Acad. Sci. U. S. A.* 107, 545-550.
49. Yang, X. O., Le Brun, N. E., Thomson, A. J., Moore, C. R., and Chasteen, N. D. (2000) The iron oxidation and hydrolysis chemistry of *Escherichia coli* bacterioferritin, *Biochemistry* 39, 4915-4923.
50. Almiron, M., Link, A. J., Furlong, D., and Kolter, R. (1992) A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*, *Genes & Development* 6, 2646-2654.
51. Grant, R. A., Filman, D. J., Finkel, S. E., Kolter, R., and Hogle, J. M. (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA, *Nature Structural Biology* 5, 294-303.
52. Ilari, A., Savino, C., Stefanini, S., Chiancone, E., and Tsernoglou, D. (1999) Crystallization and preliminary X-ray crystallographic analysis of the unusual ferritin from *Listeria innocua*, *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 55, 552-553.
53. Ceci, P., Ilari, A., Falvo, E., and Chiancone, E. (2003) The Dps protein of *Agrobacterium tumefaciens* does not bind to DNA but protects it toward oxidative cleavage, *J. Biol. Chem.* 278, 20319-20326.
54. Papinutto, E., Dundon, W. G., Pitulis, N., Battistutta, R., Montecucco, C., and Zanotti, G. (2002) Structure of two iron-binding proteins from *Bacillus anthracis*, *J. Biol. Chem.* 277, 15093-15098.

55. Kauko, A., Haataja, S., Pulliainen, A. T., Finne, J., and Papageorgiou, A. C. (2004) Crystal structure of *Streptococcus suis* Dps-like peroxide resistance protein Dpr: Implications for iron incorporation, *J. Mol. Biol.* 338, 547-558.
56. Wiedenheft, B., Mosolf, J., Willits, D., Yeager, M., Dryden, K. A., Young, M., and Douglas, T. (2005) An archaeal antioxidant: Characterization of a Dps-like protein from *Sulfolobus solfataricus*, *Proc. Natl. Acad. Sci. U. S. A.* 102, 10551-10556.
57. Ramsay, B., Wiedenheft, B., Allen, M., Gauss, G. H., Lawrence, C. M., Young, M., and Douglas, T. (2006) Dps-like protein from the hyperthermophilic archaeon *Pyrococcus furiosus*, *J. Inorg. Biochem.* 100, 1061-1068.
58. Reindel, S., Schmidt, C. L., Anemuller, S., and Matzanke, B. F. (2002) Characterization of a non-haem ferritin of the Archaeon *Halobacterium salinarum*, homologous to Dps (starvation-induced DNA-binding protein), *Biochem. Soc. Trans.* 30, 713-715.
59. Liu, X. F., and Theil, E. C. (2005) Ferritins: Dynamic management of biological iron and oxygen chemistry, *Accounts of Chemical Research* 38, 167-175.
60. Martinez, A., and Kolter, R. (1997) Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps, *Journal of Bacteriology* 179, 5188-5194.
61. Su, M. H., Cavallo, S., Stefanini, S., Chiancone, E., and Chasteen, N. D. (2005) The so-called *Listeria innocua* ferritin is a Dps protein. Iron incorporation, detoxification, and DNA protection properties, *Biochemistry* 44, 5572-5578.
62. Bozzi, M., Mignogna, G., Stefanini, S., Barra, D., Longhi, C., Valenti, P., and Chiancone, E. (1997) A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in *Listeria innocua*, *J. Biol. Chem.* 272, 3259-3265.
63. Zhao, G. H., Ceci, P., Ilari, A., Giangiacomo, L., Laue, T. M., Chiancone, E., and Chasteen, N. D. (2002) Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells - A ferritin-like DNA-binding protein of *Escherichia coli*, *J. Biol. Chem.* 277, 27689-27696.
64. Ilari, A., Ceci, P., Ferrari, D., Rossi, G. L., and Chiancone, E. (2002) Iron incorporation into *Escherichia coli* Dps gives rise to a ferritin-like microcrystalline core, *J. Biol. Chem.* 277, 37619-37623.
65. Chiancone, E., and Ceci, P. (2010) The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding, *Biochim Biophys Acta*.
66. Caspar, D. L. D., and Klug, A. (1962) Physical principles in construction of regular viruses, *Cold Spring Harbor Symposia on Quantitative Biology* 27, 1-24.
67. Zlotnick, A. (2005) Theoretical aspects of virus capsid assembly, *Journal of Molecular Recognition* 18, 479-490.
68. Katen, S., and Zlotnick, A. (2009) The thermodynamics of virus capsid assembly, in *Methods in Enzymology: Biothermodynamics, Vol 455, Part A*, pp 395-417.
69. Mannige, R. V., and Brooks, C. L. (2010) Periodic table of virus capsids: Implications for natural selection and design, *Plos One* 5.

70. Speir, J. A., Munshi, S., Baker, T. S., and Johnson, J. E. (1993) Preliminary-X-ray data-analysis of crystalline cowpea chlorotic mottle virus, *Virology* 193, 234-241.
71. Lin, T. W., Chen, Z. G., Usha, R., Stauffacher, C. V., Dai, J. B., Schmidt, T., and Johnson, J. E. (1999) The refined crystal structure of cowpea mosaic virus at 2.8 angstrom resolution, *Virology* 265, 20-34.
72. Kaper, J. M. (1975) The chemical basis of virus structure, dissociation and reassembly, *Front Biol* 39, 1-485.
73. Shipway, A. N., Katz, E., and Willner, I. (2000) Nanoparticle arrays on surfaces for electronic, optical, and sensor applications, *Chemphyschem* 1, 18-52.
74. Rosi, N. L., and Mirkin, C. A. (2005) Nanostructures in biodiagnostics, *Chemical Reviews* 105, 1547-1562.
75. Yamashita, I. (2008) Biosupramolecules for nano-devices: biomineralization of nanoparticles and their applications, *J. Mater. Chem.* 18, 3813-3820.
76. Dickerson, M. B., Sandhage, K. H., and Naik, R. R. (2008) Protein- and peptide-directed syntheses of inorganic materials, *Chemical Reviews* 108, 4935-4978.
77. Mann, S., and Meldrum, F. C. (1991) Controlled synthesis of inorganic materials using supramolecular assemblies, *Advanced Materials* 3, 316-318.
78. Tsukamoto, R., Iwahori, K., Muraoka, M., and Yamashita, I. (2005) Synthesis of Co₃O₄ nanoparticles using the cage-shaped protein, apoferritin, *Bulletin of the Chemical Society of Japan* 78, 2075-2081.
79. Tsukamoto, R., Muraoka, M., Fukushige, Y., Nakagawa, H., Kawaguchi, T., Nakatsuji, Y., and Yamashita, I. (2008) Improvement of Co₃O₄ nanoparticle synthesis in apoferritin cavity by outer surface PEGylation, *Bulletin of the Chemical Society of Japan* 81, 1669-1674.
80. Okuda, M., Kobayashi, Y., Suzuki, K., Sonoda, K., Kondoh, T., Wagawa, A., Kondo, A., and Yoshimura, H. (2005) Self-organized inorganic nanoparticle arrays on protein lattices, *Nano Letters* 5, 991-993.
81. Mackle, P., Charnock, J. M., Garner, C. D., Meldrum, F. C., and Mann, S. (1993) Characterization of the manganese core of reconstituted ferritin by X-ray-absorption spectroscopy, *Journal of the American Chemical Society* 115, 8471-8472.
82. Meldrum, F. C., Douglas, T., Levi, S., Arosio, P., and Mann, S. (1995) Reconstitution of manganese oxide cores in horse spleen and recombinant ferritins, *J. Inorg. Biochem.* 58, 59-68.
83. Douglas, T., and Stark, V. T. (2000) Nanophase cobalt oxyhydroxide mineral synthesized within the protein cage of ferritin, *Inorganic Chemistry* 39, 1828-1830.
84. Okuda, M., Iwahori, K., Yamashita, I., and Yoshimura, H. (2003) Fabrication of nickel and chromium nanoparticles using the protein cage of apoferritin, *Biotechnology and Bioengineering* 84, 187-194.
85. Klem, M. T., Mosolf, J., Young, M., and Douglas, T. (2008) Photochemical mineralization of europium, titanium, and iron oxyhydroxide nanoparticles in the ferritin protein cage, *Inorganic Chemistry* 47, 2237-2239.
86. Li, M., Viravaidya, C., and Mann, S. (2007) Polymer-mediated synthesis of ferritin-encapsulated inorganic nanoparticles, *Small* 3, 1477-1481.

87. Wong, K. K. W., and Mann, S. (1996) Biomimetic synthesis of cadmium sulfide-ferritin nanocomposites, *Advanced Materials* 8, 928-932.
88. Yamashita, I., Hayashi, J., and Hara, M. (2004) Bio-template synthesis of uniform CdSe nanoparticles using cage-shaped protein, apoferritin, *Chem. Lett.* 33, 1158-1159.
89. Iwahori, K., Yoshizawa, K., Muraoka, M., and Yamashita, I. (2005) Fabrication of ZnSe nanoparticles in the apoferritin cavity by designing a slow chemical reaction system, *Inorganic Chemistry* 44, 6393-6400.
90. Iwahori, K., Morioka, T., and Yamashita, I. (2006) The optimization of CdSe nanoparticles synthesis in the apoferritin cavity, *Physica Status Solidi a-Applications and Materials Science* 203, 2658-2661.
91. Hennequin, B., Turyanska, L., Ben, T., Beltran, A. M., Molina, S. I., Li, M., Mann, S., Patane, A., and Thomas, N. R. (2008) Aqueous near-infrared fluorescent composites based on apoferritin-encapsulated PbS quantum dots, *Advanced Materials* 20, 3592-3596.
92. Xing, R. M., Wang, X. Y., Yan, L. L., Zhang, C. L., Yang, Z., Wang, X. H., and Guo, Z. J. (2009) Fabrication of water soluble and biocompatible CdSe nanoparticles in apoferritin with the aid of EDTA, *Dalton Trans.*, 1710-1713.
93. Warne, B., Kasyutich, O. I., Mayes, E. L., Wiggins, J. A. L., and Wong, K. K. W. (2000) Self assembled nanoparticulate Co : Pt for data storage applications, *Ieee Transactions on Magnetism* 36, 3009-3011.
94. Ensign, D., Young, M., and Douglas, T. (2004) Photocatalytic synthesis of copper colloids from Cu(II) by the ferrihydrite core of ferritin, *Inorganic Chemistry* 43, 3441-3446.
95. Kramer, R. M., Li, C., Carter, D. C., Stone, M. O., and Naik, R. R. (2004) Engineered protein cages for nanomaterial synthesis, *Journal of the American Chemical Society* 126, 13282-13286.
96. Ueno, T., Suzuki, M., Goto, T., Matsumoto, T., Nagayama, K., and Watanabe, Y. (2004) Size-selective olefin hydrogenation by a Pd nanocluster provided in an apo-ferritin cage, *Angewandte Chemie-International Edition* 43, 2527-2530.
97. Galvez, N., Sanchez, P., and Dominguez-Vera, J. M. (2005) Preparation of Cu and CuFe Prussian Blue derivative nanoparticles using the apoferritin cavity as nanoreactor, *Dalton Trans.*, 2492-2494.
98. Klem, M. T., Willits, D., Solis, D. J., Belcher, A. M., Young, M., and Douglas, T. (2005) Bio-inspired synthesis of protein-encapsulated CoPt nanoparticles, *Advanced Functional Materials* 15, 1489-1494.
99. Galvez, N., Sanchez, P., Dominguez-Vera, J. M., Soriano-Portillo, A., Clemente-Leon, M., and Coronado, E. (2006) Apoferritin-encapsulated Ni and Co superparamagnetic nanoparticles, *Journal of Materials Chemistry* 16, 2757-2761.
100. Clemente-Leon, M., Coronado, E., Soriano-Portillo, A., Galvez, N., and Dominguez-Vera, J. M. (2007) Permanent magnetism in apoferritin-encapsulated Pd nanoparticles, *Journal of Materials Chemistry* 17, 49-51.
101. Dominguez-Vera, J. M., Galvez, N., Sanchez, P., Mota, A. J., Trasobares, S., Hernandez, J. C., and Calvino, J. J. (2007) Size-controlled water-soluble Ag nanoparticles, *Eur. J. Inorg. Chem.*, 4823-4826.

102. Butts, C. A., Swift, J., Kang, S. G., Di Costanzo, L., Christianson, D. W., Saven, J. G., and Dmochowski, I. J. (2008) Directing noble metal ion chemistry within a designed ferritin protein, *Biochemistry* 47, 12729-12739.
103. Ceolin, M., Galvez, N., and Dominguez-Verab, J. M. (2008) Thermal induced phase transitions and structural relaxation in apoferritin encapsulated copper nanoparticles, *Physical Chemistry Chemical Physics* 10, 4327-4332.
104. Ceolin, M., Galvez, N., Sanchez, P., Fernandez, B., and Dominguez-Vera, J. M. (2008) Structural aspects of the growth mechanism of copper nanoparticles inside apoferritin, *Eur. J. Inorg. Chem.*, 795-801.
105. Galvez, N., Fernandez, B., Valero, E., Sanchez, P., Cuesta, R., and Dominguez-Vera, J. M. (2008) Apoferritin as a nanoreactor for preparing metallic nanoparticles, *C. R. Chim.* 11, 1207-1212.
106. Kasyutich, O., Ilari, A., Fiorillo, A., Tatchev, D., Hoell, A., and Ceci, P. (2010) Silver ion incorporation and nanoparticle formation inside the cavity of *Pyrococcus furiosus* ferritin: Structural and size-distribution analyses, *Journal of the American Chemical Society* 132, 3621-3627.
107. Pead, S., Durrant, E., Webb, B., Larsen, C., Heaton, D., Johnson, J., and Watt, G. D. (1995) Metal-ion binding to apo, holo, and reconstituted horse spleen ferritin, *J. Inorg. Biochem.* 59, 15-27.
108. Naik, R. R., Stringer, S. J., Agarwal, G., Jones, S. E., and Stone, M. O. (2002) Biomimetic synthesis and patterning of silver nanoparticles, *Nature Materials* 1, 169-172.
109. Yoshimura, H. (2006) Protein-assisted nanoparticle synthesis, *Colloids and Surfaces a-Physicochemical and Engineering Aspects* 282, 464-470.
110. Allen, M., Willits, D., Young, M., and Douglas, T. (2003) Constrained synthesis of cobalt oxide nanomaterials in the 12-subunit protein cage from *Listeria innocua*, *Inorganic Chemistry* 42, 6300-6305.
111. Resnick, D. A., Gilmore, K., Idzerda, Y. U., Klem, M. T., Allen, M., Douglas, T., Arenholz, E., and Young, M. (2006) Magnetic properties of Co₃O₄ nanoparticles mineralized in *Listeria innocua* Dps, *Journal of Applied Physics* 99.
112. Iwahori, K., Enomoto, T., Furusho, H., Miura, A., Nishio, K., Mishima, Y., and Yamashita, I. (2007) Cadmium sulfide nanoparticle synthesis in Dps protein from *Listeria innocua*, *Chemistry of Materials* 19, 3105-3111.
113. Douglas, T., and Young, M. (1998) Host-guest encapsulation of materials by assembled virus protein cages, *Nature* 393, 152-155.
114. Douglas, T., Strable, E., Willits, D., Aitouchen, A., Libera, M., and Young, M. (2002) Protein engineering of a viral cage for constrained nanomaterials synthesis, *Advanced Materials* 14, 415-418.

CHAPTER 2 Investigation into the structural energetic consequences of helix swapping between a ferritin and a ferritin-like protein^{[1]*}

2.1 Introduction

Directly or indirectly, investigations into the structure of biomacromolecules have resulted in ~25% of the chemistry Nobel Prize in the past 50 years.^[2] Despite these accomplishments and this level of enquiry, a complete understanding of protein folding remains an unsolved problem and one of the “big unanswered scientific questions”.^[3] The field has advanced enough that now it is possible for scientists to design proteins *de novo*.^[4] However, the majority of the resulting designed proteins have been small and monomeric. Surprisingly, considering that complex biostructural information is often built up by the spontaneous assembly of subunits, the rational design of multimeric protein structures has only recently begun to be studied.^[5-8] In addition, the fundamentals of self-assembly have been explored on the mesoscale^[9] and as a pillar of supramolecular chemistry.^[10]

In nature, there are many examples of proteins that consist of multiple subunits and self-assemble into well-defined, cage-like structure.^[11] These protein cages have been the focus of much recent attention as part of bioorthogonal methodology development,^[12] as part of drug delivery studies,^[13-14] and as platforms for nanostructured materials.^[15-16] To explore and extend applications of protein cages, a complete understanding of the rules which control the association

* Parts of this chapter have been published previously with the author of this thesis as the first author. *Biochemistry*, **2009**, 48, 5623.

of subunits in protein cages has become essential. Previous analysis of the contact between protein subunit interfaces reveals that symmetry is important in defining the multiplicity of protein cages.^[7]

2.1.1 The ferritins – model protein cages

The ferritins make up a superfamily of well studied protein cages that are involved in iron storage and are ubiquitously found in both prokaryotes and eukaryotes.^[17-18] This wide distribution in nearly all organisms reflects the essential role of sequestering iron in the cells.^[19-20] The hollow protein cages of the ferritins store excess cytosolic iron in a mineralized form as insoluble hydrous ferric oxide. Due to the universal importance of the ferritins, they have been extensively studied, and many high-resolution structures of ferritins have been determined. This structural data, the cellular importance of these proteins, and the amenability of the ferritins to manipulation with standard molecular biological techniques, make this superfamily an ideal model system for developing the fundamentals of how to expand the dimensionality of rational protein engineering.

2.1.1.1 The structure of bacterioferritin

Ferritins that are isolated from bacteria are called bacterioferritins and most of bacterioferritins contain heme groups. Bacterioferritins have been successfully isolated from diverse bacteria strains such as *Rhodobacter capsulatus*,^[21] *Desulfovibrio desulfuricans*,^[22-23] *Mycobacterium smegmatis*,^[24] and *Escherichia coli*.^[25-27] It has been found that the main function of bacterioferritin is to detoxify iron in cells which resembles that of ferritin. However, the function of heme groups in bacterioferritins still remains unclear.^[28]

Many crystal structures of bacterioferritins have been established, and the most extensively studied bacterioferritin is from *Escherichia coli*. Bacterioferritin

from *Escherichia coli* (BFR) is composed of twenty-four identical subunits which self-assemble into a slightly distorted spherical cage with 432 octahedral symmetry.^[27] The ferritin cage has an exterior diameter of ~12 nm and an interior cavity of ~7.5 nm. Each monomer of BFR folds into a left-handed four-helix bundle which is composed of the A (Thr5-Trp35), B (Lys38-Phe64), C (Val83-Val111) and D (Try114-Met144) helices. Adjacent helix pairs AB and CD have an antiparallel relationship and are linked by a long random coil loop (Leu65-Asp82) between the B and C helices which spans the length of the bundle (Figure 2. 1). Each BFR subunit possesses an additional short fifth helix (i.e., the E helix, Leu146-Gln151) at the C-terminus, which lies at an angle of 78° with respect to the bundle axis. When the subunits assemble into a protein cage, the E helix points toward the core and the A and C helices are positioned on the outer surface, whereas the B and D helices face the inner cavity. In addition, the BC loop is primarily on the outside surface.

Figure 2. 1 (left) Ribbon diagram representation of bacterioferritin from *Escherichia coli* (Protein Data Bank ID is 1bfr) and (right) the structure of one monomer, which consists of the A (purple), B (cyan), C (yellow), D (orange) and E (red) helices and the BC loop (green).

2.1.1.2 The structure of DNA-binding protein from starved cells

DNA-binding proteins from starved cells are ferritin-like proteins which belong to the ferritin superfamily and are expressed in many bacteria under stress conditions.^[29] They have been shown to play important roles in protecting DNA

from oxidative damage.^[30-31] For the DNA-binding protein from starved *Escherichia coli* (DPS), twelve subunits self-assemble into a 32 symmetric tetrahedral, ferritin-like protein cage with an exterior diameter of ~9 nm and an interior cavity diameter of ~4.5 nm.^[30] The subunits of DPS, like those of BFR, fold into four-helix bundles. Similarly to BFR in the DPS monomer, the A (Asp23-Trp52), B (Phe59-Gln86), C (Val114-Glu138) and D (Asp142-Asn165) helices form two antiparallel helix pairs AB and CD, which are linked by short loops. Unlike BFR, the loop between the B and C helices that stretches along the length of the bundle possesses secondary structure as expressed in a short helix (i.e., the BC helix, Thr95-Lys101). The BC helix runs nearly orthogonal to the bundle axis and is exposed on the outside of the assembled protein cage (Figure 2. 2).^[32-37]

Figure 2. 2 (left) Ribbon diagram representation of DNA-binding protein from starved *Escherichia coli* (Protein Data Bank ID is 1dps) and (right) its monomer structure, which consists of the A (purple), B (cyan), C (yellow), D (orange) and BC (green) helices.

2.1.2 Structural comparison between BFR and DPS

Both BFR and DPS are members of the ferritin superfamily and structural studies have revealed that the monomers of BFR and DPS fold into similar four-helix bundle tertiary structures. However, BFR forms a cage with octahedral symmetry made up of twenty-four identical monomers and DPS forms a cage with tetrahedral symmetry composed of twelve identical monomers. Although comparative analyses have been proposed, the principles that control the

nanoarchitecture of these proteins as well as the role the extra helices play in protein assembly are still unknown. Grant and coworkers^[30] have compared the crystal structures of BFR and DPS by structural alignment and demonstrated that the helices of the bundle are positioned in a remarkably similar manner (RMSD = 1.1 Å) (Figure 2. 3). Moreover, of the amino acids that superimpose in the structure alignment, twenty-one are identical (Figure 2. 4).

The BFR monomers join at their E helices to define the four-fold axes of the octahedral protein cage, while in the DPS nanocapsule, the BC helix of one monomer interacts with the symmetry-related copy of another monomer's BC helix at the two-fold axes. In addition, the A and D helices of BFR are slightly longer at their C-termini than those of DPS. The N-terminus of DPS stretches beyond the four-helix bundle and thrusts from the three-fold channel toward the outside of the assembled cage. The positively charged N-termini of the DPS monomers encourage non-sequence specific interactions with DNA, and the assembled cages promote DNA condensation through the formation of large protein-DNA complexes.^[29, 38]

Figure 2. 3 Superimposition of BFR (grey) and DPS (black) monomers. The structural alignment is generated using DaliLite online program which can be found at <http://www.ebi.ac.uk/Tools/dalilite/index.html>.

```

BFR  .....MKGDTKVINYLNKLLGNELVAINQYFLHARMFKNWGLKRLNDVEYH
ident                               ||
DPS  SKATNLLYTRNDVSDSEKKATVELLNQVIQFIDLSLITKQAHWNMRGANFIAVHEMLDG

BFR  ESIDEMKHADRYIERILFLEGLPNLQ.....DLGKLNIG...EDVEEMLRSDLALELDG
ident  | | | | |
DPS  FRTALIDHLDTMAERAVQLGGVALGTTQVINSKTPKSYPLDIHNVQDHLKELADRYAIV

BFR  AKNLREAIGYADSVHDYVSRDMMIEILRDEEGHIDWLETELDLIQKMGLQNYLQAQIREEG
ident  | | | | |
DPS  ANDVRKAIGEAK...DDDTADILTAASRDLDKFLWFIECNIE.....

```

Figure 2. 4 Sequence comparison of the structural alignment of BFR and DPS. The grey shading indicates alpha helix regions and the identical residues are highlighted in red. The structural alignment is generated using DaliLite online program which can be found at <http://www.ebi.ac.uk/Tools/dalilite/index.html>.

Hydrophobic and hydrophilic interactions which stabilize the subunit interfaces to make up the quaternary structure of the BFR and DPS protein cages were analyzed to understand the assembly. (The summary of hydrophobic and hydrophilic interactions in BFR and DPS is listed in Table 2. 1 and Table 2. 2 respectively. More detailed analysis of all inter-subunit interfaces in BFR and DPS are presented in appendix 3 and 4 respectively.) Hydrophobic interactions usually involve a large number of residues buried within the interfaces and have the largest contribution to subunit-subunit association.^[39]

Dimer interfaces along the two-fold symmetry axes exist in both BFR and DPS, (Figure 2. 5a, d) and the interfaces are formed mainly by interactions between the A and B helices and the BC loop. Additionally, in DPS, the short BC helix is positioned in the middle of this interface and two BC helices have an antiparallel relationship. The surface area buried upon dimerization of two BFR monomers is $836 \text{ \AA}^2/\text{monomer}$ ^[40] whereas this value is nearly doubled in DPS ($1539 \text{ \AA}^2/\text{monomer}$).^[32-33] The additional surface burial could be a consequence of the interactions between the two BC helices. The BC loop in BFR is employed, albeit to a lesser extent, at the dimer interface; there are hydrogen bonds between

the backbone near Gln72 of one loop and the backbone near Leu77 and Gly75 of the other monomer's BC loop (Figure 2A. 1).^[27] Near the cavity surface, salt bridges are formed between Arg30, Asp56 and Glu60 (Figure 2A. 2).^[27] In DPS, the dimer interface is further stabilized by salt bridges formed between residues Lys48 on the A helix or Arg70 on the B helix of one subunit and Asp78 on the B helix of two-fold symmetry-related monomer (Figure 2A. 5).^[41]

DPS has two different types of three-fold axes. One has the similar packing manner to the three-fold axis in BFR (Figure 2. 5b, e). This axis is therefore called the ferritin-like three-fold axis. Interestingly, the majority of the amino acids along these axes are hydrophilic in both BFR and DPS. Along the three-fold axis in BFR, monomers are stabilized by an inter-subunit salt bridge between Arg61 and Glu128 (Figure 2A. 3).^[27] Moreover, the residue Glu128 also contributes an intra-subunit salt bridge with Arg102 (Figure 2A. 4).^[27] In DPS, the ferritin-like three-fold axis is stabilized by three salt bridges.^[33, 41] One of them is between Arg18 on the loop of the N-terminus and Asp123 in the C helix (Figure 2A. 6), the second is formed between Asp20 and Arg133 (Figure 2A. 7), and the third is formed between Arg83 and Asp156 at the ends of the B and D helices respectively (Figure 2A. 8). The ferritin-like interfaces in both BFR and DPS are not only stabilized by hydrophilic interactions, but also exhibit hydrophobic interactions. The buried surface area involves mostly the CD loop, the N-terminus of the D helix and the C-terminus of the B helix. The degree of burial is comparable for both proteins and is 1025 and 1375 Å²/monomer for BFR and DPS respectively.^[32-33, 40]

The four-fold axis in octahedral BFR is governed by the C-terminal E helix (Figure 2. 5c). However this symmetry related interface does not exist in tetrahedral DPS. The C-terminus of DPS along with parts of the A helix and the

BC loop form a second type of three-fold axis (Figure 2. 5f) which is referred to as the DPS-like three-fold axis. The buried surface area at this interface is 793 Å²/monomer which is significantly less than that at the dimer and ferritin-like three-fold interfaces.^[32-33] In BFR, the buried surface area is 1170 Å²/monomer^[40] at the four-fold interface. The residues around four-fold channels in BFR are hydrophilic, whereas the DPS-like three-fold channel is lined by hydrophobic residues.

Figure 2. 5 Comparison of the BFR and DPS protein cages. Views of BFR along a) two-fold axis b) three-fold axis c) four-fold axis. Views of DPS along d) two-fold axis e) ferritin-like three-fold axis f) DPS-like three-fold axis.

Table 2. 1 Buried surface area (Å²) per monomer in the interfaces of BFR and DPS. The data was taken from Vijayan *et al.*^[32-33, 40]

BFR	Dimer interface	Three-fold interface	Four-fold interface
Buried surface area, (Å ²)	836	1025	1170
DPS	Dimer interface	Ferritin-like three-fold interface	Dps-like three-fold interface
Buried surface area, (Å ²)	1539	1375	793

Table 2. 2 Hydrophilic interactions stabilizing the interfaces in BFR and DPS.^[33, 41] See appendix 2 for structural depictions of these interactions.

BFR	Dimer interface	Three-fold interface	Four-fold interface
	Gln72–Gly75–Leu77	Arg61–Glu128	
	Arg30–Asp56	Arg102–Glu128	
	Arg30–Glu60		
DPS	Dimer interface	Ferritin-like three-fold interface	Dps-like three-fold interface
	Lys48–Asp78	Arg18–Asp123	
	Arg70–Asp78	Arg83–Asp156	
		Arg133–Asp20	

2.1.3 Previous studies to understand ferritin self-assembly

Some limited mutagenesis work on ferritins has been performed with the goal of understanding their assembly. Andrews and coworkers engineered the C-terminus of BFR with a 14-residue extension.^[42] Gel filtration studies demonstrated that even though the wild type is a mixture of 24-mer and dimer, the mutant solely exists as a 24-mer, suggesting that the C-terminus and perhaps the E helix (which is located at the C-terminus and defines the four-fold axis), plays a role in controlling the oligomerization state in BFR.

Other researchers engineered human H-chain ferritin, which forms an octahedral 24-mer in analogy to BFR, by deleting the last twenty-two amino acids at the C-terminus or removing the first thirteen residues at the N-terminus.^[43–44] The two mutants showed similar motilities on native gel electrophoresis to that of the native ferritin, and they maintained the ability to catalyze iron oxidation. These data indicated that the mutants assembled into native-like protein cages and perhaps the interactions along the four-fold axes in mammalian ferritin were different from those in bacterioferritin as the E helix does not make a significant contribution to the energetics of human H-chain ferritin self-assembly.

A study conducted by Santambrogio and coworkers also involving human H-chain ferritin.^[45] They utilized a mutagenesis strategy to identify key residues at

the symmetry-related interfaces. While only a small number of mutants were investigated, many of these mutations drastically altered the oligomerization state and stability of the proteins. Analysis of renaturation which monitored by native PAGE resulted in the conclusion that mutants focused on the two-fold axis had a more pronounced effect on 24-mer oligomerization than that along the four- and three-fold axes. This conclusion reinforces to the proposed ferritin assembly mechanism, for this system at least, which is initiated by dimerization followed by the step-wise aggregation of additional dimers to form the 24-mer.

Recently, researchers have attached a KPAAALEHHHHH tag onto the C-terminal of *M. Smegmatis* DPS.^[33] The C-terminal tag was responsible for the DNA interaction and did not affect protein association. The crystal structure (2.8Å resolution) of the mutant revealed that the last twenty-two residues with the C-terminal tag were unstructured and did not participate in subunit-subunit interactions. Moreover, it was possible to disassociate the protein cage from 12-mer to dimer, then to the monomer by lowering the pH. This further supported the hypothesis that the dimer might be the early building blocks in the formation of ferritin-like protein cages.

2.1.4 Aims and objectives

As the E and BC helices are important for controlling the size and symmetry of ferritin assembly, in this chapter, we sought to clarify their roles in the assembly and stability of BFR and DPS by swapping these domains between their parent proteins. Although they are short helices (2 - 2.5 helical turns), we postulated that this swapping strategy would result in significant changes in stability and assembled structure since they are the major tertiary structural differences between BFR and DPS. Additionally, because the E helix is associated

with the four-fold axis in BFR and the BC helix of DPS with the two-fold axis, switching these elements would also allow us to probe the importance of the symmetry-related interactions. Moreover, as we are not introducing any mutations within the four-helix bundles themselves (with the exception of some minor modifications within the D helix to better align the shorter DPS with BFR), it is thought that this experimental design would least likely disrupt the tertiary structure while providing an avenue for exploring assembly as independently as possible in a cooperatively folding multiple-subunit protein. We therefore designed six mutant proteins that either added (DPS+E, BFR+BC), replaced (DPS-BC+E, BFR+BC-E), or deleted (DPS-BC, BFR-E) these extra-bundle helices and set out to study their biophysical and self-assembly properties (Figure 2. 6).

Figure 2. 6 Top) Crystal structures of the DPS (PDB ID is 1dps) and BFR (PDB ID is 1bfr) protein nanocages with highlighted monomers, associated four-helix bundle monomers that make them up (not to scale) viewed down the helix bundle axis, and schematics indicating the A-D helices associated with the bundle (circles) and the additional BC and E helices (rounded rectangles) (white=DPS; black=BFR). **(Bottom)** Schematic design for proteins with added or

deleted BC and E domains following the convention from above. This figure is taken from Fan *et al.*^[1]

2.2 Experimental

2.2.1 Total gene synthesis of BFR and DPS

2.2.1.1 Codons optimization and oligonucleotides (oligos) design

The protein sequences of bacterioferritin from *E. coli* (BFR, PDB ID is 1bfr) and DNA binding protein from starved *E. coli* (DPS, PDB ID is 1dps) are obtained from the Protein Data Bank (PDB). The protein sequences of wild type and mutant (Note: the details of protein design are discussed in the results and discussion.) were converted into gene sequences by utilizing optimal codons for each amino acid to maximize expression level in *E. coli* and codons that collectively preserved an overall gene GC content of genes within 40 to 60%.^[46] The optimized codons are listed as below (Table 2. 3)^[46]. Most of the target gene sequences were divided into 27 nucleotide (nt) oligos, which overlap with adjacent oligos by 15 nt at their 5' end and 12 nt at their 3' end. However, a small number of the oligos were in different length due to the variable mutant gene sequences (NOTE: all of the oligo sequences are listed in appendix 1). Oligos were synthesized using solid-phase phosphoramidite technology by 1st Base Pte. Ltd. (Singapore). The initial concentration used in molecular biological reactions is 100 μ M before dilution unless otherwise stated.

Table 2. 3 Optimal codons used to translate protein sequences to gene sequences. The table is adapted from Dong *et al.*^[46]

codon	a.a. *	codon	a.a. *	codon	a.a. *	codon	a.a. *
AAA	Lys	ATG	Met	CTG	Leu	GTG	Val
AAC	Asn	CAG	Gln	GAA	Glu	TAC	Tyr
ACC	Thr	CAT	His	GAT	Asp	TGG	Trp
AGC	Ser	CCG	Pro	GCA	Ala	TGC	Cys
ATC	Ile	CGT	Arg	GGT	Gly	TTC	Phe

* a.a. is the abbreviation for amino acids.

2.2.1.2 Assembly, amplification and extension PCR (Figure 2. 7)

All the genes were synthesized by a short oligo assembly approach described by Stemmer *et al* (Figure 2. 7).^[47] The total gene synthesis reactions were performed by adding the oligos (0.25 μ L of each, 10 μ M) to a solution made up of a dNTP solution (Fermentas, 2.5 μ L, 2 mM of each dATP, dCTP, dTTP, dGTP), Pfu buffer (Fermentas, 2.5 μ L, 10 $\times$), Pfu DNA polymerase (Fermentas, 0.5 μ L, 2.5 U/ μ L,) and deionized water to bring the total volume to 25 μ L. The gene was constructed in a programmable thermal cycler (TaKaRa, Japan). The program consisted of thirty cycles made up of a denaturation step (94 $^{\circ}$ C, 0.5 min), an annealing step (55 $^{\circ}$ C, 0.5 min) and an extension step (72 $^{\circ}$ C, 1 min). These thirty cycles were followed by a final extension step (72 $^{\circ}$ C, 5 min).

The assembled gene was amplified by PCR. The unpurified assembled template (1 μ L) was added to a solution of forward and reverse primers (these being the end primers from the assembly PCR, 2 μ L of each, 10 μ M), dNTP solution (2.5 μ L, 2 mM of each dATP, dCTP, dTTP, dGTP), Pfu DNA polymerase (0.5 μ L, 2.5 U/ μ L), Pfu buffer (2.5 μ L, 10 $\times$) and deionized water (14.5 μ L) to bring the total volume to 25 μ L. The amplification program consisted of an initial denaturation step (94 $^{\circ}$ C, 5 min) followed by thirty cycles consisting of a denaturation step (94 $^{\circ}$ C, 0.5 min), an annealing step (55 $^{\circ}$ C, 0.5 min) and an extension step (72 $^{\circ}$ C, 1 min). These thirty cycles were followed by a final extension step (72 $^{\circ}$ C, 5 min).

The DNA template (1 μ L) was added to a solution of sense and antisense of Ligation Independent Cloning (LIC) primers (2 μ L of each, 10 μ M) dNTP solution (2.5 μ L, 2 mM of each dATP, dCTP, dTTP, dGTP), Pfu buffer (2.5 μ L, 10 $\times$), Pfu DNA polymerase (0.5 μ L, 2.5 U/ μ L) and deionized water (14.5 μ L) to bring the

total volume to 25 μ L. The PCR program consisted of an initial denaturation step (94 $^{\circ}$ C, 5 min) followed by thirty cycles made up of a denaturation step (94 $^{\circ}$ C, 0.5 min), an annealing step (55 $^{\circ}$ C, 0.5 min) and an extension step (72 $^{\circ}$ C, 1 min). These thirty cycles were followed by a final extension step (55 $^{\circ}$ C, 5 min). The product was isolated by gel purification (Eppendorf Perfectprep Gel Cleanup kit).

The products from the assembly, amplification and extension PCRs (2 μ L) were mixed with loading dye (1 μ L, Fermentas, consist of 10 mM Tris pH 7.6, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, and 60 mM EDTA) before loading on a 1% agarose gel which was prepared by dissolving agarose in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA). The gel was run in TAE buffer at 120 V for 30 min.

Figure 2. 7 Schematic of assembly, amplification and extension PCR. This figure is taken from Fan *et al.*^[1]

2.2.2 Preparation of extended single stranded complementary ends and annealing to pET-32 Ek/LIC vector (Figure 2. 8)

The purified DNA product from the extension PCR (8 μ L) was added to a solution consisting of dATP solution (Novagen, 2 μ L, 25 mM), DTT (Novagen, 1 μ L, 100 mM), T4 DNA polymerase buffer (Novagen, 2 μ L, 10x), T4 DNA polymerase (Novagen, 0.4 μ L, 2.5U/ μ L) and deionized water (6.6 μ L) to bring the

total volume to 20 μ L. The sample was incubated (22 $^{\circ}$ C, 30 min) after which the temperature was increased (75 $^{\circ}$ C, 30 min) to inactivate the enzyme.

The gene insert (2 μ L) was then added to a solution of deionized water (3.5 μ L) containing the pET-32 Ek/LIC vector (Novagen, 0.5 μ L). The solution was incubated (22 $^{\circ}$ C, 5 min), and EDTA (Novagen Chemicals, 2 μ L, 25 mM) was then added and the solution was incubated for an additional hour.

Figure 2. 8 Schematic of the cloning strategy. This figure is taken from Fan *et al.*^[1]

2.2.3 Colony screening by PCR and sequencing

The resulting vector construct was transformed by electroporation into electrocompetent Novablue cells (Novagen). The transformants were plated on LB plates (containing 50 μ g/mL carbenicillin, 2.5 μ g/mL tetracycline) and were grown (37 $^{\circ}$ C) overnight.

Each resulting carbenicillin-resistant colony (total 10-20 colonies) was picked up and transferred into both sterile water (50 μ L) and LB (3 mL, containing 50 μ g/mL carbenicillin, 12.5 μ g/mL tetracycline). Colonies in water were dispersed by vortexing and boiled (99 $^{\circ}$ C, 5 min) to lyse the bacteria. Cell debris was removed by centrifugation (12000 rpm, 1 min) and the supernatant (5 μ L) was used as a template for colony PCR. The template was added to a solution of

forward and reverse primers (these being the end primers from gene synthesis), (2 μ L of each, 10 μ M), dNTP solution (2.5 μ L, 2 mM of each dATP, dCTP, dTTP, dGTP), Pfu DNA polymerase (0.5 μ L, 2.5 U/ μ L), Pfu buffer (2.5 μ L, 10 $\times$) and deionized water (14.5 μ L) to bring the total volume to 25 μ L. The amplification program consisted of an initial denaturation step (94 $^{\circ}$ C, 5 min) followed by thirty cycles consisting of a denaturation step (94 $^{\circ}$ C, 0.5 min), an annealing step (55 $^{\circ}$ C, 0.5 min) and an extension step (72 $^{\circ}$ C, 1 min). These thirty cycles were followed by a final extension step (72 $^{\circ}$ C, 5 min). Colony PCR products (2 μ L) were loaded on a 1% agarose gel and run in TAE buffer at 120 V for 30 min. Positive colonies which had PCR products of the correct size were grown in LB (37 $^{\circ}$ C) overnight. The recombinant constructs were extracted from bacteria culture by miniprep (Qiagen) and confirmed by dideoxy sequencing (1st Base).

2.2.4 Protein expression and extraction

The positive recombinant gene constructs were subsequently transformed into BL21(DE3) by electroporation. The transformants were plated on LB plates (containing 50 μ g/mL carbenicillin) and incubated (37 $^{\circ}$ C) overnight. Single colonies were inoculated and grown in LB (3 mL, 50 μ g/mL carbenicillin) overnight. The culture (3 mL) of BL21(DE3) (Novagen) cells containing the positive expression construct was added to LB (250 mL) supplemented with carbenicillin (50 μ g/mL). The culture was incubated (150 rpm, 37 $^{\circ}$ C) until OD₆₀₀ ~ 0.4-0.6 upon which protein expression was induced (IPTG, Fermentas, 1 mL, 100 mM) and the culture was incubated further (150 rpm, 30 $^{\circ}$ C, 3 h).

The cells were harvested by centrifugation (12000 rpm, 4 $^{\circ}$ C, 15 min) and the pellet, re-suspended in equilibrium buffer (300 mM NaCl, 50 mM NaH₂PO₄ (USB Chemicals), pH 7.0, 7.5 mL) was lysed by sonication (SONICS Vibra Cell

sonicator, cycle of 3 s pulse at 40% amplitude followed by a 3 s rest repeated for a 4 min period). The lysis debris was cleared by centrifugation (12000 rpm, 4 °C, 15 min) and the supernatant was filtered through a 0.2 µm membrane.

2.2.5 Protein purification under native conditions (Figure 2. 9)

The his-tagged proteins were purified by incubation of the filtered protein solution with metal affinity resin (TALON, Clontech) in batch mode as per the manufacturer's instructions. The proteins were eluted from the resin using affinity tag cleavage by the addition of enterokinase (New England Biolabs, 8 µL, 2 µg/mL) and the solution was incubated with gentle inversion (4 °C, 48 h). The protein was then concentrated and the buffer was exchanged using ultrafiltration (Centricon). A 12% SDS-PAGE gel was run to verify the purity of the proteins.

2.2.6 Protein purifications under denaturing conditions (Figure 2. 9)

After induction and cell isolation, the bacterial pellet was resuspended in lysis buffer (100 mM NaH₂PO₄, 10 mM Tris, 8 M urea, pH 8.0, 4 mL). The solution was centrifuged (10000 rpm, 4 °C, 20 min), and the pellet discarded. Resin (1 mL) was added to the supernatant and incubated with gentle inversion (1 h). The supernatant was then discarded and the resin washed with buffer (100 mM NaH₂PO₄, 10 mM Tris, 8 M urea, pH 6.3 (for BFR+BC-E), to elute any nonspecifically bound protein. The resin was then washed with buffers containing decreasing amounts of urea (6 M, 4 M, 2 M, 1 M, 0.5 M) to refold the protein. The buffer was then replaced with enterokinase buffer (20 mM Tris, 50 mM NaCl, 2 mM CaCl₂, pH 8.0, 2 mL), enterokinase (2 µL) was added, and the solution was incubated with gentle inversion (4 °C, 48 h).

Figure 2. 9 Schematic of protein purification under (left) native and (right) denaturing conditions. This figure is taken from Fan *et al.*^[1]

2.2.7 Temperature dependent circular dichroism (CD) analysis

Purified protein solution was diluted with phosphate buffer (50 mM NaH₂PO₄, pH 7.2) and then filtered (0.2 μm). The concentration of the protein solution was determined (BCA, Novagen), and it was diluted to 100 μg/mL.

CD analysis was performed in 1 mm quartz cells, using a Jasco J-710 spectropolarimeter fitted with a Peltier temperature controller. Spectra were recorded from 250-200 nm at 5 °C intervals over a temperature range of 20-90 °C with an equilibration time of 10 min at each temperature. At least three replicates were performed. For each experiment, the signal at 222 nm was plotted versus temperature and fitted to the equation: $CD = CD_{fold} + (CD_{unfold} - CD_{fold}) / (1 + (T/T_m)^B)$ where CD_{fold} is the signal of the fully folded protein, CD_{unfold} is the signal of the fully unfolded protein, T is temperature, T_m is the melting point, and B is a fitting

parameter which gives a sense of the cooperativity of the transition. After melting, the protein solutions were cooled slowly to 20 °C over 30 min and the resulting spectra were compared to those obtained before the unfolding experiment.

2.2.8 Dynamic light scattering (DLS)

The concentration of the protein was determined (BCA, Novagen), and the solution was diluted to 100 µg/mL with phosphate buffer. DLS analysis was performed using a Brookhaven 90 Plus Particle Sizer and a 1 cm path length quartz cell. The data set was collected for 3 min, and three data sets were collected for each protein. An average hydrodynamic diameter was calculated from at least three replicates.

2.2.9 Size exclusion chromatography (SEC)

Size exclusion was performed on a GE Akta FPLC system monitored at 280 nm using a Superdex 200 10/300 GL gel filtration column at a flow rate of 0.5 mL/min (running buffer: 50 mM phosphate, 0.15 M NaCl, pH 7.0). Calibration was performed by injection of blue dextran, thyroglobin, horse spleen ferritin, aldolase, conalbumin, and ovalbumin (GE Biosystems). The elution volumes (V_e , average of two runs) were used to determine the gel-phase distribution constant (K_{av}) by the following relation: $K_{av} = (V_e - V_o) / (V_c - V_o)$ where V_o is the void volume (assumed to be the elution volume of the blue dextran), and V_c is the column volume (24 mL). These K_{av} values were correlated to the Stokes radii of the proteins. Conalbumin was excluded from the Stokes radii correlation due to the fact that its radius is unknown. (GE Biosystems Calibration Kit Manual)

2.2.10 Transmission electron microscopy (TEM)

The solution of the purified protein (10 µL, 20-100 µg/mL) was dropped onto a sheet of parafilm and the grid (formvar/carbon, copper 300 mesh) was

placed on the drop (1 min) after which filter paper was used to wick away excess solution. The grid was then placed on a solution of uranyl acetate stain (7 μ L, 1% w/v, 1 min). The grid was dried with filter paper then air dried in a closed box.

TEM data were obtained at the National University of Singapore Medical School using a FEI EM 208S transmission electron microscope operating at 100 keV.

Particles sizes were measured from the micrographs using Image J (National Institutes of Health, Bethesda, Maryland, USA). An ellipse was drawn around the particle and the diameter of this ellipse was calculated. At least 50 particles were measured from each image, except in the cases where there were fewer than 50 particles in a particular image, in which case the diameter of every particle in the image was measured.

2.2.11 Analytical ultracentrifuge sedimentation equilibrium (SE)

Sedimentation equilibrium experiments were performed on a Beckman Optima XL-A analytical ultracentrifuge. Solutions of DPS and DPS+E were filtered (0.2 μ m) and diluted to 0.2 mg/ml with gel filtration running buffer (50 mM phosphate, 0.15 M NaCl, pH 7.0). Protein samples were centrifuged at 5000, 6500, 9000, 12000 rpm. The equilibrium data were collected at each speed for 4, 8, 12, 16, 20 h at 20 °C with 0.0001 cm step resolution, 20 averages/scan. “Ultrascan” software was used to calculate buffer density (1.0089 g/mL) and partial specific volume of proteins from amino acid composition (the partial specific volume of DPS and DPS+E are 0.73915 mL/g and 0.73907 mL/g respectively)

2.3 Results and discussion

2.3.1 Protein design

Inspection of the BFR and DPS four-helix bundle monomer structures reveals that the two primarily differ in the placement of an additional fifth helix: DPS has a BC helix between the second and third helices of the bundle and BFR has an E helix at its C-terminus. The BC helix is found at the interface between two dimers near the two-fold axis in DPS whereas the E helix is located at the tetramerization interface and is positioned near the four-fold symmetry axis in BFR. The goal of this study is to determine what role the E and BC helices play in BFR and DPS respectively by determining changes in monomer structure, thermodynamic stability, folding reversibility, size of the oligomer, and oligomerization state when these mini-helices are swapped between the two proteins. We therefore designed six mutant proteins that explore all possible permutations. The proteins that were studied were DPS-BC and BFR-E which have neither of the extra-bundle helices, DPS and BFR+BC-E which possess only the BC helix, DPS-BC+E and BFR which only incorporate the E helix, and DPS+E and BFR+BC each of which have both BC and E helices (Figure 2. 6). Protein sequences of the wild type proteins and mutants are listed below (Figure 2. 10).

Figure 2. 10 Amino acid sequences of proteins used in this study. Grey shading indicates each of the helices (A-D in order from the N- to C-termini) in the four-helix bundle from the parent protein. A black box indicates the E helix from BFR, and a white box indicates the BC helix from DPS. This figure is taken from Fan *et al.*^[1]

2.3.2 Verification of PCR products

Genes of all targets were synthesized by using PCR to assemble a number of short oligos which are partially complementary to each other. To increase the number of copies of the genes, amplification PCR was conducted with the two end primers from the assembly PCR (Figure 2. 7). The products of these two rounds of PCR were verified by agarose gel electrophoresis to determine success of gene synthesis (Figure 2. 11). The gel images indicated that the products of assembly PCR showed smeared bands which were ~100-200 base pairs (bp) and they had lower molecular weight compared to fully assembled target genes which is ~500

bp (Figure 2. 11). This might due to the low efficiency of producing entire target gene by one step assembly PCR. However, the low population of the target could be amplified using a PCR with the two end primers. The resulting distinct bands have molecular weight of approximately 500 bp which correspond to the correct length of the target genes (Figure 2. 11). In conclusion, the desired genes were successfully synthesized by PCR.

Figure 2. 11 DNA electrophoretic analysis of the products of gene assembly and amplification. (See Figure 2. 7 for a schematic of the gene synthesis strategy.) This figure is taken from Fan *et al.*^[1]

The amplification products were elongated by extension PCR using Ek/LIC primers which incorporated specific sequences at the ends of the target genes to

enable the ligation of the genes into the expression vector. The products of extension PCR were also analyzed by agarose gel electrophoresis (Figure 2. 12), and the resulting ~500 bp bands were excised from the gel and purified.

Figure 2. 12 DNA electrophoretic analysis of the products of extension PCR with LIC primers. The bands of appropriate size (~500 bp) were excised and purified. This figure is taken from Fan *et al.*^[1]

The purified extension PCR products were treated with T4 DNA polymerase to generate single stranded overhangs which were annealed with the complementary sticky ends on the vector by ligation independent cloning. The resulting constructs were electroporated into competent cells and the transformants were selected on plates containing carbenicillin. The antibiotic resistant colonies were picked and screened by colony PCR. The amplified colony PCR products were analyzed by agarose gel electrophoresis and bands corresponding to the expected size of the products (~500 bp) were observed (Figure 2. 13). To further confirm the inserts, the plasmids in the positive colonies were extracted by miniprep and sequenced.

Figure 2. 13 DNA electrophoretic analysis of the products of single-colony PCR screens for vectors containing the insert genes. This figure is taken from Fan *et al.*^[1]

In conclusion, the genes for expression were successfully generated through the assembly of synthetic oligos by means of a double PCR procedure. These genes were efficiently conjugated to a pET-32 Ek/LIC expression vector.

This cloning method is proven to be general and high yielding, and should be amenable to eventual higher throughput production.

2.3.3 Protein expression and purification

The plasmids with the correct insert sequences were electroporated into *E. coli* BL21(DE3) for protein over-expression. The advantages of using the pET-32 Ek/LIC expression vector are a) it uses ligation-independent cloning which is high yielding and therefore could prove useful in making libraries, b) it provides proteins fused to a series of affinity tags to simplify purification, c) these tags can be removed through cleavage at an enterokinase cleavage site, and d) this site abuts the natural start site of the protein, leaving no additional residues, an essential requirement for self-assembly studies.

The recombinant proteins were over-expressed in *E. coli* by inducing with addition of IPTG. The cells were lysed by sonication and recombinant proteins were released into the lysis buffer. The his-tagged recombinant protein was purified through immobilization on affinity resin followed by on-resin enterokinase cleavage and protein release (Figure 2. 9). The benefit of this strategy is that because the monomer is bound to the resin and only self-assembles upon release (the large affinity tags presumably would prevent self-assembly), it would conceivably be relatively simple to load for drug delivery studies. The SDS-PAGE analysis of each protein expression (Figure 2. 14) showed a large band of protein with molecular weight (MW) between 25 and 37 kDa after IPTG induction, which indicated that the recombinant protein with fusion tags was expressed in all cases. After lysis of the bacteria with sonication, most of the fusion proteins predominantly resided in the soluble fraction and the soluble lysate was applied to Ni affinity resin. After affinity resin isolation and enterokinase cleavage, pure

proteins bands of ~20 kDa (expected MW 18-19 kDa) were observed on SDS-PAGE gels. This procedure appeared to be general in that it worked well for all the proteins with the exception of BFR+BC-E, where after over-expression all the protein was taken up in insoluble inclusion bodies (Figure 2. 14 Top gel of BFR+BC-E). After unsuccessfully attempting to optimize the expression of BFR+BC-E, an on-resin refolding strategy was employed, where the inclusion bodies were solubilized in urea, applied to the resin, and the denatured protein was refolded by gradual buffer exchange into enterokinase buffer and cleaved as normal (Figure 2. 14 Bottom gel of BFR+BC-E).

Figure 2. 14 Expression and purification of ferritins and ferritin mutants. (from left to right, with the exception of BFR+BC-E) Lane 1: Protein ladder. Lane 2: Uninduced. Lane 3: Induced. Lane 4: Insoluble fraction. Lane 5: Soluble fraction. Lane 6: Flow through after applying soluble fraction to resin. Lane 7: Flow through from washing the resin. Lane 8: Purified protein after Ek cleavage of the affinity tags and release from the resin. Note: The bottom gel for BFR+BC-E was done after on-resin refolding. Lane 1: Ladder. Lane 2: Uninduced. Lane 3: Induced. Lane 4: Solubilized lysate. Lane 5: Flow through after applying solubilized lysate to resin. Lane 6: Flow through from washing the resin. Lane 7: Purified protein after Ek cleavage of the affinity tags and release from the resin. This figure is taken from Fan *et al.*^[1]

2.3.4 Circular dichroism-based studies

Circular dichroism (CD) is a spectroscopic method which has been largely used to characterize the overall secondary structure of peptides and proteins. Chiral samples (like proteins) absorb left and right circularly polarized light differently resulting in CD signals. The characteristic CD spectra of α -helices have been well studied and exhibit double minima at 208 nm and 222 nm indicating π - π^* and n - π^* electron transition respectively.^[48] Ellipticity (θ) is usually used for CD signals and is measured in millidegrees (mdeg). The signals are typically normalized for concentration and number of residues and defined as mean residue ellipticity θ_{mrw} ($\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$).

$$\theta_{mrw} = \frac{\theta \times M}{10 \times c \times l \times n}$$

Where M is the molecular weight of protein, n is the number of residues, c is the concentration of protein in mg/mL and l is the path length of the cell in cm.

BFR and DPS are composed of α -helical bundles. The circular dichroism (CD) spectra of all the proteins in phosphate buffer at 20 °C indicate that they all fold into highly helical secondary structures and exhibit two negative minima at 222 nm and 208 nm. The mutations did not strongly affect the overall helicities of the proteins (Figure 2. 16 top) suggesting that their secondary structure is intact. Interestingly the spectra of the mutants at 20 °C resemble those of the parent proteins (Figure 2. 16 top) suggesting that the structures of the mutants derived from BFR are more similar to BFR than they are to DPS and vice versa and that the tertiary structure of the mutants are intact.

To determine the role of the two mini-helices play in thermodynamic stability of the proteins, temperature denaturation monitored by the CD 222 nm band was conducted (Figure 2. 15). In all cases, the α -helicity decreased with

increasing temperature. As would be expected by making such gross, domain-based mutations without optimization for packing, the thermal stability of the mutants was depressed, although they all exhibit highly cooperative unfolding suggesting well packed folded initial structures (Figure 2. 16 bottom). The maxi-ferritin, BFR, was more sensitive to mutation than DPS. For the BFR series, either the deletion of the E helix (BFR-E) or the addition of the BC helix (BFR+BC) resulted in ~ 11 °C depression in T_m indicating that the E helix is essential for protein overall stability and that introduction of the BC helix disrupts the stability (Figure 2. 16 bottom, Table 2. 4). This trend is reinforced by the observation that removal of the E helix and inclusion of the BC helix (i.e., BFR+BC-E) has the largest affect on BFR stability, reducing the T_m by ~ 27 °C (Figure 2. 16 bottom, Table 2. 4). Taken together, these data demonstrate the importance of the interactions of the E helix across the four-fold axis and imply that the introduction of additional interactions across the two-fold symmetry axis is detrimental rather than advantageous as would be expected if dimerization has been highly optimized to drive the assembly. As this maybe because the BC helix and the BC loop in BFR have different, non-analogous packing roles as they are played in DPS. For the mutants derived from DPS, as we expected, the removal of the BC helix (DPS-BC) has the largest melting point depression (~ 16 °C) presumably due to the loss of essential contacts across the two-fold axis (Figure 2. 16 bottom, Table 2. 4). However, addition of the E helix (DPS-BC+E) nearly fully recovers the stability as this mutant has T_m that is depressed by only ~ 5 °C from that of DPS (Figure 2. 16 bottom, Table 2. 4). The protein with both helices (DPS+E) preserves the trend with a T_m depression of ~ 7 °C (Figure 2. 16 bottom, Table 2. 4). It is postulated that the E helix in DPS+E plays an essential role in making interactions at the

DPS-like three-fold axis and stabilizes the mutant (DPS-BC+E) which lacks the natural two-fold interface interactions.

Figure 2. 15 Temperature dependent circular dichroism (CD) spectra of the ferritins and ferritin mutants. CD spectra of the proteins at different temperatures for the BFR (left) and

the DPS series (right). Spectra were taken for each protein at temperatures from 20-90 °C at 5 °C intervals. These spectra were used to determine the loss of helical structure with temperature by following the signal at 222nm and fitting the resulting data to establish a melting point. These spectra are the average of at least three replicate melts. This figure is taken from Fan *et al.*^[1]

Figure 2. 16 (Top) Circular dichroism (CD) spectra of the cage proteins and mutants at 20 °C before thermal denaturation (solid lines) and at 20 °C after thermal denaturation (dotted lines) to indicate level of unfolding reversibility. (Bottom) The CD signal of the cage proteins and mutants at 222 nm during thermal denaturation. This figure is taken from Fan *et al.*^[1]

Table 2. 4 Melting temperatures of proteins used in temperature dependent CD study. ΔT_m is the change in melting temperature with respect to the parent protein. This table is taken from Fan *et al.*^[1]

Protein	T_m (°C)	ΔT_m (°C)
BFR	65.3±0.7	NA
BFR-E	53.9±0.2	11.4±0.9
BFR+BC-E	38.1±2.7	27.2±3.4
BFR+E	53.6±0.2	11.7±0.9
DPS	67.6±0.3	NA
DPS-BC	50.1±0.8	17.5±1.1
DPS-BC+E	63.4±1.4	4.2±1.7
DPS+E	60.3±0.4	7.3±0.7

Controlled refolding experiments result in an additional surprise. Although the DPS series is more thermally stable, slow cooling after denaturation indicated that DPS and its mutants were much less likely to return to their native fold than that of the BFR mutants (Figure 2. 16). These results imply that the DPS series either possesses a larger kinetic barrier to refolding or can refold along a non-helical pathway. Understanding the nature of this effect warrants further investigation. Taken together with the observation that BFR is more thermally sensitive to mutation, these data suggest that the forces governing the folding of the octahedral ferritin are possibly dissimilar to those that control the tetrahedral ferritin-like protein.

2.3.5 Transmission electron microscopy investigation

With the knowledge that all the proteins form cooperatively folded, thermodynamically stable, helical structures, we explored their ability to assemble into hollow cages by transmission electron microscopy (TEM) (Figure 2. 17). Single ferritins and ferritin-like protein cages can be easily visualized by TEM. Commonly, specimens are deposited on a thin carbon film supported by copper grid and the contrast surrounding the ferritin is achieved after negative staining which provides an outline image of the particles. Consistent with the literature,^[49] our proteins appeared as light rings surrounding dark cores. The light rings indicated the ferritin protein shell and the dark cores are evidence of the uptake of the negative stain in the hollow protein cavities (Figure 2. 17). The electron micrographs of the native proteins BFR and DPS revealed structures resembling the shape and size of those reported in the literature (Figure 2. 18). Of the mutants, only the DPS series formed any uniform structure visible by TEM. This is consistent with the CD data which demonstrated the greater sensitivity of BFR to

mutagenesis relative to that of DPS. Image analysis of the DPS-BC micrographs indicates that the size of the mutant is slightly smaller than that of DPS. Interestingly, DPS+E produced nano-structured particles of a size intermediate between that of BFR and DPS (Figure 2. 18). Although their outer diameters were of different size, the sizes of their inner cavities remain unchanged from that of the native DPS (Figure 2. 18). Many of the negatively stained proteins have polygonal appearance. This could be a result of 1) some of the randomly orientated ferritins exhibiting angular profiles when viewed in projection or 2) the thickness of the negative stain is uneven.^[49]

Figure 2. 17 Transmission electron micrographs of ferritin and ferritin mutant negatively stained with uranyl acetate. The scale bars are 50 nm. This figure is taken from Fan *et al.*^[1]

Figure 2. 18 Particle sizes of BFR and DPS mutants from the image analysis of TEM micrographs. This figure is taken from Fan *et al.*^[1]

2.3.6 Dynamic light scattering and size exclusion chromatography

To gain further insight into the nature of the nano-structures and their relative populations in solution, and to assess self-assembled state of the mutants that do not exhibit cage structures in the TEM, two solution-phase techniques were employed. Dynamic light scattering (DLS) revealed particles with diameters of ~13 and ~10 nm for BFR and DPS respectively (Figure 2. 20). These values are slightly larger than those determined by TEM, however DLS measures hydrodynamic radii and so the measured diameters would be expected to be larger than those obtained by TEM or AFM analysis due to bound solvent. The sizes (~1-3 nm) of all BFR mutants are much smaller than their parent implying that those mutants have lost their cage structure, which is consistent with the TEM data. The DPS series formed larger particles than the BFR mutants and ranged from ~6 to ~13 nm suggesting that some sort of oligomers were formed. Analysis of the mutants showed that DPS+E formed particles larger than the parent, whereas DPS-BC formed nanostructures that were smaller which is also consistent with the electron microscopy data.

Size exclusion chromatography (SEC) is a simple and nondestructive method which allows the exploration of both small and large oligomers, and differences between populations of small particles (such as monomer and dimer) can be clearly observed (Figure 2. 19). SEC therefore can sharpen our perception of the oligomerization state and therefore help to further refine the role the BC and E helices play in assembly of the protein cages by estimating the hydrated radii (stokes radii) through comparing the elution volume with those of protein standards of known diameter (Figure 2. 19). (Note: the calculation is specified in the experimental part.)

Figure 2. 19 Size exclusion chromatography calibration. Top: Chromatogram of calibration proteins. Bottom: Stokes radius (R_{st}) correlation to K_{av} as defined in the Experimental Section. Conalbumin was excluded from the Stokes radii correlation due to the fact that its radius is unknown. This figure is taken from Fan *et al.*^[1]

Size exclusion experiments, possibly because of the associated solvation sphere surrounding these proteins, suggested that BFR and DPS (BFR: ~13.0 nm; DPS: ~10.5 nm) were slightly larger than those seen in the TEM (Figure 2. 20). Consistent with the literature^[42], the BFR 24-mer in our hands exists as a mixture with a much smaller oligomer which correlates to a dimer. Isolation of one of these peaks, followed by concentration, results in reappearance of both peaks, suggesting

that they are near equilibrium. None of the BFR mutants form the larger oligomer, again emphasizing the sensitivity of BFR to mutation. Removing the E helix from BFR (BFR-E) results in a protein that exists entirely in the dimeric state, further emphasizing the great importance of the helix E in controlling the assembly of the protein cage. Addition of the helix BC to both BFR (BFR+BC) and the E helix deleted mutant (BFR+BC-E) results in mixtures of the dimer and a protein which correlates to the monomer. Light is shed on the sensitivity of BFR to mutagenesis through the realization that native BFR is an equilibrium mixture; this delicate balance may be easily tipped. The E helix may have near complete control over this equilibrium as evidenced by the fact that BFR-E forms a homogeneous population of the lower order complex. Since the BC helix exists at the two-fold symmetry axis between dimers and no packing optimization was performed on the mutants, it stands to reason that addition of the BC helix affects the dimer/monomer transition.

The DPS series is consistent with the conclusions established from the CD and DLS data that the structure of this protein is less sensitive to mutation. Analysis of the DPS series also supports the hypothesis that the E helix has greater oligomerization power than the BC helix. Removal of the BC helix (DPS-BC) results in no observable change in the oligomerization state, whereas addition of the E helix to this mutant (DPS-BC+E) leads to a protein that exists as a complex mixture of states composed of the dimer, the natural 12-mer oligomerization state, and intriguingly, a well-populated population (54%) of a state that is intermediate between the two (Figure 2. 20). Most striking of all is DPS+E which, consistent with the TEM and DLS data, forms a unique, homogenous, and stable population (as evidenced by the melting data) of an oligomer intermediate in size (~11.5 nm)

between BFR and DPS, supporting the notion that the E helix domain of BFR controls higher order oligomerization (Figure 2. 20).

Figure 2. 20 Dynamic light scattering (upper left) and size exclusion chromatograms (right) of cage proteins and mutants. The chromatogram peaks were correlated to elution volumes of standard proteins (Figure 2. 19) to determine Stokes diameters (bottom left). The chromatograms were normalized to 1 and the populations of particles of each diameter are indicated. This figure is taken from Fan *et al.*^[1]

2.3.7 Sedimentation equilibrium calculation

Due to the intriguing and unexpected TEM, DLS, and SEC results suggesting that DPS+E forms capsules that are intermediate in size between those of native DPS and BFR, we sought to determine whether this size difference is due to a different oligomerization state or to some alteration in the assembled conformation. While SEC can be used to determine difference in size of protein

oligomers, sedimentation equilibrium (SE) is routinely used to determine differences in molecular weight.^[50] Analysis of the SE data for DPS and DPS+E suggests that DPS forms an oligomer with a molecular weight of 220 kDa whereas the molecular weight of the DPS+E protein cage is 240 kDa (Figure 2. 21). These data are consistent with the formation of a DPS+E multimer with a stoichiometry similar to that of the DPS 12-mer (Note: the DPS monomer molecular mass is 18.7 KDa; the DPS+E molecular mass is 20.9 KDa). Therefore we conclude that DPS+E, instead of assembling into a unique and novel oligomerization state, forms a uniquely sized protein cage of the same order as the native protein. We are actively seeking a broader understanding regarding the origin of this alteration in size and whether it can be exploited to broaden the fundamental understanding of protein self-assembly, and whether it can be applied to the formation of novel nanomaterials.

Figure 2. 21 Sedimentation equilibrium for DPS (bottom left) and DPS+E (bottom right) at four angular velocities. The data are best fit to DPS=12-mer and DPS+E=12-mer and the residuals of these fits are presented above the data traces. This figure is taken from Fan *et al.*^[1]

2.4 Conclusion

Table 2. 5 Summarize the properties of BFR and DPS proteins which determined by CD, TEM, DLS, and SEC analysis.

Protein	T _m (°C)	ΔT_m (°C)	TEM Outer Diameter (nm)	TEM Cavity Diameter (nm)	DLS Hydrodynamic Diameter (nm)	SEC	
						Stokes Diameter (nm)	Oligomer percentage
BFR	65.3±0.7	NA	12.4±0.3	6.7±0.5	13.4±0.5	13.3	50%
BFR-E	53.9±0.2	11.4±0.9	NA	NA	1.3±0.3	5.8	50%
BFR+BC-E	38.1±2.7	27.2±3.4	NA	NA	3.2±0.5	6.1	100%
BFR+E	53.6±0.2	11.7±0.9	NA	NA	3±2.2	6.0	33%
DPS	67.6±0.3	NA	9.7±0.2	3.9±0.2	10.2±0.1	5.0	66%
DPS-BC	50.1±0.8	17.5±1.1	9.1±0.2	3.7±0.6	6.3±1.1	5.9	40%
DPS-BC+E	63.4±1.4	4.2±1.7	NA	NA	11.4±0.4	5.1	60%
DPS+E	60.3±0.4	7.3±0.7	10.4±0.4	3.6±0.5	12.7±0.1	10.4	100%
						10.5	100%
						10.3	28%
						9.0	54%
						5.6	18%
						11.9	100%

The E helix from BFR and the BC helix from DPS were swapped between the two parent proteins resulting in a series of mutants made up of all possible permutations. Although all mutants exhibited decreased thermostability compared to their parent (Table 2. 5), they all formed helical structures and unfolded cooperatively suggesting tight, near-native tertiary folds. All BFR mutants exhibited a large decrease in thermostability suggesting a higher sensitivity of this protein to mutation compared to that of DPS. Although BFR is more thermodynamically sensitive to mutation, all members of the BFR series exhibited a higher degree of thermal folding reversibility than did the DPS series, implying that DPS is blocked kinetically from returning to the folded state. Of the mutants, only DPS+E and DPS-BC could form protein capsules as evidenced by TEM. Techniques to measure the size of the oligomers that did form (DLS and SEC) demonstrated that none of the BFR mutants formed structures larger in size than a dimer, whereas all members of the DPS series exhibited some populations of higher order oligomers (Table 2. 5). The weak oligomerization power of the BC helix was demonstrated by DPS-BC which formed a nano-structure identical in size to the native DPS and by BFR+BC and BFR+BC-E both of which formed

smaller populations of dimer than that of BFR and BFR-E (Table 2. 5). However, the greater power exhibited by the E helix in controlling oligomerization is made evident by the removal of the E helix from BFR (BFR-E) which results in obliteration of the 24-mer to form a dimer as a homogenous population (Table 2. 5). In addition, a comparison between the DPS-BC+E and DPS-BC mutants further demonstrates the high importance of the E helix in oligomerization, as the addition of the E helix in the former resulted in a mixed population of 12-mer, dimer, and a majority of ~6-mer, which was not seen in DPS-BC (Table 2. 5). Finally, addition of the E helix to native DPS (DPS+E) dramatically emphasizes the importance of the E-helix. The techniques of TEM, DLS, and SEC all show that this mutant forms a unique and homogenous population of a protein that is intermediate in size between the BFR and DPS native proteins (Table 2. 5). The determination of its molecular weight using SEC supports the running hypothesis that this protein is not self-assembling into an oligomer with a greater number of subunits, but that is in fact just larger, most likely due to a change in assembly structure, monomer conformation, or both. Taken together, these data indicate that the E helix in BFR plays a stronger role in the formation of quaternary structure than does the BC helix in DPS and that stabilization of the dimer may not always be the most important self-assembly step in cage formation as was previously assumed.

We are actively seeking a broader understanding of these systems, attempting to further characterize some of the novel structures that were formed, and to apply these mutants to the templation of novel nano-scaled materials. As a larger goal of this project, we are attempting to build more comprehensive systems to investigate other factors that determine protein cage self-assembly, such as grafting bundle surface residues, varying amino acid content, and changing helix

length. Additionally, we hope that this research can provide preliminary data for the establishment of these cage proteins as model systems for exploring self-assembly and quaternary structure formation within a protein design context.

2.5 References

1. Fan, R., Boyle, A. L., Cheong, V. V., Ng, S. L., and Orner, B. P. (2009) A Helix Swapping Study of Two Protein Cages, *Biochemistry* 48, 5623-5630.
2. Seringhaus, M., and Gerstein, M. (2007) Chemistry Nobel rich in structure, *Science* 315, 40-41.
3. (2005) So much more to know, *Science* 309, 78-102.
4. Hill, R. B., Raleigh, D. P., Lombardi, A., and Degrado, W. F. (2000) De novo design of helical bundles as models for understanding protein folding and function, *Accounts of Chemical Research* 33, 745-754.
5. Padilla, J. E., Colovos, C., and Yeates, T. O. (2001) Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments, *Proceedings of the National Academy of Sciences of the United States of America* 98, 2217-2221.
6. Papapostolou, D., Smith, A. M., Atkins, E. D. T., Oliver, S. J., Ryadnov, M. G., Serpell, L. C., and Woolfson, D. N. (2007) Engineering nanoscale order into a designed protein fiber, *Proceedings of the National Academy of Sciences of the United States of America* 104, 10853-10858.
7. Grueninger, D., Treiber, N., Ziegler, M. O. P., Koetter, J. W. A., Schulze, M. S., and Schulz, G. E. (2008) Designed protein-protein association, *Science* 319, 206-209.
8. Klug, A. (1999) The tobacco mosaic virus particle: structure and assembly, *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 354, 531-535.
9. Whitesides, G. M., and Grzybowski, B. (2002) Self-assembly at all scales, *Science* 295, 2418-2421.
10. Lehn, J. M. (1988) Supramolecular chemistry - Scope and perspectives molecules, supermolecules, and molecular devices, *Angewandte Chemie-International Edition in English* 27, 89-112.
11. Liu, X. F., and Theil, E. C. (2005) Ferritins: Dynamic management of biological iron and oxygen chemistry, *Accounts of Chemical Research* 38, 167-175.
12. Wang, Q., Lin, T. W., Tang, L., Johnson, J. E., and Finn, M. G. (2002) Icosahedral virus particles as addressable nanoscale building blocks, *Angewandte Chemie-International Edition* 41, 459-462.
13. Destito, G., Yeh, R., Rae, C. S., Finn, M. G., and Manchester, M. (2007) Folic acid-mediated targeting of cowpea mosaic virus particles to tumor cells, *Chemistry & Biology* 14, 1152-1162.
14. Ren, Y., Wong, S. M., and Lim, L. Y. (2007) Folic acid-conjugated protein cages of a plant virus: A novel delivery platform for doxorubicin, *Bioconjugate Chemistry* 18, 836-843.

15. Meldrum, F. C., Wade, V. J., Nimmo, D. L., Heywood, B. R., and Mann, S. (1991) Synthesis of inorganic nanophase materials in supramolecular protein cages, *Nature* 349, 684-687.
16. Uchida, M., Klem, M. T., Allen, M., Suci, P., Flenniken, M., Gillitzer, E., Varpness, Z., Liepold, L. O., Young, M., and Douglas, T. (2007) Biological containers: Protein cages as multifunctional nanoplateforms, *Advanced Materials* 19, 1025-1042.
17. Andrews, S. C., Smith, J. M. A., Yewdall, S. J., Guest, J. R., and Harrison, P. M. (1991) Bacterioferritins and ferritins are distantly related in evolution - conservation of ferroxidase-center residues, *Febs Letters* 293, 164-168.
18. Harrison, P. M., and Arosio, P. (1996) Ferritins: Molecular properties, iron storage function and cellular regulation, *Biochimica Et Biophysica Acta-Bioenergetics* 1275, 161-203.
19. Arosio, P., and Levi, S. (2002) Ferritin, iron homeostasis, and oxidative damage, *Free Radical Biology and Medicine* 33, 457-463.
20. Andrews, N. C. (2008) Forging a field: the golden age of iron biology, *Blood* 112, 219-230.
21. Cobessi, D., Huang, L. S., Ban, M., Pon, N. G., Daldal, F., and Berry, E. A. (2002) The 2.6 angstrom resolution structure of *Rhodobacter capsulatus* bacterioferritin with metal-free dinuclear site and heme iron in a crystallographic 'special position', *Acta Crystallographica Section D-Biological Crystallography* 58, 29-38.
22. Romao, C. V., Regalla, M., Xavier, A. V., Teixeira, M., Liu, M. Y., and Le Gall, J. (2000) A bacterioferritin from the strict anaerobe *Desulfovibria desulfuricans* ATCC 27774, *Biochemistry* 39, 6841-6849.
23. Macedo, S., Romao, C. V., Mitchell, E., Matias, P. M., Liu, M. Y., Xavier, A. V., LeGall, J., Teixeira, M., Lindley, P., and Carrondo, M. A. (2003) The nature of the di-iron site in the bacterioferritin from *Desulfovibrio desulfuricans*, *Nature Structural Biology* 10, 285-290.
24. Janowski, R., Auerbach-Nevo, T., and Weiss, M. S. (2008) Bacterioferritin from *Mycobacterium smegmatis* contains zinc in its di-nuclear site, *Protein Science* 17, 1138-1150.
25. Yariv, J., Kalb, A. J., Sperling, R., Bauminger, E. R., Cohen, S. G., and Ofer, S. (1981) The composition and the structure of bacterioferritin of *Escherichia coli*, *Biochemical Journal* 197, 171-175.
26. Frolow, F., Kalb, A. J., and Yariv, J. (1994) Structure of a unique twofold symmetrical heme-binding site, *Nature Structural Biology* 1, 453-460.
27. Dautant, A., Meyer, J. B., Yariv, J., Precigoux, G., Sweet, R. M., Kalb, A. J., and Frolow, F. (1998) Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from *E. coli*, *Acta Crystallographica Section D-Biological Crystallography* 54, 16-24.
28. Carrondo, M. A. (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint, *Embo Journal* 22, 1959-1968.
29. Martinez, A., and Kolter, R. (1997) Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps, *Journal of Bacteriology* 179, 5188-5194.
30. Grant, R. A., Filman, D. J., Finkel, S. E., Kolter, R., and Hogle, J. M. (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA, *Nature Structural Biology* 5, 294-303.

31. Almiron, M., Link, A. J., Furlong, D., and Kolter, R. (1992) A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*, *Genes & Development* 6, 2646-2654.
32. Roy, S., Gupta, S., Das, S., Sekar, K., Chatterji, D., and Vijayan, M. (2004) X-ray analysis of *Mycobacterium smegmatis* Dps and a comparative study involving other Dps and Dps-like molecules, *Journal of Molecular Biology* 339, 1103-1113.
33. Ceci, P., Ilari, A., Falvo, E., Giangiacomo, L., and Chiancone, E. (2005) Reassessment of protein stability, DNA binding, and protection of *Mycobacterium smegmatis* Dps, *Journal of Biological Chemistry* 280, 34776-34785.
34. Bellapadrona, G., Chiaraluce, R., Consalvi, V., Ilari, A., Stefanini, S., and Chiancone, E. (2007) The mutations Lys 114 -> Gln and Asp 126 -> Asn disrupt an intersubunit salt bridge and convert *Listeria innocua* Dps into its natural mutant *Listeria monocytogenes* Dps. effects on protein stability at low pH, *Proteins-Structure Function and Bioinformatics* 66, 975-983.
35. Roy, S., Saraswathi, R., Gupta, S., Sekar, K., Chatterji, D., and Vijayan, M. (2007) Role of N and C-terminal tails in DNA binding and assembly in Dps: Structural studies of *Mycobacterium smegmatis* Dps deletion mutants, *Journal of Molecular Biology* 370, 752-767.
36. Roy, S., Saraswathi, R., Chatterji, D., and Vijayan, M. (2008) Structural studies on the second *Mycobacterium smegmatis* Dps: Invariant and variable features of structure, assembly and function, *Journal of Molecular Biology* 375, 948-959.
37. Chowdhury, R. P., Vijayabaskar, M. S., Vishveshwara, S., and Chatterji, D. (2008) Molecular mechanism of in vitro oligomerization of Dps from *Mycobacterium smegmatis*: Mutations of the residues identified by "interface cluster" analysis, *Biochemistry* 47, 11110-11117.
38. Ceci, P., Cellai, S., Falvo, E., Rivetti, C., Rossi, G. L., and Chiancone, E. (2004) DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus, *Nucleic Acids Research* 32, 5935-5944.
39. Kellis, J. T., Nyberg, K., Sali, D., and Fersht, A. R. (1988) Contribution of hydrophobic interactions to protein stability, *Nature* 333, 784-786.
40. Maziar, S. A. (2010) Master Thesis: Structural stabilization of protein nano-capsules via modification of protein-protein interactions.
41. Franceschini, S., Ceci, P., Alaleona, F., Chiancone, E., and Ilari, A. (2006) Antioxidant Dps protein from the thermophilic cyanobacterium *Thermosynechococcus elongatus* - An intrinsically stable cage-like structure endowed with enhanced stability, *Febs Journal* 273, 4913-4928.
42. Andrews, S. C., Smith, J. M. A., Hawkins, C., Williams, J. M., Harrison, P. M., and Guest, J. R. (1993) Overproduction, purification and characterization of the bacterioferritin of *Escherichia coli* and a C-terminally extended variant, *European Journal of Biochemistry* 213, 329-338.
43. Levi, S., Luzzago, A., Cesareni, G., Cozzi, A., Franceschinelli, F., Albertini, A., and Arosio, P. (1988) 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 two H-chain deletion mutants, *Journal of Biological Chemistry* 263, 18086-18092.
44. Arosio, P., Cozzi, A., Ingrassia, R., Levi, S., Luzzago, A., Ruggeri, G., Iacobello, C., Santambrogio, P., and Albertini, A. (1990) A mutational analysis of the epitopes of recombinant human H-ferritin, *Biochimica Et Biophysica Acta* 1039, 197-203.
 45. Santambrogio, P., Pinto, P., Levi, S., Cozzi, A., Rovida, E., Albertini, A., Artymiuk, P., Harrison, P. M., and Arosio, P. (1997) Effects of modifications near the 2-, 3- and 4-fold symmetry axes on human ferritin renaturation, *Biochemical Journal* 322, 461-468.
 46. Young, L., and Dong, Q. H. (2004) Two-step total gene synthesis method, *Nucleic Acids Research* 32.
 47. Stemmer, W. P. C., Cramer, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides, *Gene* 164, 49-53.
 48. Holzwarth, G., and Doty, P. (1965) Ultraviolet circular dichroism of polypeptides, *Journal of the American Chemical Society* 87, 218-228.
 49. Massover, W. H. (1993) Ultrastructure of ferritin and apoferritin - A review, *Micron* 24, 389-437.
 50. Laue, T. M., and Rhodes, D. G. (1990) Determination of size, molecular-weight, and presence of subunits, *Methods in Enzymology* 182, 566-587.

CHAPTER 3 Fabrication of gold nanoparticles inside unmodified horse spleen apoferritin^{[1]*}

3.1 Introduction

Metallic nanomaterials have attracted extensive attention as a result of their emerging properties and their resulting potential for a multitude of applications in electronics, medical diagnosis, sensors and as a fundamental component of catalysts.^[2-6] Many of the applications for metallic nanomaterials, especially those composed of gold and silver, have exploited their unique surface plasmon resonance (SPR) properties.^[2] This property is quite sensitive to the size, shape, composition and chemical environment of the metallic nanomaterials.^[4, 6]

Based on the unique requirements of various applications, metallic nanomaterials have been fabricated through a variety of physical and chemical methods. Many of these methods require extreme conditions such as high temperature, elevated pressure, organic solvents, caustic pH and strong reducing reagents. However, nanomaterial syntheses based on biomolecules have the potential to be more 'green' than some of the traditional methods due to the required mild conditions such as near ambient temperatures, near neutral pH, and the fact that aqueous reaction solutions are often employed. Moreover, bio-assemblies tend to have exquisite nanostructures and their components often can be manipulated easily using molecular biological techniques. As the advantage of the biotemplate approach, new functional groups on the biomolecules could be projected onto the growing nanomaterials enabling the potential for further applications without the need of adding stabilizing agent. Moreover, the

* Parts of this chapter have been published previously with the author of this thesis as the first author. *Small*, **2010**, 6, 1483.

biomolecules can prevent the agglomeration of nanomaterials in high salt solutions and increase the solubility of nanomaterials in aqueous solutions.

In this chapter, we describe a new biotemplated synthetic method to fabricate gold nanomaterials. During the last few decades, quite a few approaches have been developed to synthesize gold nanoparticles due to their many useful applications. Many of the chemical methods to synthesize gold nanoparticles can be placed in two categories. One utilizes citrate to reduce gold ions,^[7] while the other one employs strong binding ligands to transfer gold ions from water to an organic layer where the gold ions are reduced by NaBH_4 in the presence of surfactants.^[8-9] Gold nanoparticles synthesized by these two approaches usually require further functionalization to be utilized in future applications. To avoid additional modification steps, biomolecules including amino acids, peptides and proteins have been employed to fabricate gold nanoparticles and some of these reports have shown that the morphology, size and crystal structure can be regulated by different biomolecules.^[10]

3.1.1 Amino acids-induced the formation of gold nanomaterials

Amino acids are small biomolecules which serve as the building blocks for protein synthesis. Every amino acid consists of a carboxyl group, an amine group and a side chain which varies between different amino acids. At neutral pH condition, the carboxyl group and amine group are ionized. The side chains of some amino acids are charged and some others contain aromatic rings. Amino acids have been employed as simple biotemplates to prepare gold nanomaterials and played an essential role in stabilizing the nanomaterials. Dong and coworkers have developed a technique to incubate various amino acids such as lysine, arginine, tryptophan, tyrosine, and aspartic acid, with tetrachloroaurate in aqueous

solution at room temperature for 12 h.^[11] Their results revealed that spherical gold nanoparticles can be produced with lysine, arginine or tryptophan, while both spherical and rod-shape nanoparticles were formed with tyrosine (Figure 3. 1). Interestingly, it was found that aspartic acid containing conditions alone could result in single crystal hexagonal and truncated triangular shaped gold nanoplates in (111) facet. They postulated that aspartic acid can specifically interact with the (111) gold crystalline face, thereby promoting local reduction at the interaction region. Taken together these results indicate that shape and crystal preference could be regulated by different amino acids.

Figure 3. 1 TEM micrographs of gold nanoparticles prepared with different amino acids (a) lysine, (b) tryptophan, (c) arginine, and (d) tyrosine. The scale bars for the images are (a) 5 nm, (b) 50 nm, (c) 20 nm, and (d) 100 nm respectively. Selected area diffraction patterns of gold nanoparticles synthesized by each amino acids are presented in the insets. This figure is taken from Dong *et al.*^[11]

In the synthesis of gold nanomaterials with amino acids, the amino acids act as the reducing agents to form the gold nanostructures from the chloroaurate ions; therefore the amino acids are spontaneously oxidized during the reaction. To confirm this mechanism, experiments were conducted to characterize the tryptophan reduced gold nanoparticles.^[12] Compared to pure tryptophan, tryptophan reduced gold nanoparticles exhibit 1) a blue shift of the π - π^* region of their UV spectra, and 2) a large change in their proton NMR spectra. These results suggested that the structure of tryptophan has undergone a significant change after reducing the chloroaurate ions implying that the tryptophan is oxidized.

The synthesis of tyrosine-, glycyl-L-tyrosine-, and arginine- reduced gold nanoparticles in alkaline conditions has also been reported by other researchers.^[13] The size of the tyrosine-prepared nanoparticles was larger, on average, compared to those reduced by glycyl-L-tyrosine in identical conditions. L-arginine-prepared particles exhibited a larger size and unique morphology. In addition it was found that the size of the particles could be increased by lowering the gold concentration. Together these data indicate that the size and morphology of the resulting nanomaterials could be manipulated by using amino acids with presumed different reduction potentials and by varying gold ion concentrations. More generally, these studies demonstrate that some amino acids can serve as reducing agents to initiate the growth of gold nanomaterials and stabilize them in aqueous solution.

3.1.2 Peptide-mediated the synthesis of gold nanoparticles

Peptides are small polyamides which are composed of less than ~25 amino acids. Although peptides are composed of amino acids, the amino acid-based gold nanoparticle synthesis can not be directly applied to conditions using peptides. These differences could be due to altered projection of functionality which is a result of secondary structure formation in the peptides. As these conformationally related effects along with the opportunity to combine functional groups with varied chemical properties could potentially afford a degree of nuanced control not available with amino acids, methods involving peptides have been the focus of much scientific interest. A histidine rich peptide (AHHAHHAAD) was reported to possess a high affinity for metal ions. Slocik and coworkers have utilized it to mediate the growth of gold nanoparticles in aqueous solution.^[14-15] Gold ions were mixed with the peptide in a 1:1 molar ratio and nanoparticles were formed by addition of reducing agent (sodium citrate or sodium borohydride). The UV spectra

of the particles showed the characteristic plasmon resonance peak at 524 nm. In addition, the average diameter of the particles was 9.5 nm. Immunoassays have been conducted and antibodies fused to the histidine rich peptides were able to detect the gold nanoparticles suggesting that the peptides mediate the growth of gold nanoparticles through surface recognition. With confirmation that these peptides act as ligands, nanoparticles can be produced and functionalized efficiently in one-pot synthesis without performing ligand exchange steps. In a later paper, Matsui and coworkers immobilized histidine rich peptides on nanotubes to mediate the growth of uniform sized gold nanocrystals (Figure 3. 2).^[16] The peptide-functionalized nanotubes were mixed with organic gold complex ($\text{Cl}_3\text{AuPMe}_3$). The gold ions were slowly trapped by the histidine rich peptides and crystallized by reduction with NaBH_4 . This procedure resulted in highly monodisperse gold nanocrystals with an average diameter of 6 nm.^[16] The diffraction pattern of these particles indicated that the crystalline phase was in (111) and (220).^[16] In a subsequent report, it was found that the packing density of gold nanocrystals could be manipulated by altering the pH and ion concentration, however, the diameter of gold nanoparticles was retained.^[17] As the packing density on the nanotubes has significant influence on conductivity, this technology could be utilized as a conductivity-tunable building block in electronics and sensor devices.^[17]

Figure 3. 2 Scheme of synthesizing Au particles on nanotubes. (a) Immobilized histidine-rich peptide on nanotubes. (b) Immobilized Au ions on histidine-rich peptide. (c) Au nanoparticles grown on the nanotubes after reduction. (d) TEM image of Au nanoparticles grown on the nanotubes and the scale bar is 200 nm. This figure is taken from Matsui *et al.*^[17]

Some other peptides like the Flag and A3 peptide (DYKDDDDK and AYSSGAPPMPPF respectively) have been used to synthesize gold nanoparticles in HEPES buffer.^[18] Biotinylated anti-Flag antibody could successfully recognize Flag peptides which were ligated to the surface of gold nanoparticles during the particle formation and the complexes were mixed with streptavidin coated quantum dots to form bio-assembled hybrid nanostructures.^[19] To conclude, with the assistance of high throughput screening techniques like phage display, peptides that have high affinity to metal ions/particles can be easily selected and employed for further bio-system applications.

3.1.3 Protein-assisted fabrication of gold nanoparticle

Proteins are essentially large peptides. Their linear length, functional diversity, and amino acid sequence cause them to adopt thermally stable folds, thereby defining the proteins' biological and molecular recognition activity. Various proteins have been directly utilized in the fabrication process of gold nanoparticles. The common protein, bovine serum albumin (BSA) is a suitable candidate for the synthesis of gold nanoparticles as it possesses many sulfur-, oxygen-, and nitrogen- containing residues all of which have a high affinity for gold ions.^[20] After mixing BSA with gold ions (AuCl_4^-), the reducing reagent NaBH_4 was added resulting in well dispersed gold nanoparticles with an average diameter less than 2 nm.^[20] Their results indicated that the backbone and the functional groups on the side chain remained intact. Whereas, the disulfide bonds in BSA were broken resulting in free thiol groups available to make strong interactions with the gold nanoparticles. This study demonstrated that proteins could be conjugated to gold nanoparticles during their formation.

Another protein, α -amylase, was reported to reduce gold ions (AuCl_4^-) while maintaining its enzymatic activity.^[21] It was speculated that the protein was interacting with the particle surface through free thiol groups. Interestingly, the active site of the enzyme which is positioned opposite these implicated cysteines was not affected during the gold particle formation as enzymatic turnover could still be observed. Although many other enzymes have been screened for protein-assisted gold nanoparticle synthesis, only *EcoR* I could successfully produce gold nanoparticles. The only structural similarity between α -amylase and *EcoR* I is that both of the proteins have free cysteines which were presumably essential in the reduction of gold ions to gold nanoparticles.

Amino acids, peptides, and the above mentioned proteins all seem to function by surface binding of growing gold nanoparticles. The size and shape of the resulting particles is depending on the affinity of this binding along with how the binding on and off rates related to the chemical rate of the reduction of the gold. Much optimization is therefore required to produce the particles and their resulting size and shape can rarely be predicted rationally from the outset. However, some proteins assemble into unique quaternary structures and therefore could serve as platforms to template gold nanoparticle formation. In these cases the resulting material reflects the size and 3D shape of the templating protein.

Mann and coworkers have reported the use of tobacco mosaic virus (TMV), a rod shape virus with 18 nm in diameter and 300 nm in length, to perform this templation by incubating the virus with gold ions (AuCl_4^-) at acidic pH.^[22] After adding the reducing agent hydrazine hydrate, gold ions which bound to the TMV surface acted as nucleation sites to promote nanoparticle growth. Spherical gold nanoparticles with diameters of 8.6 ± 3 nm densely covered the external surface of

the viral capsids. In a subsequent report, they optimized the conditions to deposit gold nanoparticles homogeneously on TMV through repeated the addition of gold ions and reducing agent in aliquots in the presence of poly-L-lysine (Figure 3. 3).^[23]

Figure 3. 3 TEM characterization of gold nanoparticles densely fabricated on TMV capsids by TEM. (a) Low and (b) high magnification TEM images. This figure is taken from Mann *et al.*^[23]

As we have discussed above, ferritin is a well characterized cage-like protein which consist of twenty-four subunits. The hollow protein cage has an exterior diameter of ~12 nm and an interior cavity of ~8 nm. Due to its unique structure and high stability, it is considered as an excellent template to fabricate nanoparticles with uniform size. With that said, the mineralization of gold nanoparticles in the presence of ferritin cages has only recently been reported in a series of two papers from Dmochowski and coworkers. Reaction between either monoanionic AuCl_4^- or neutral AuCl_3 and unmodified horse spleen apoferritin (HSAFn) resulted only in gold mineralization on the outside of the protein, but the size of these deposits could be controlled by the choice of the reductant.^[24] In a subsequent report, human heavy chain ferritin was modified by removing solvent-exposed gold binding amino acids, such as cysteine and histidine, from the outer surface and by lining the interior surface with cysteine residues (Figure 3. 4). Gold

nanoparticles were successfully incorporated inside the cavity of this modified protein by addition of AuCl_3 followed by reduction with MOPS.^[25]

Figure 3. 4 (a) Scheme of mutated human heavy chain ferritin mineralized gold nanoparticles inside the protein cage. (b) TEM image of gold nanoparticles synthesized within the cavity of ferritin. This figure is taken from Dmochowski *et al.*^[25]

3.1.4 Aims and objectives

In this chapter, a novel strategy has been developed to mineralize gold nanoparticles inside native horse spleen apoferritin (HSAFn) without manipulating or modifying the protein shell. In comparison with methods involving modified protein cage, this method eliminates the need of protein engineering. In addition, this general method can be readily applied to other protein cages.

A challenge of this undertaking is that gold ions have a poor natural preference for the ferritin interior over that of the exterior which thereby favors nanoparticle formation on the outside of the cage. Therefore, to encourage formation of particles only on the inside, we designed a strategy to trap a small number of gold ions inside the ferritin cavity and then isolate the encapsulated gold away from solution gold which would have had the potential of mineralizing on the outside of the protein. Subsequently the entrapped gold ions could be rapidly reduced to form gold nanoclusters. Then in a second reduction step, additional gold ions could be added along with a weak reductant that preferentially permits mineralization on the nanocluster seed. Thus, the nanoparticle would grow until it matches the size of the inner cavity. The two step reduction strategy should be

amenable to nanoparticle growth inside almost any nanocage protein and this growth should be limited by the size of the protein shell. The resulting particles size can be rationally predicted prior to synthesis with virtually little optimization. However, with the application of this strategy, challenges it faces are: 1) How to keep the initial number of gold ions low inside the ferritin so that the majority of the particle growth happens during the second reduction. 2) How to isolate the ferritin encapsulated ions. 3) How to control the selectivity of the reductions in each of the two steps.

3.2 Experimental

3.2.1 Protein purification

Horse spleen apoferritin (HSAFn, Sigma-aldrich) was stored in glycerol solution (50% glycerol, 75 mM NaCl) at a high concentration (25 mg/mL). The stock solution was applied to a gel filtration column (Superdex 200 10/300 GL, GE Healthcare) equilibrated with phosphate buffer (50 mM NaH₂PO₄, 150 mM NaCl, pH 7.0) to remove aggregates. The elution was monitored by UV (at 280 nm) with a flow rate of 0.5 mL/min. The fractions were collected, pooled together and the concentration of the protein was determined by Bicinchoninic acid (BCA) protein assay kit (Merck).

3.2.2 Gold nanoparticles mineralization in horse spleen ferritin

Chloroauric acid (HAuCl₄, 22.6 µL of a 0.1 M solution) was added to a solution of purified HSAFn (1 mL of a 1 mg/mL solution) with the molar ratio of 1000:1. The mixture was incubated at room temperature for 3 h and centrifuged at 10000 rpm for 10 min. The supernatant was applied to a desalting column (Sephadex G-25, GE Healthcare) equilibrated with Tris buffer (50 mM Tris, pH

7.5) at a flow rate of 1 mL/min. The fractions containing protein were combined and sodium borohydride (NaBH_4 , 20 μL of a 0.1 M solution) was added into the solution. The resulting mixture was stirred slowly for 10 min before unagitated incubation at room temperature for 3 h. Freshly dissolved ascorbic acid (30 μL of a 0.1 M solution) and additional HAuCl_4 (10 μL of a 0.1 M solution) were added to the protein solution, and it was incubated with no agitation overnight at room temperature. The resulting solution was centrifuged at 10000 rpm for 10 min and the supernatant was purified by a gel filtration column (Superdex 200 10/300 GL, GE Healthcare) equilibrated with phosphate buffer (50 mM NaH_2PO_4 , 150 mM NaCl , pH 7.0). To isolate a homogeneous preparation of the gold nanoparticles mineralized in horse spleen ferritin, the sample was then applied to a 20% to 60% sucrose gradient and centrifuged at 40,000 rpm for 5 h. The bottom layer was collected and subjected to buffer exchange (50 mM Tris, pH 7.5).

3.2.3 Ultraviolet-Visible spectroscopy (UV)

Protein concentration of apoferritin and mineralized ferritin were determined by BCA protein assay kit (Merck). Samples were diluted to 100 $\mu\text{g/mL}$ with phosphate buffer (50 mM NaH_2PO_4 , 150 mM NaCl , pH 7.0) and recorded on a Cary 100 UV-visible spectrophotometer with a wavelength range from 250 nm to 700 nm.

3.2.4 Size exclusion chromatography (SEC)

Apoferritin and mineralized ferritin were analyzed by fast protein liquid chromatography (FPLC, GE Healthcare, ÄKTAexplorer 10 S) using a gel filtration column (Superdex 200 10/300 GL, GE Healthcare) which was equilibrated with phosphate buffer (50 mM NaH_2PO_4 , 150 mM NaCl , pH 7.0). All samples (100

µg/mL) were monitored at 280 nm (for protein) and 520 nm (for gold nanoparticles).

3.2.5 Native polyacrylamide gel electrophoresis

Apo ferritin and mineralized ferritin were analyzed by native polyacrylamide gel electrophoresis with 4% stacking gel and 7% resolving gel. The gel was stained with Coomassie Brilliant Blue.

3.2.6 Transmission electron microscopy (TEM)

Sample solutions in phosphate buffer (50 mM NaH₂PO₄, 150 mM NaCl, pH 7.0) were dropped on a piece of parafilm. Copper grids (formvar/carbon coated copper, 300 mesh) were gently placed on top of the samples for 1 min. After the samples had adhered to the grids, filter paper was used to wick the excess solution. Subsequently, the grids were placed on drops of negative stain solution (Nano W, Nanoprobes) for 1 min. The grids were blotted with filter papers again and left to air dry. TEM data were collected on a JEOL JEM-1400 microscopy operating at 120 keV. Particle sizes were determined from 100 measurements by using image analysis software Image J (National Institutes of Health). The yield of mineralized nanoparticles was calculated from the number of gold nanoparticles mineralized horse spleen ferritin divided by the total ferritin cages observed in TEM images.

3.2.7 High-resolution transmission electron microscopy (HRTEM), energy-dispersive X-ray spectroscopy (EDX) and selected area diffraction (SAD) analysis

Gold nanoparticles mineralized in horse spleen ferritin were observed at high magnifications without negative staining and lattice images of cores were recorded on a JEOL JEM-2100F microscope operating at 200 keV. Energy-dispersive X-ray (EDX) spectra were obtained on the same sample to determine

the elemental composition. Selected area diffraction (SAD) patterns were collected with the camera length of 20 cm and d-spacings were measured.

3.3 Results and Discussion

3.3.1 Design of a novel two-step method to fabricate gold nanoparticles inside unmodified horse spleen ferritin

In the experiment, anionic AuCl_4^- was utilized as the gold source to ensure that the number of gold ions was low inside the ferritin cage as the interior of the protein cage is negatively charged.^[26-27] Moreover, calculations of the ferritin electrostatic potential revealed that the field at the three-fold axis is directed through the channels toward the inside of the cavity presumably to facilitate the transport of cations.^[9] Therefore, transport of gold anions into the cavity with a mechanism analogous to cations would be disfavored and thus ensure a low amount of ions trapped inside the cage. A mixture of AuCl_4^- and protein solution reached equilibrium after incubating for several hours and resulted in a homogeneous, pale yellow solution. To isolate the unmineralized, ferritin-entrapped gold ions we took advantage of protein biochemical techniques. Size exclusion chromatography permitted the desalting of the protein/gold ion solution while isolating the protein with trapped ions in a mild way that kept the protein cage intact. The first reduction was executed with the strong reductant, NaBH_4 , to form small gold clusters (HSFn-Au clusters) which gave a light reddish-brown solution while the second reduction of the gold clusters to fill up the protein cavities with gold nanoparticles (HSFn-AuNPs) was achieved with the mild reducing agent, ascorbic acid, which is unable to reduce gold in the absence of mineralized seeds.^[28] The solution of gold-mineralized ferritin quickly became

ruby-red. The scheme of the mineralization process is shown below (Figure 3. 5). To isolate pure and uniform size HSFn-AuNPs, gel filtration was applied to remove protein aggregations. Subsequently, ultracentrifugation, another mild biochemical technique through a sucrose gradient which separates based on density was utilized to remove empty or incompletely mineralized ferritin. This mild, simple, and robust procedure yields highly homogeneous gold nanoparticles mineralized inside the unmodified ferritin.

Figure 3. 5 Schematic representation of the two-step reduction procedure using native horse spleen ferritin (HSAFn) to mineralize gold nanoparticles. Gold clusters (HSFn-Au clusters) are mineralized inside HSAFn by reducing AuCl_4^- trapped inside the ferritin cavity. The formation of gold nanoparticles (HSFn-AuNPs) is promoted by addition of ascorbic acid and additional AuCl_4^- . This figure is taken from Fan *et al.*^[1]

3.3.2 Characterization of the gold nanoparticles mineralized inside unmodified horse spleen ferritin by UV-Vis absorption

Characterization was conducted by UV-Vis spectroscopy (Figure 3. 6). The spectra of HSAFn, HSFn-Au clusters, and HSFn-AuNPs, all revealed strong absorbance at 280 nm mainly due to the presence of the protein. For the HSFn-

AuNPs, an additional absorption peak at 519 nm was observed indicating the characteristic surface plasmon resonance (SPR) of spherical gold nanoparticles which are less than 20 nm in diameter.^[2, 29] In contrast, the HSFn-Au clusters exhibited no identifiable SPR band, suggesting that the size of mineral cores was less than 2 nm in diameter.^[4]

Figure 3. 6 UV-Visible spectra of horse spleen apoferritin (HSAFn, black), gold clusters mineralized in horse spleen ferritin (HSFn-Au clusters, blue) and gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs, red). This figure is taken from Fan *et al.*^[1]

3.3.3 Characterization of gold nanoparticles inside unmodified horse spleen ferritin by SEC

SEC was carried out to monitor apoferritin and mineralized ferritin at 280 nm (protein) and 520 nm (gold nanoparticles) and to determine if the gold nanoparticle were mineralized inside the ferritin. The HSAFn eluted at 11.5 mL which is consistent with a self-assembled 24-mer protein cage. As expected it exhibited no peak in the 520 nm channel. The elution profiles monitored at 280 nm for both HSFn-Au clusters and HSFn-AuNPs resulted in peaks with identical elution volumes as HSAFn which implied that the protein cages were intact and endured no major alternations during mineralization (Figure 3. 7). In addition, the chromatograms of both HSFn-Au clusters and HSFn-AuNPs showed coelution of

protein and the gold suggesting that mineralized had occurred inside the ferritin. Small shoulder peaks (~10 mL elution volume) were observed in the elution profiles from both the HSFn-Au clusters and HSFn-AuNPs. Isolation and characterization of these peaks demonstrated that a small amount of protein aggregation formed during the reduction steps. This is not surprising as these conditions are, of course, non-native. However, the fact that this aggregation is so slight is a testament to the mildness of this procedure (Figure 3. 8).

Figure 3. 7 Size exclusion chromatography of horse spleen apoferritin (HSAFn, top), gold clusters mineralized in horse spleen ferritin (HSFn-Au clusters, middle) and gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs, bottom). Elution profiles were measured at 280 nm (black line) and 520 nm (red line). This figure is taken from Fan *et al.*^[1]

Figure 3. 8 Transmission electron microscopy images of protein aggregations in the shoulder peaks (at ~10 mL elution volume) of (left) HSFn-Au cluster and (right) HSFn-AuNPs after size exclusion chromatography. This figure is taken from Fan *et al.*^[1]

3.3.4 Characterization of gold nanoparticles inside unmodified horse spleen ferritin by native PAGE gel

To further confirm the composite nature of gold nanoparticles mineralized ferritin, HSAFn and HSFn-AuNPs were electrophoresed in a 7% polyacrylamide gel under native conditions. Before staining, the band of HSFn-AuNPs was red due to the mineralized gold nanoparticles. After staining with Coomassie Brilliant Blue, bands of HSAFn and HSFn-AuNPs were observed with the same electrophoretic mobility suggesting that the overall surface charge of the ferritin was unchanged and the protein cage remained intact during the mineralization process (Figure 3. 9).

Figure 3. 9 Native gel electrophoresis of horse spleen apoferritin (HSAFn, left) and gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs, right). Gel was stained with Coomassie blue. This figure is taken from Fan *et al.*^[1]

3.3.5 Characterization of gold nanoparticles inside unmodified horse spleen ferritin by TEM

TEM was conducted to visualize HSAFn, HSFn-Au clusters, and HSFn-AuNPs with negative staining. The micrographs revealed that HSAFn self-assembled into cage-like structures (Figure 3. 10a) with an average diameter of 13.1 ± 0.7 nm which was consistent with the literature (12 nm).^[30] The HSFn-Au clusters were difficult to observe in TEM operating at 120 keV due to their small size and low electron density (Figure 3. 10b), however, the HSFn-AuNPs could be

visualized easily, and the gold particles appeared to be surrounded by intact protein shells (Figure 3. 10 c-d). In addition, the shape of the nanoparticles was spherical and they were uniform. The highly homogenous population had an average diameter of 6.3 ± 0.8 nm which was consistent with the interior dimension (7 nm)^[30] of the protein cage suggesting that the protein was responsible for controlling the size.

Figure 3. 10 Transmission electron micrographs of a) negatively stained horse spleen apoferritin (HSAFn) b) negatively stained gold clusters mineralized in horse spleen ferritin (HSFn-Au clusters) c) and d) negatively stained gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs). The scale bar is 50 nm for all micrographs. This figure is taken from Fan *et al.*^[1]

3.3.6 Characterization of gold nanoparticles inside unmodified horse spleen ferritin by element composition analysis

Elemental composition analysis of the HSFn-AuNPs was determined using energy-dispersive X-ray (EDX) techniques (Figure 3. 11). The spectrum showed $M\alpha$, $L\alpha$ and $L\beta$, Au peaks (2.1, 9.7, and 11.4 keV respectively) which confirmed the presence of Au as the element of the mineral cores. Moreover, a C peak (at

0.27 keV) and three Cu peaks (0.9, 8.0, and 8.9 keV) were observed which could be attributed to the protein shell and the TEM grid.

Figure 3. 11 EDX spectrum of unstained gold nanoparticles mineralized in horse spleen ferritin (HSFn-AuNPs). The spectrum was obtained on a copper grid. This figure is taken from Fan *et al.*^[1]

3.3.7 Gold nanoparticles crystalline analysis by TEM and SAED

The structure of the gold nanoparticles mineralized in horse spleen ferritin was investigated at an atomic scale by high-resolution transmission electron microscopy (HRTEM). The micrograph showed a clear lattice image indicating that some gold particles are single crystals (Figure 3. 12a). However, other micrographs of the nanoparticles from our preparation indicated that some of them are polycrystalline (Figure 3. 12b). This polycrystallinity suggests that in some cases, the gold crystals might be grown from one or more nucleation sites. Gold usually forms face central cubic (*fcc*) crystals. Measurement of the lattice spacings from the HRTEM image was 2.37 Å, corresponding to the value of the (111) facet of *fcc* gold crystals (2.36 Å). To further understand the crystal structure of the gold particles, selected-area electron diffraction (SAED) patterns from a large number of particles were obtained (Figure 3. 12c). The observed d-spacings at 2.4, 2.1, 1.5, 1.3 and 0.97 Å correspond to miller indices of (111), (200), (220), (311) and (331) in gold crystals respectively (Table 3. 1) which further confirmed that the

mineralized nanoparticles are *fcc* (face centered cubic) crystallized gold nanoparticles. Other approaches to synthesize gold nanoparticles could only form *fcc* lattice and particles might be single crystalline or polycrystalline.^[9, 31-32]

Figure 3. 12 High resolution transmission electron microscopy (HRTEM) of a) single crystal and b) polycrystal gold nanoparticles mineralized in horse spleen ferritin without negative staining. c) Selected area electron diffraction pattern (SAED) for HSFn-AuNPs and the camera length is 20 cm. This figure is taken from Fan *et al.*^[1]

Table 3. 1 Theoretical and observed d-spacings for gold nanoparticles mineralized in horse spleen ferritin. This table is taken from Fan *et al.*^[1]

d-spacing [Å]	Measured d-spacing [Å]	hkl
2.36	2.4	111
2.04	2.1	200
1.44	1.5	220
1.23	1.3	311
0.94	0.97	331

3.3.8 Mechanistic study to determine the necessity of the sucrose gradient ultracentrifugation step

After the HSFn-AuNPs were synthesized, gel filtration chromatography was applied to remove the small amount of protein aggregates. However, as this technique separates based on size differences, the resulting HSFn-AuNPs preparation had $40 \pm 3\%$ empty or incompletely mineralized ferritin (Figure 3. 13 a-b). A homogeneous preparation of HSFn-AuNPs was obtained after sucrose gradient centrifugation, which resulted in $94 \pm 2\%$ of the ferritins filled with nanoparticles with average diameter of 6.3 ± 0.8 nm (Figure 3. 13 c-f). Therefore,

the sucrose gradient ultracentrifugation step is essential to remove unmineralized ferritin and obtain the gold nanoparticles mineralized in ferritin with uniform size.

Figure 3. 13 TEM images of HSFn-AuNPs a) and b) before and c)-f) after sucrose gradient ultracentrifugation. This figure is taken from Fan *et al.*^[1]

3.4 Conclusion

In conclusion, we report a new two-step reduction method to produce gold nanoparticles inside ferritin protein cages that requires no modification of the protein. The formation of nanoparticles on the outside could be suppressed by first loading gold anions into the cage followed by removal of the excess ions in solution immediately before reacting with the first reductant. The resulting encapsulated clusters are then used to seed nanoparticle formation by slow reduction. We confirmed that highly monodisperse nanocrystals are formed within intact protein shells and the size of the nanoparticles correlates with the interior diameter of the cages. Gold nanoparticles are of intense interest due to their multiple applications and our mild and flexible fabrication method could aid in that research. In addition, the protein surface could be easily modified through chemical and genetic methods to direct other functions for further applications. We also think that our approach to first form clusters within the protein cages to

nucleate nanoparticle growth could be used to synthesize core/shell or alloy nanoparticles inside ferritins. Furthermore, the fact that our procedure does not require any modification of the protein, gives this method an added advantage of being readily amenable to any cage protein whether they are altered structural variants of the ferritins or viruses.

3.5 References

1. Fan, R., Chew, S. W., Cheong, V. V., and Orner, B. P. (2010) Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin, *Small* 6, 1483-1487.
2. Daniel, M. C., and Astruc, D. (2004) Gold nanoparticles: Assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology, *Chemical Reviews* 104, 293-346.
3. Cushing, B. L., Kolesnichenko, V. L., and O'Connor, C. J. (2004) Recent advances in the liquid-phase syntheses of inorganic nanoparticles, *Chemical Reviews* 104, 3893-3946.
4. Burda, C., Chen, X. B., Narayanan, R., and El-Sayed, M. A. (2005) Chemistry and properties of nanocrystals of different shapes, *Chemical Reviews* 105, 1025-1102.
5. Rosi, N. L., and Mirkin, C. A. (2005) Nanostructures in biodiagnostics, *Chemical Reviews* 105, 1547-1562.
6. Stewart, M. E., Anderton, C. R., Thompson, L. B., Maria, J., Gray, S. K., Rogers, J. A., and Nuzzo, R. G. (2008) Nanostructured plasmonic sensors, *Chemical Reviews* 108, 494-521.
7. Turkevich, J., Stevenson, P. C., and Hillier, J. (1951) A study of the nucleation and growth processes in the synthesis of colloidal gold, *Discussions of the Faraday Society*, 55-75.
8. Brust, M., Walker, M., Bethell, D., Schiffrin, D. J., and Whyman, R. (1994) Synthesis of thiol-derivatized gold nanoparticles in a 2-phase liquid-liquid system, *Journal of the Chemical Society-Chemical Communications*, 801-802.
9. Brust, M., Fink, J., Bethell, D., Schiffrin, D. J., and Kiely, C. (1995) Synthesis and reactions of functionalized gold nanoparticles, *Journal of the Chemical Society-Chemical Communications*, 1655-1656.
10. Dickerson, M. B., Sandhage, K. H., and Naik, R. R. (2008) Protein- and peptide-directed syntheses of inorganic materials, *Chemical Reviews* 108, 4935-4978.
11. Shao, Y., Jin, Y. D., and Dong, S. J. (2004) Synthesis of gold nanoplates by aspartate reduction of gold chloride, *Chemical Communications*, 1104-1105.
12. Selvakannan, P., Mandal, S., Phadtare, S., Gole, A., Pasricha, R., Adyanthaya, S. D., and Sastry, M. (2004) Water-dispersible tryptophan-

- protected gold nanoparticles prepared by the spontaneous reduction of aqueous chloroaurate ions by the amino acid, *Journal of Colloid and Interface Science* 269, 97-102.
13. Bhargava, S. K., Booth, J. M., Agrawal, S., Coloe, P., and Kar, G. (2005) Gold nanoparticle formation during bromoaurate reduction by amino acids, *Langmuir* 21, 5949-5956.
 14. Slocik, J. M., Moore, J. T., and Wright, D. W. (2002) Monoclonal antibody recognition of histidine-rich peptide encapsulated nanoclusters, *Nano Letters* 2, 169-173.
 15. Slocik, J. M., and Wright, D. W. (2003) Biomimetic mineralization of noble metal nanoclusters, *Biomacromolecules* 4, 1135-1141.
 16. Djalali, R., Chen, Y., and Matsui, H. (2002) Au nanowire fabrication from sequenced histidine-rich peptide, *Journal of the American Chemical Society* 124, 13660-13661.
 17. Djalali, R., Chen, Y. F., and Matsui, H. (2003) Au nanocrystal growth on nanotubes controlled by conformations and charges of sequenced peptide templates, *Journal of the American Chemical Society* 125, 5873-5879.
 18. Slocik, J. M., Stone, M. O., and Naik, R. R. (2005) Synthesis of gold nanoparticles using multifunctional peptides, *Small* 1, 1048-1052.
 19. Slocik, J. M., Govorov, A. O., and Naik, R. R. (2006) Optical characterization of bio-assembled hybrid nanostructures, *Supramolecular Chemistry* 18, 415-421.
 20. Burt, J. L., Gutierrez-Wing, C., Miki-Yoshida, M., and Jose-Yacaman, M. (2004) Noble-metal nanoparticles directly conjugated to globular proteins, *Langmuir* 20, 11778-11783.
 21. Rangnekar, A., Sarma, T. K., Singh, A. K., Deka, J., Ramesh, A., and Chattopadhyay, A. (2007) Retention of enzymatic activity of alpha-amylase in the reductive synthesis of gold nanoparticles, *Langmuir* 23, 5700-5706.
 22. Dujardin, E., Peet, C., Stubbs, G., Culver, J. N., and Mann, S. (2003) Organization of metallic nanoparticles using tobacco mosaic virus templates, *Nano Letters* 3, 413-417.
 23. Bromley, K. M., Patil, A. J., Perriman, A. W., Stubbs, G., and Mann, S. (2008) Preparation of high quality nanowires by tobacco mosaic virus templating of gold nanoparticles, *Journal of Materials Chemistry* 18, 4796-4801.
 24. Zhang, L., Swift, J., Butts, C. A., Yerubandi, V., and Dmochowski, I. J. (2007) Structure and activity of apoferritin-stabilized gold nanoparticles, *Journal of Inorganic Biochemistry* 101, 1719-1729.
 25. Butts, C. A., Swift, J., Kang, S. G., Di Costanzo, L., Christianson, D. W., Saven, J. G., and Dmochowski, I. J. (2008) Directing noble metal ion chemistry within a designed ferritin protein, *Biochemistry* 47, 12729-12739.
 26. Yang, X. K., and Chasteen, N. D. (1996) Molecular diffusion into horse spleen ferritin: A nitroxide radical spin probe study, *Biophysical Journal* 71, 1587-1595.
 27. Yang, X. K., Arosio, P., and Chasteen, N. D. (2000) Molecular diffusion into ferritin: Pathways, temperature dependence, incubation time, and concentration effects, *Biophysical Journal* 78, 2049-2059.
 28. Jana, N. R., Gearheart, L., and Murphy, C. J. (2001) Wet chemical synthesis of high aspect ratio cylindrical gold nanorods, *Journal of Physical Chemistry B* 105, 4065-4067.

29. Kelly, K. L., Coronado, E., Zhao, L. L., and Schatz, G. C. (2003) The optical properties of metal nanoparticles: The influence of size, shape, and dielectric environment, *Journal of Physical Chemistry B* 107, 668-677.
30. Harrison, P. M., Andrews, S. C., Artymiuk, P. J., Ford, G. C., Guest, J. R., Hirzmann, J., Lawson, D. M., Livingstone, J. C., Smith, J. M. A., Treffry, A., and Yewdall, S. J. (1991) Probing structure-function relations in ferritin and bacterioferritin, *Adv. Inorg. Chem.* 36, 449-486.
31. Luo, Y. L. (2006) Preparation of single-crystalline gold nanoparticles through a thermal process, *Materials Letters* 60, 3361-3363.
32. Moon, S. Y., Tanaka, S., and Sekino, T. (2010) Crystal Growth of Thiol-Stabilized Gold Nanoparticles by Heat-Induced Coalescence, *Nanoscale Research Letters* 5, 813-817.

CHAPTER 4 **Conclusion**^{*}

Protein cages have been the focus of studies across multiple scientific disciplines. They have been used to deliver drugs, as templates for nanostructured materials, as substrates in the development of bio-orthogonal chemistry, and to restrict diffusion to study spatially confined reactions.

In chapter 2, two cage proteins, DPS and BFR which self-assemble to form a 12-mer with tetrahedral symmetry and an octahedrally symmetric 24-mer respectively have been applied to understand the fundamentals of protein quaternary structure formation. Their monomers fold into four-helix bundle structures and share strong similarities of both sequence and tertiary structure. However, they differ in the presence of a short additional helix. In BFR, the fifth helix is at the C-terminus and is positioned along the four-fold symmetry axis, whereas with DPS, an extra helix helps to define the two-fold axis in the cage and is located between the second and third helices in the monomer bundle. In an attempt to investigate if these short helices govern protein assembly, mutants were designed and produced that delete and swap these minidomains. All mutants were successfully cloned, expressed and purified and they form highly helical structures that unfold cooperatively as evidenced by thermal melting monitored by circular dichroism. The stability of the DPS mutants was less sensitive to mutation than that of BFR, however, the refolding of the DPS mutants was less reversible indicating a large kinetic barrier to refolding. Dynamic light scattering, size exclusion chromatography, and sedimentation equilibrium experiments demonstrated that although many of the BFR mutants do not self-assemble and form lower-order complexes, many DPS mutants could form cages despite their

^{*} Parts of this chapter have been published previously with the author of this thesis as the first author. *Biochemistry*, **2009**, 48,5623. and *Small*, **2010**, 6, 1483.

unnatural design. Taken together, our data indicate that the BC helix is less important than the E helix for overall cage self-assembly, suggesting that dimerization in nanostructure formation is not as important as previously assumed. Additionally, we found that fusing the minidomain from BFR onto DPS results in a mutant that assembles into a homogeneous population of a novel protein oligomer. This assembled cage while still formed from twelve subunits is larger in overall shape than that of the native protein. We are currently attempting to investigate this using crystallography.

In chapter 3, native horse spleen ferritin was utilized to successfully synthesize size-constrained gold nanoparticles within the cavity by using a two-step process relying on initial nanocluster formation. The resulting gold nanoparticles were analyzed by UV-Vis spectroscopy to demonstrate characteristic absorption properties, and size exclusion chromatography which showed that the particles coelute with the nanocage ferritin and that the nanoparticles are associated with the protein cages. Further characterization by TEM results revealed that the monodisperse and crystalline gold nanoparticles are synthesized in the ferritin protein cages and the growth of the particles is limited by the protein cages. This study provides the possibilities to synthesize gold/silver-core/shell nanoparticles and fabricate metal alloys. Moreover, we are currently expanding this general method to other metals and other proteins. The mineralized ferritins can be exploited for a wide range of applications such as bioimaging, cell targeting, drug delivery, and electronics.

APPENDIXES

Appendix 1 Details on gene design of BFR and DPS proteins

BFR	
A MKGDTKUIINYLNKLLGNELUAINQYFLHARMFKNWGLKRLNDUEYHESIDEMKHADRYIERILFLEGLPNLQDLGKLNIGEDUEEMLRSDLALELDGAKNLRERIGYADSVHDYVSRDMMIEILRDEEGHIDWLETELDLIQKMGKQNYLQAQIREEG	
B BFR (474bp) 5'-3' <div> <div> F1 ATG-AAA-GGT-GAT-ACC-AAA-GTG-ATC-AAC F2 TAC-CTG-AAC-AAA-CTG-CTG-GGT-AAC-GAA F3 CTG-GTG-GCA-ATC-AAC-CAG-TAC-TTC-CTG F4 CAT-GCA-CGT-ATG-TTC-AAA-AAC-TGG-GGT F5 CTG-AAA-CGT-CTG-AAC-GAT-GTG-GAA-TAC F6 CAT-GAA-AGC-ATC-GAT-GAA-ATG-AAA-CAT F7 GCA-GAT-CGT-TAC-ATC-GAA-CGT-ATC-CTG F8 TTC-CTG-GAA-GGT-CTG-CCG-AAC-CTG-CAG F9 GAT-CTG-GGT-AAA-CTG-AAC-ATC-GGT-GAA F10 GAT-GTG-GAA-GAA-ATG-CTG-CGT-AGC-GAT F11 CTG-GCA-CTG-GAA-CTG-GAT-GGT-GCA-AAA F12 AAC-CTG-CGT-GAA-GCA-ATC-GGT-TAC-GCA F13 GAT-AGC-GTG-CAT-GAT-TAC-GTG-AGC-CGT F14 GAT-ATG-ATG-ATC-GAA-ATC-CTG-CGT-GAT F15 GAA-GAA-GGT-CAT-ATC-GAT-TGG-CTG-GAA F16 ACC-GAA-CTG-GAT-CTG-ATC-CAG-AAA-ATG F17 GGT-CTG-CAG-AAC-TAC-CTG-CAG-GCA-CAG </div> <div> R1 CAG-TTT-GTT-CAG-GTA-GTT-GAT-CAC-TTT R2 GTT-GAT-TGC-CAC-CAG-TTC-GTT-ACC-CAG R3 GAA-CAT-ACG-TGC-ATG-CAG-GAA-GTA-CTG R4 GTT-CAG-ACG-TTT-CAG-ACC-CCA-GTT-TTT R5 ATC-GAT-GCT-TTC-ATG-GTA-TTC-CAC-ATC R6 GAT-GTA-ACG-ATC-TGC-ATG-TTT-CAT-TTC R7 CAG-ACC-TTC-CAG-GAA-CAG-GAT-ACC-TTC R8 CAG-TTT-ACC-CAG-ATC-CTG-CAG-GTT-CGG R9 CAT-TTC-TTC-CAC-ATC-TTC-ACC-GAT-GTT R10 CAG-TTC-CAG-TGC-CAG-ATC-GCT-ACC-CAG R11 TGC-TTC-ACG-CAG-GTT-TTT-TGC-ACC-ATC R12 ATC-ATG-CAG-GCT-ATC-TGC-GTA-ACC-GAT R13 TTC-GAT-CAT-CAT-ATC-ACG-GCT-CAC-GTA R14 GAT-ATG-ACC-TTC-TTC-ATC-ACG-CAG-GAT R15 CAG-ATC-CAG-TTC-GGT-TTC-CAG-CCA-ATC R16 GTA-GTT-CTG-CAG-ACC-CAT-TTT-CTG-GAT R17 ACC-TTC-TTC-ACG-GAT-CTG-TGC-CTG-CAG </div> </div>	
C <div> <div>sense 5'-3' GAC-GAC-GAC-AAG-ATG-AAA-GGT-GAT-ACC-AAA-GTG</div> <div>antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TCA-ACC-TTC-TTC-ACG-GAT</div> </div>	

Figure A1. 1 BFR gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning.

BFR-E	
A MKGDTKUIINYLNKLLGNELUAINQYFLHARMFKNWGLKRLNDUEYHESIDEMKHADRYIERILFLEGLPNLQDLGKLNIGEDUEEMLRSDLALELDGAKNLRERIGYADSVHDYVSRDMMIEILRDEEGHIDWLETELDLIQKMGKQNYLQAQIREEG	
B BFR-E(456bp) 5'-3' <div> <div> F1 ATG-AAA-GGT-GAT-ACC-AAA-GTG-ATC-AAC F2 TAC-CTG-AAC-AAA-CTG-CTG-GGT-AAC-GAA F3 CTG-GTG-GCA-ATC-AAC-CAG-TAC-TTC-CTG F4 CAT-GCA-CGT-ATG-TTC-AAA-AAC-TGG-GGT F5 CTG-AAA-CGT-CTG-AAC-GAT-GTG-GAA-TAC F6 CAT-GAA-AGC-ATC-GAT-GAA-ATG-AAA-CAT F7 GCA-GAT-CGT-TAC-ATC-GAA-CGT-ATC-CTG F8 TTC-CTG-GAA-GGT-CTG-CCG-AAC-CTG-CAG F9 GAT-CTG-GGT-AAA-CTG-AAC-ATC-GGT-GAA F10 GAT-GTG-GAA-GAA-ATG-CTG-CGT-AGC-GAT F11 CTG-GCA-CTG-GAA-CTG-GAT-GGT-GCA-AAA F12 AAC-CTG-CGT-GAA-GCA-ATC-GGT-TAC-GCA F13 GAT-AGC-GTG-CAT-GAT-TAC-GTG-AGC-CGT F14 GAT-ATG-ATG-ATC-GAA-ATC-CTG-CGT-GAT F15 GAA-GAA-GGT-CAT-ATC-GAT-TGG-CTG-GAA F16 ACC-GAA-CTG-GAT-CTG-ATC-CAG-AAA-ATG - GGT-GCA-CAG </div> <div> R1 CAG-TTT-GTT-CAG-GTA-GTT-GAT-CAC-TTT R2 GTT-GAT-TGC-CAC-CAG-TTC-GTT-ACC-CAG R3 GAA-CAT-ACG-TGC-ATG-CAG-GAA-GTA-CTG R4 GTT-CAG-ACG-TTT-CAG-ACC-CCA-GTT-TTT R5 ATC-GAT-GCT-TTC-ATG-GTA-TTC-CAC-ATC R6 GAT-GTA-ACG-ATC-TGC-ATG-TTT-CAT-TTC R7 CAG-ACC-TTC-CAG-GAA-CAG-GAT-ACC-TTC R8 CAG-TTT-ACC-CAG-ATC-CTG-CAG-GTT-CGG R9 CAT-TTC-TTC-CAC-ATC-TTC-ACC-GAT-GTT R10 CAG-TTC-CAG-TGC-CAG-ATC-GCT-ACC-CAG R11 TGC-TTC-ACG-CAG-GTT-TTT-TGC-ACC-ATC R12 ATC-ATG-CAG-GCT-ATC-TGC-GTA-ACC-GAT R13 TTC-GAT-CAT-CAT-ATC-ACG-GCT-CAC-GTA R14 GAT-ATG-ACC-TTC-TTC-ATC-ACG-CAG-GAT R15 CAG-ATC-CAG-TTC-GGT-TTC-CAG-CCA-ATC R16 ACC-TTC-TTC-ACG-GAT-CTG-TGC-ACC-CAT - TTT-CTG-GAT </div> </div>	
C <div> <div>sense 5'-3' GAC-GAC-GAC-AAG-ATG-AAA-GGT-GAT-ACC-AAA-GTG</div> <div>antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TCA-ACC-TTC-TTC-ACG-GAT</div> </div>	

Figure A1. 2 BFR-E gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning.

BFR+BC	
A MKGDTKVIINVLNKLGNELUAINQYFLHARMFKNWGLKRLNDUEYHESIDEMKHADRYIERILFLGGVALGTTQUINSKTPKLS VPLDIHNVEEMLRSDLALELDGAKNLREAIGYADSUHDYUSRDMHIEILRDEEGHIDWLETELDLIQKMGLQNYLQAQIREEG	
B BFR+BC(501bp) 5'-3' <div> <div> F1 ATG-AAA-GGT-GAT-ACC-AAA-GTG-ATC-AAC F2 TAC-CTG-AAC-AAA-CTG-CTG-GGT-AAC-GAA F3 CTG-GTG-GCA-ATC-AAC-CAG-TAC-TTC-CTG F4 CAT-GCA-CGT-ATG-TTC-AAA-AAC-TGG-GGT F5 CTG-AAA-CGT-CTG-AAC-GAT-GTG-GAA-TAC F6 CAT-GAA-AGC-ATC-GAT-GAA-ATG-AAA-CAT F7 GCA-GAT-CGT-TAC-ATC-GAA-CGT-ATC-CTG F8 TTC-CTG-GGT-GGT-GTG-GCA-CTG-GGT-ACC F9 ACC-CAG-GTG-ATC-AAC-AGC-AAA-ACC-CCG F10 CTG-AAA-AGC-TAC-CCG-CTG-GAT-ATC-CAT F11 AAC-GTG-GAA-GAA-ATG-CTG-CGT-AGC-GAT F12 CTG-GCA-CTG-GAA-CTG-GAT-GGT-GCA-AAA F13 AAC-CTG-CGT-GAA-GCA-ATC-GGT-TAC-GCA F14 GAT-AGC-GTG-CAT-GAT-TAC-GTG-AGC-CGT F15 GAT-ATG-ATG-ATC-GAA-ATC-CTG-CGT-GAT F16 GAA-GAA-GGT-CAT-ATC-GAT-TGG-CTG-GAA F17 ACC-GAA-CTG-GAT-CTG-ATC-CAG-AAA-ATG F18 GGT-CTG-CAG-AAC-TAC-CTG-CAG-GCA-CAG </div> <div> R1 CAG-TTT-GTT-CAG-GTA-GTT-GAT-CAC-TTT R2 GTT-GAT-TGC-CAC-CAG-TTC-GTT-ACC-CAG R3 GAA-CAT-ACG-TGC-ATG-CAG-GAA-GTA-CTG R4 GTT-CAG-ACG-TTT-CAG-ACC-CCA-GTT-TTT R5 ATC-GAT-GCT-TTC-ATG-GTA-TTC-CAC-ATC R6 GAT-GTA-ACG-ATC-TGC-ATG-TTT-CAT-TTC R7 CAC-ACC-ACC-CAG-GAA-CAG-GAT-ACG-TTC R8 GTT-GAT-CAC-CTG-GGT-GGT-ACC-CAG-TGC R9 CGG-GTA-GCT-TTT-CAG-CGG-GGT-TTT-GCT R10 CAT-TTC-TTC-CAC-GTT-ATG-GAT-ATC-CAG R11 CAG-TTC-CAG-TGC-CAG-ATC-GCT-ACC-CAG R12 TGC-TTC-ACG-CAG-GTT-TTT-TGC-ACC-ATC R13 ATC-ATG-CAC-GCT-ATC-TGC-GTA-ACC-GAT R14 TTC-GAT-CAT-CAT-ATC-ACG-GCT-CAC-GTA R15 GAT-ATG-ACC-TTC-TTC-ATC-ACG-CAG-GAT R16 CAG-ATC-CAG-TTC-GGT-TTC-CAG-CCA-ATC R17 GTA-GTT-CTG-CAG-ACC-CAT-TTT-CTG-GAT R18 ACC-TTC-TTC-ACG-GAT-CTG-TGC-CTG-CAG </div> </div>	
C <div> sense 5'-3' GAC-GAC-GAC-AAG-ATG-AAA-GGT-GAT-ACC-AAA-GTG antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TCA-ACC-TTC-TTC-ACG-GAT </div>	

Figure A1. 3 BFR+BC gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning

BFR+BC-E	
A MKGDTKVIINVLNKLGNELUAINQYFLHARMFKNWGLKRLNDUEYHESIDEMKHADRYIERILFLGGVALGTTQUINSKTP LKSYPLDIHNVEEMLRSDLALELDGAKNLREAIGYADSUHDYUSRDMHIEILRDEEGHIDWLETELDLIQKMGAQIREEG	
B BFR+BC-E(483bp) 5'-3' <div> <div> F1 ATG-AAA-GGT-GAT-ACC-AAA-GTG-ATC-AAC F2 TAC-CTG-AAC-AAA-CTG-CTG-GGT-AAC-GAA F3 CTG-GTG-GCA-ATC-AAC-CAG-TAC-TTC-CTG F4 CAT-GCA-CGT-ATG-TTC-AAA-AAC-TGG-GGT F5 CTG-AAA-CGT-CTG-AAC-GAT-GTG-GAA-TAC F6 CAT-GAA-AGC-ATC-GAT-GAA-ATG-AAA-CAT F7 GCA-GAT-CGT-TAC-ATC-GAA-CGT-ATC-CTG F8 TTC-CTG-GGT-GGT-GTG-GCA-CTG-GGT-ACC F9 ACC-CAG-GTG-ATC-AAC-AGC-AAA-ACC-CCG F10 CTG-AAA-AGC-TAC-CCG-CTG-GAT-ATC-CAT F11 AAC-GTG-GAA-GAA-ATG-CTG-CGT-AGC-GAT F12 CTG-GCA-CTG-GAA-CTG-GAT-GGT-GCA-AAA F13 AAC-CTG-CGT-GAA-GCA-ATC-GGT-TAC-GCA F14 GAT-AGC-GTG-CAT-GAT-TAC-GTG-AGC-CGT F15 GAT-ATG-ATG-ATC-GAA-ATC-CTG-CGT-GAT F16 GAA-GAA-GGT-CAT-ATC-GAT-TGG-CTG-GAA F17 ACC-GAA-CTG-GAT-CTG-ATC-CAG-AAA-ATG GGT-GCA-CAG </div> <div> R1 CAG-TTT-GTT-CAG-GTA-GTT-GAT-CAC-TTT R2 GTT-GAT-TGC-CAC-CAG-TTC-GTT-ACC-CAG R3 GAA-CAT-ACG-TGC-ATG-CAG-GAA-GTA-CTG R4 GTT-CAG-ACG-TTT-CAG-ACC-CCA-GTT-TTT R5 ATC-GAT-GCT-TTC-ATG-GTA-TTC-CAC-ATC R6 GAT-GTA-ACG-ATC-TGC-ATG-TTT-CAT-TTC R7 CAC-ACC-ACC-CAG-GAA-CAG-GAT-ACG-TTC R8 GTT-GAT-CAC-CTG-GGT-GGT-ACC-CAG-TGC R9 CGG-GTA-GCT-TTT-CAG-CGG-GGT-TTT-GCT R10 CAT-TTC-TTC-CAC-GTT-ATG-GAT-ATC-CAG R11 CAG-TTC-CAG-TGC-CAG-ATC-GCT-ACC-CAG R12 TGC-TTC-ACG-CAG-GTT-TTT-TGC-ACC-ATC R13 ATC-ATG-CAC-GCT-ATC-TGC-GTA-ACC-GAT R14 TTC-GAT-CAT-CAT-ATC-ACG-GCT-CAC-GTA R15 GAT-ATG-ACC-TTC-TTC-ATC-ACG-CAG-GAT R16 CAG-ATC-CAG-TTC-GGT-TTC-CAG-CCA-ATC R17 ACC-TTC-TTC-ACG-GAT-CTG-TGC-ACC-CAT -TTT-CTG-GAT </div> </div>	
C <div> sense 5'-3' GAC-GAC-GAC-AAG-ATG-AAA-GGT-GAT-ACC-AAA-GTG antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TCA-ACC-TTC-TTC-ACG-GAT </div>	

Figure A1. 4 BFR+BC-E gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning.

DPS	
A MSTAKLUKSKATNLLYTRNDUSDSEKKATUELLNRQUIQFIDLSLITKQAHWNMRGANFIAVHEMLDGFRTALIDHLDTMAERA UQLGGVALGTTQVINSKTPKSYPLDIHNUQDHLKELADRYAIVANDURKAIGAEKDDDTADILTAASRDLDKFLWFIECNIE	
B DPS(501bp) 5' -3' <div> <div> F1 ATG-AGC-ACC-GCA-AAA-CTG-GTG-AAA-AGC F2 AAA-GCA-ACC-AAC-CTG-CTG-TAC-ACC-CGT F3 AAC-GAT-GTG-AGC-GAT-AGC-GAA-AAA-AAA F4 GCA-ACC-GTG-GAA-CTG-CTG-AAC-CGT-CAG F5 GTG-ATC-CAG-TTC-ATC-GAT-CTG-AGC-CTG F6 ATC-ACC-AAA-CAG-GCA-CAT-TGG-AAC-ATG F7 CGT-GGT-GCA-AAC-TTC-ATC-GCA-GTG-CAT F8 GAA-ATG-CTG-GAT-GGT-TTC-CGT-ACC-GCA F9 CTG-ATC-GAT-CAT-CTG-GAT-ACC-ATG-GCA F10 GAA-CGT-GCA-GTG-CAG-CTG-GGT-GGT-GTG F11 GCA-CTG-GGT-ACC-ACC-CAG-GTG-ATC-AAC F12 AGC-AAA-ACC-CCG-CTG-AAA-AGC-TAC-CCG F13 CTG-GAT-ATC-CAT-AAC-GTG-CAG-GAT-CAT F14 CTG-AAA-GAA-CTG-GCA-GAT-CGT-TAC-GCA F15 ATC-GTG-GCA-AAC-GAT-GTG-CGT-AAA-GCA F16 ATC-GGT-GAA-GCA-AAA-GAT-GAT-GAT-ACC F17 GCA-GAT-ATC-CTG-ACC-GCA-GCA-AGC-CGT F18 GAT-CTG-GAT-AAA-TTC-CTG-TGG-TTC-ATC </div> <div> R1 CAG-GTT-GGT-TGC-TTT-GCT-TTT-CAC-CAG R2 ATC-GCT-CAC-ATC-GTT-ACG-GGT-GTA-CAG R3 CAG-TTC-CAC-GGT-TGC-TTT-TTT-TTC-GCT R4 GAT-GAA-CTG-GAT-CAC-CTG-ACG-GTT-CAG R5 TGC-CTG-TTT-GGT-GAT-CAG-GCT-CAG-ATC R6 GAA-GTT-TGC-ACC-ACG-CAT-GTT-CCA-ATG R7 ACC-ATC-CAG-CAT-TTC-ATG-CAC-TGC-GAT R8 CAG-ATG-ATC-GAT-CAG-TGC-GGT-ACG-GAA R9 CTG-CAC-TGC-ACG-TTC-TGC-CAT-GGT-ATC R10 GGT-GGT-ACC-CAG-TGC-CAC-ACC-ACC-CAG R11 CAG-CGG-GGT-TTT-GCT-GTT-GAT-CAC-CTG R12 GTT-ATG-GAT-ATC-CAG-CGG-GTA-GCT-TTT R13 TGC-CAG-TTC-TTT-CAG-ATG-ATC-CTG-CAC R14 ATC-GTT-TGC-CAC-GAT-TGC-GTA-ACG-ATC R15 TTT-TGC-TTC-ACC-GAT-TGC-TTT-ACG-CAC R16 GGT-CAG-GAT-ATC-TGC-GGT-ATC-ATC-ATC R17 GAA-TTT-ATC-CAG-ATC-ACG-GCT-TGC-TGC R18 TTC-GAT-GTT-GCA-TTC-GAT-GAA-CCA-CAG </div> </div>	
C sense 5'-3' GAC-GAC-GAC-AAG-ATG-AGC-ACC-GCA-AAA-CTG-GTG antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TAT-TCG-ATG-TTG-CAT-TCG	

Figure A1. 5 DPS gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning.

DPS-BC	
A MSTAKLUKSKATNLLYTRNDUSDSEKKATUELLNRQUIQFIDLSLITKQAHWNMRGANFIAVHEMLDGFRTALIDHLDTM AERAUQLGGVALGTTPLKSYPLDIHNUQDHLKELADRYAIVANDURKAIGAEKDDDTADILTAASRDLDKFLWFIECNIE	
B DPS-BC(480bp) 5' -3' <div> <div> F1 ATG-AGC-ACC-GCA-AAA-CTG-GTG-AAA-AGC F2 AAA-GCA-ACC-AAC-CTG-CTG-TAC-ACC-CGT F3 AAC-GAT-GTG-AGC-GAT-AGC-GAA-AAA-AAA F4 GCA-ACC-GTG-GAA-CTG-CTG-AAC-CGT-CAG F5 GTG-ATC-CAG-TTC-ATC-GAT-CTG-AGC-CTG F6 ATC-ACC-AAA-CAG-GCA-CAT-TGG-AAC-ATG F7 CGT-GGT-GCA-AAC-TTC-ATC-GCA-GTG-CAT F8 GAA-ATG-CTG-GAT-GGT-TTC-CGT-ACC-GCA F9 CTG-ATC-GAT-CAT-CTG-GAT-ACC-ATG-GCA F10 GAA-CGT-GCA-GTG-CAG-CTG-GGT-GGT-GTG F11 GCA-CTG-GGT-ACC-ACC-CCG-CTG-AAA-AGC -TAC-CCG F12 CTG-GAT-ATC-CAT-AAC-GTG-CAG-GAT-CAT F13 CTG-AAA-GAA-CTG-GCA-GAT-CGT-TAC-GCA F14 ATC-GTG-GCA-AAC-GAT-GTG-CGT-AAA-GCA F15 ATC-GGT-GAA-GCA-AAA-GAT-GAT-GAT-ACC F16 GCA-GAT-ATC-CTG-ACC-GCA-GCA-AGC-CGT F17 GAT-CTG-GAT-AAA-TTC-CTG-TGG-TTC-ATC </div> <div> R1 CAG-GTT-GGT-TGC-TTT-GCT-TTT-CAC-CAG R2 ATC-GCT-CAC-ATC-GTT-ACG-GGT-GTA-CAG R3 CAG-TTC-CAC-GGT-TGC-TTT-TTT-TTC-GCT R4 GAT-GAA-CTG-GAT-CAC-CTG-ACG-GTT-CAG R5 TGC-CTG-TTT-GGT-GAT-CAG-GCT-CAG-ATC R6 GAA-GTT-TGC-ACC-ACG-CAT-GTT-CCA-ATG R7 ACC-ATC-CAG-CAT-TTC-ATG-CAC-TGC-GAT R8 CAG-ATG-ATC-GAT-CAG-TGC-GGT-ACG-GAA R9 CTG-CAC-TGC-ACG-TTC-TGC-CAT-GGT-ATC R10 CAG-CGG-GGT-GGT-ACC-CAG-TGC-CAC-ACC -ACC-CAG R11 GTT-ATG-GAT-ATC-CAG-CGG-GTA-GCT-TTT R12 TGC-CAG-TTC-TTT-CAG-ATG-ATC-CTG-CAC R13 ATC-GTT-TGC-CAC-GAT-TGC-GTA-ACG-ATC R14 TTT-TGC-TTC-ACC-GAT-TGC-TTT-ACG-CAC R15 GGT-CAG-GAT-ATC-TGC-GGT-ATC-ATC-ATC R16 GAA-TTT-ATC-CAG-ATC-ACG-GCT-TGC-TGC R17 TTC-GAT-GTT-GCA-TTC-GAT-GAA-CCA-CAG </div> </div>	
C sense 5'-3' GAC-GAC-GAC-AAG-ATG-AGC-ACC-GCA-AAA-CTG-GTG antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TAT-TCG-ATG-TTG-CAT-TCG	

Figure A1. 6 DPS-BC gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning.

DPS+E	
A MSTAKLUKSKATNLLYTRNDUSDSEKKATUELLNRQUIQFIDLSLITKQAHWNMRGANFIAUHEMLDGFRTALIDHLDTMA ERAUQLGGVALGTTPLKSYPLDIHNNUQDHLKELADRYAIVANDURKAIGAEKDDDTADILTAASRDLDKFLWFI ECNLDLIQKMGLOQNYLQAQIREEG	
B DPS+E (558bp) 5'-3' <div> <div> F1 ATG-AGC-ACC-GCA-AAA-CTG-GTG-AAA-AGC F2 AAA-GCA-ACC-AAC-CTG-CTG-TAC-ACC-CGT F3 AAC-GAT-GTG-AGC-GAT-AGC-GAA-AAA-AAA F4 GCA-ACC-GTG-GAA-CTG-CTG-AAC-CGT-CAG F5 GTG-ATC-CAG-TTC-ATC-GAT-CTG-AGC-CTG F6 ATC-ACC-AAA-CAG-GCA-CAT-TGG-AAC-ATG F7 CGT-GGT-GCA-AAC-TTC-ATC-GCA-GTG-CAT F8 GAA-ATG-CTG-GAT-GGT-TTC-CGT-ACC-GCA F9 CTG-ATC-GAT-CAT-CTG-GAT-ACC-ATG-GCA F10 GAA-CGT-GCA-GTG-CAG-CTG-GGT-GGT-GTG F11 GCA-CTG-GGT-ACC-ACC-CAG-GTG-ATC-AAC F12 AGC-AAA-ACC-CCG-CTG-AAA-AGC-TAC-CCG F13 CTG-GAT-ATC-CAT-AAC-GTG-CAG-GAT-CAT F14 CTC-AAA-GAA-CTG-GCA-GAT-CGT-TAC-GCA F15 ATC-GTG-GCA-AAC-GAT-GTG-CGT-AAA-GCA F16 ATC-GGT-GAA-GCA-AAA-GAT-GAT-GAT-ACC F17 GCA-GAT-ATC-CTG-ACC-GCA-GCA-AGC-CGT F18 GAT-CTG-GAT-AAA-TTC-CTG-TGG-TTC-ATC F19 GAA-TGC-AAC-CTG-GAT-CTG-ATC-CAG-AAA F20 ATG-GGT-CTG-CAG-AAC-TAC-CTG-CAG-GCA </div> <div> R1 CAG-GTT-GGT-TGC-TTT-GCT-TTT-CAC-CAG R2 ATC-GCT-CAC-ATC-GTT-ACG-GGT-GTA-CAG R3 CAG-TTC-CAC-GGT-TGC-TTT-TTT-TTC-GCT R4 GAT-GAA-CTG-GAT-CAC-CTG-ACG-GTT-CAG R5 TGC-CTG-TTT-GGT-GAT-CAG-GCT-CAG-ATC R6 GAA-GTT-TGC-ACC-ACG-CAT-GTT-CCA-ATG R7 ACC-ATC-CAG-CAT-TTC-ATG-CAC-TGC-GAT R8 CAG-ATG-ATC-GAT-CAG-TGC-GGT-ACG-GAA R9 CTG-CAC-TGC-ACG-TTC-TGC-CAT-GGT-ATC R10 GGT-GGT-ACC-CAG-TGC-CAC-ACC-ACC-CAG R11 CAG-CGG-GGT-TTT-GCT-GTT-GAT-CAG-CTG R12 GTT-ATG-GAT-ATC-CAG-CGG-GTA-GCT-TTT R13 TGC-CAG-TTC-TTT-CAG-ATG-ATC-CTG-CAC R14 ATC-GTT-TGC-CAC-GAT-TGC-GTT-ACG-ATC R15 TTT-TGC-TTC-ACC-GAT-TGC-TTT-ACG-CAC R16 GGT-CAG-GAT-ATC-TGC-GGT-ATC-ATC-ATC R17 GAA-TTT-ATC-CAG-ATC-ACG-GCT-TGC-TGC R18 ATC-CAG-GTT-GCA-TTC-GAT-GAA-CCA-CAG R19 GTT-CTG-CAG-ACC-CAT-TTT-CTG-GAT-CAG R20 ACC-TTC-TTC-ACG-GAT-CTG-TGC-CTG-CAG -GTA </div> </div>	
C <div> sense 5'-3' GAC-GAC-GAC-AAG-ATG-AGC-ACC-GCA-AAA-CTG-GTG antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TCA-ACC-TTC-TTC-ACG-GAT </div>	

Figure A1. 7 DPS+E gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning.

DPS-BC+E	
A MSTAKLUKSKATNLLYTRNDUSDSEKKATUELLNRQUIQFIDLSLITKQAHWNMRGANFIAUHEMLDGFRTALIDHLDTMAERA UQLGGVALGTTPLKSYPLDIHNNUQDHLKELADRYAIVANDURKAIGAEKDDDTADILTAASRDLDKFLWFI ECNLDLIQKMGLOQNYLQAQIREEG	
B DPS-BC+E (537bp) 5'-3' <div> <div> F1 ATG-AGC-ACC-GCA-AAA-CTG-GTG-AAA-AGC F2 AAA-GCA-ACC-AAC-CTG-CTG-TAC-ACC-CGT F3 AAC-GAT-GTG-AGC-GAT-AGC-GAA-AAA-AAA F4 GCA-ACC-GTG-GAA-CTG-CTG-AAC-CGT-CAG F5 GTG-ATC-CAG-TTC-ATC-GAT-CTG-AGC-CTG F6 ATC-ACC-AAA-CAG-GCA-CAT-TGG-AAC-ATG F7 CGT-GGT-GCA-AAC-TTC-ATC-GCA-GTG-CAT F8 GAA-ATG-CTG-GAT-GGT-TTC-CGT-ACC-GCA F9 CTG-ATC-GAT-CAT-CTG-GAT-ACC-ATG-GCA F10 GAA-CGT-GCA-GTG-CAG-CTG-GGT-GGT-GTG F11 GCA-CTG-GGT-ACC-ACC-CCG-CTG-AAA-AGC -TAC-CCG F12 CTG-GAT-ATC-CAT-AAC-GTG-CAG-GAT-CAT F13 CTG-AAA-GAA-CTG-GCA-GAT-CGT-TAC-GCA F14 ATC-GTG-GCA-AAC-GAT-GTG-CGT-AAA-GCA F15 ATC-GGT-GAA-GCA-AAA-GAT-GAT-GAT-ACC F16 GCA-GAT-ATC-CTG-ACC-GCA-GCA-AGC-CGT F17 GAT-CTG-GAT-AAA-TTC-CTG-TGG-TTC-ATC F18 GAA-TGC-AAC-CTG-GAT-CTG-ATC-CAG-AAA F19 ATG-GGT-CTG-CAG-AAC-TAC-CTG-CAG-GCA </div> <div> R1 CAG-GTT-GGT-TGC-TTT-GCT-TTT-CAC-CAG R2 ATC-GCT-CAC-ATC-GTT-ACG-GGT-GTA-CAG R3 CAG-TTC-CAC-GGT-TGC-TTT-TTT-TTC-GCT R4 GAT-GAA-CTG-GAT-CAC-CTG-ACG-GTT-CAG R5 TGC-CTG-TTT-GGT-GAT-CAG-GCT-CAG-ATC R6 GAA-GTT-TGC-ACC-ACG-CAT-GTT-CCA-ATG R7 ACC-ATC-CAG-CAT-TTC-ATG-CAC-TGC-GAT R8 CAG-ATG-ATC-GAT-CAG-TGC-GGT-ACG-GAA R9 CTG-CAC-TGC-ACG-TTC-TGC-CAT-GGT-ATC R10 CAG-CGG-GGT-GGT-ACC-CAG-TGC-CAC-ACC -ACC-CAG R11 GTT-ATG-GAT-ATC-CAG-CGG-GTA-GCT-TTT R12 TGC-CAG-TTC-TTT-CAG-ATG-ATC-CTG-CAC R13 ATC-GTT-TGC-CAC-GAT-TGC-GTA-ACG-ATC R14 TTT-TGC-TTC-ACC-GAT-TGC-TTT-ACG-CAC R15 GGT-CAG-GAT-ATC-TGC-GGT-ATC-ATC-ATC R16 GAA-TTT-ATC-CAG-ATC-ACG-GCT-TGC-TGC R17 ATC-CAG-GTT-GCA-TTC-GAT-GAA-CCA-CAG R18 GTT-CTG-CAG-ACC-CAT-TTT-CTG-GAT-CAG R19 ACC-TTC-TTC-ACG-GAT-CTG-TGC-CTG-CAG -GTA </div> </div>	
C <div> sense 5'-3' GAC-GAC-GAC-AAG-ATG-AGC-ACC-GCA-AAA-CTG-GTG antisense 5'-3' GAG-GAG-AAG-CCC-GGT-TCA-ACC-TTC-TTC-ACG-GAT </div>	

Figure A1. 8 DPS-BC+E gene assembly and amplification: A) The amino acid sequence of the protein. B) The DNA sequence of the forward and reverse oligonucleotides used in gene assembly. C) The DNA sequence of the sense and antisense primers used in the extension PCR to provide LIC termini for cloning.

Appendix 2 Some interactions in BFR and DPS

Figure 2A. 1 View along the two-fold axis of BFR subunits in ribbon diagrams showing that the residues involved in inter-subunit hydrogen bonds are Gln72, Gly75 and Leu77. The figure is generated by using UCSF Chimera.

Figure 2A. 2 View along the two-fold axis of BFR subunits from the inner cavity indicating that a salt bridge is formed between Arg30, Asp56 and Glu60. The figure is generated by using UCSF Chimera.

Figure 2A. 3 View along the three-fold axis of BFR subunits representing an inter-subunit salt bridge between Arg 61 and Glu128. The figure is generated by using UCSF Chimera.

Figure 2A. 4 View along three-fold axis of BFR indicating an intra-subunit salt bridge between Arg102 and Glu128. The figure is generated by using UCSF Chimera.

Figure 2A. 5 View along the two-fold axis of DPS subunits from the inner cavity indicating that salt bridges are formed between Arg70, Lys48 and Asp78. The figure is generated by using UCSF Chimera.

Figure 2A. 6 View along the three-fold axis of DPS subunits in ribbon diagrams showing that the residues involved in inter-subunit salt bridge between Arg18 and Asp123. The figure is generated by using UCSF Chimera.

Figure 2A. 7 View along the three-fold axis of DPS subunits representing an inter-subunit salt bridge between Asp20 and Arg133. The figure is generated by using UCSF Chimera

Figure 2A. 8 View along three-fold axis of BFR indicating an inter-subunit salt bridge between Arg83 and Asp156. The figure is generated by using UCSF Chimera.

Appendix 3 Analysis of all inter-subunit interfaces in BFR.

All analysis results of BFR inter-subunit interfaces are obtained from the online program PDBePISA (protein interfaces, surfaces and assemblies) with auto processing mode which can be assess at http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html.

Figure 3A. 1 Each subunit in BFR makes contacts with six neighboring subunits. For example, (in the central hexagon) subunit A (dark grey) interacts with six subunits B, H, W, X, M and L (light grey) in different symmetry related interfaces (highlighted with red circles.) On the upper left corner, the residues in subunit A that are involved in the inter-subunit interactions with subunit B around the two-fold symmetry axis are colored red. On the left, the residues in subunit A that are involved in the inter-subunit interactions with subunit H around the three-fold symmetry axis are colored red. On the bottom left corner, the residues in subunit A that are involved in the inter-subunit interactions with subunit W around the three-fold symmetry axis are colored red. On the bottom right corner, the residues in subunit A that are involved in the inter-subunit interactions with subunit X around the four-fold symmetry axis are colored red. On the right, the residues in subunit A that are involved in the inter-subunit interactions with subunit M around the four-fold symmetry axis are colored red. On the upper right corner, the residues in subunit A that are involved in the inter-subunit interactions with subunit L around the four-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the subunit numbers are taken from PDB file of 1bfr.

Figure 3A. 2 The residues in subunit A which are involved in the inter-subunit interactions with subunit B around the two-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1bfr.

Table 3A. 1 The residues in subunit A which are involved in the inter-subunit interactions with subunit B around the two-fold axis of BFR are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
19	LEU	Hydrogen bond and salt bridge
23	ASN	Hydrogen bond
26	PHE	
27	LEU	
30	ARG	Hydrogen bond and salt bridge
31	MET	
33	LYS	Salt bridge
34	ASN	
52	MET	
56	ASP	Hydrogen bond and salt bridge
59	ILE	
60	GLU	Hydrogen bond and salt bridge
63	LEU	Hydrogen bond
64	PHE	
68	LEU	
69	PRO	
71	LEU	Hydrogen bond
72	GLN	Hydrogen bond
73	ASP	
74	LEU	
75	GLY	Hydrogen bond
76	LYS	
77	LEU	Hydrogen bond
79	ILE

Figure 3A. 3 The residues in subunit A which are involved in the inter-subunit interactions with subunit H around the three-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1bfr.

Table 3A. 2 The residues in subunit A which are involved in the inter-subunit interactions with subunit H around the three-fold axis of BFR are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
1	MET	Hydrogen bond and salt bridge
61	ARG	Hydrogen bond and salt bridge
64	PHE	
66	GLU	
109	ASP	
111	VAL	
112	HIS	Hydrogen bond
113	ASP	
114	TYR	
115	VAL	
117	ARG	
118	ASP	Hydrogen bond and salt bridge
122	GLU

Figure 3A. 4 The residues in subunit A which are involved in the inter-subunit interactions with subunit W around the three-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1bfr.

Table 3A. 3 The residues in subunit A which are involved in the inter-subunit interactions with subunit W around the three-fold axis of BFR are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
95	LEU	Hydrogen bond
98	ALA	
102	ARG	
103	GLU	
105	ILE	
106	GLY	
109	ASP	
117	ARG	
118	ASP	
121	ILE	
124	LEU	Hydrogen bond and salt bridge
125	ARG	
128	GLU	Hydrogen bond and salt bridge
131	ILE	Hydrogen bond and salt bridge
132	ASP	
135	GLU

Figure 3A. 5 The residues in subunit A which are involved in the inter-subunit interactions with subunit X around the four-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1bfr.

Table 3A. 4 The residues in subunit A which are involved in the inter-subunit interactions with subunit H around the four-fold axis of BFR are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
39	ARG	
132	ASP	
136	THR	Hydrogen bond
139	ASP	
140	LEU	
142	GLN	
143	LYS	Hydrogen bond
144	MET	
145	GLY	
148	ASN	Hydrogen bond
149	TYR	
151	GLN	Hydrogen bond
152	ALA	Hydrogen bond
153	GLN	
155	ARG	Hydrogen bond
156	GLU	
157	GLU	Hydrogen bond and salt bridge
158	GLY

Figure 3A. 6 The residues in subunit A which are involved in the inter-subunit interactions with subunit M around the four-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1bfr.

Table 3A. 5 The residues in subunit A which are involved in the inter-subunit interactions with subunit M around the four-fold axis of BFR are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
148	ASN	
151	GLN

Figure 3A. 7 The residues in subunit A which are involved in the inter-subunit interactions with subunit L around the four-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1bfr.

Table 3A. 6 The residues in subunit A which are involved in the inter-subunit interactions with subunit L around the four-fold axis of BFR are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
34	ASN	Hydrogen bond
35	TRP	
36	GLY	
37	LEU	
38	LYS	Hydrogen bond and salt bridge
147	GLN	Hydrogen bond
148	ASN	Hydrogen bond
150	LEU	
151	GLN	Hydrogen bond
154	ILE	
155	ARG	Hydrogen bond and salt bridge
156	GLU	

Appendix 4 Analysis of all inter-subunit interfaces in DPS.

All analysis results of DPS inter-subunit interfaces are obtained from the online program PDBePISA (protein interfaces, surfaces and assemblies) with auto processing mode which can be assessed at http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html.

Figure 4A. 1 Each subunit in DPS makes contacts with five neighboring subunits. For example, (in the central pentagon) subunit A (colored in dark grey) in DPS interacts with five subunits C, E, I, K, and H (colored in light grey) in different symmetry related interfaces which are highlighted with red circles. On the upper left corner, the residues in subunit A that are involved in the inter-subunit interactions with subunit C around the two-fold symmetry axis are colored red. On the left, the residues in subunit A that are involved in the inter-subunit interactions with subunit E around the ferritin-like three-fold symmetry axis are colored red. On the bottom, the residues in subunit A that are involved in the inter-subunit interactions with subunit I around the ferritin-like three-fold symmetry axis are colored red. On the right, the residues in subunit A that are involved in the inter-subunit interactions with subunit K around the Dps-like three-fold symmetry axis are colored red. On the upper right corner, the residues in subunit A that are involved in the inter-subunit interactions with subunit H around the Dps-like three-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the subunit numbers are taken from PDB file of 1dps.

Figure 4A. 2 The residues in subunit A which are involved in the inter-subunit interactions with subunit C around the two-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1dps.

Table 4A. 1 The residues in subunit A which are involved in the inter-subunit interactions with subunit C around the two-fold axis of DPS are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
15	LEU	
16	TYR	
17	THR	
37	VAL	
41	ILE	
42	ASP	Hydrogen bond
45	LEU	
46	ILE	
48	LYS	Hydrogen bond and salt bridge
49	GLN	Hydrogen bond
51	HIS	
52	TRP	
53	ASN	Hydrogen bond
55	ARG	
63	HIS	
67	ASP	
70	ARG	Hydrogen bond and salt bridge
74	ILE	
75	ASP	
77	LEU	
78	ASP	Hydrogen bond and salt bridge
81	ALA	
82	GLU	
85	VAL	
88	GLY	
89	GLY	
90	VAL	
91	ALA	Hydrogen bond
92	LEU	

93	GLY	Hydrogen bond
94	THR	Hydrogen bond
95	THR	Hydrogen bond
96	GLN	
97	VAL	
98	ILE	
99	ASN	
102	THR	
104	LEU	
106	SER	
107	TYR	
109	LEU	
110	ASP	
112	HIS

Figure 4A. 3 The residues in subunit A which are involved in the inter-subunit interactions with subunit E around the ferritin-like three-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1dps.

Table 4A. 2 The residues in subunit A which are involved in the inter-subunit interactions with subunit E around the ferritin-like three-fold axis of DPS are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
16	TYR	
17	THR	Hydrogen bond
18	ARG	Hydrogen bond and salt bridge
19	ASN	
20	ASP	Hydrogen bond and salt bridge
21	VAL	
79	THR	
82	GLU	
83	ARG	Hydrogen bond and salt bridge
85	VAL	
86	GLN	
141	ASP	
142	ASP	
143	ASP	Hydrogen bond
146	ASP	Hydrogen bond and salt bridge
147	ILE	
150	ALA

Figure 4A. 4 The residues in subunit A which are involved in the inter-subunit interactions with subunit I around the ferritin-like three-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1dps.

Table 4A. 3 The residues in subunit A which are involved in the inter-subunit interactions with subunit I around the ferritin-like three-fold axis of DPS are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
119	LYS	
122	ALA	
123	ASP	Hydrogen bond and salt bridge
126	ALA	
130	ASN	Hydrogen bond
133	ARG	Hydrogen bond and salt bridge
134	LYS	Hydrogen bond and salt bridge
136	ILE	
137	GLY	
146	ASP	
149	THR	
150	ALA	
152	SER	Hydrogen bond
153	ARG	Hydrogen bond and salt bridge
156	ASP	Hydrogen bond and salt bridge
157	LYS	
159	LEU	
160	TRP	
161	PHE	
163	GLU	Hydrogen bond and salt bridge
166	ILE

Figure 4A. 5 The residues in subunit A which are involved in the inter-subunit interactions with subunit K around the Dps-like three-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1dps.

Table 4A. 4 The residues in subunit A which are involved in the inter-subunit interactions with subunit K around the Dps-like three-fold axis of DPS are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
57	ALA	
58	ASN	
61	ALA	
62	VAL	
65	MET	
160	TRP	
161	PHE	
163	GLU	
154	CYS	Hydrogen bond
165	ASN	Hydrogen bond
166	ILE	
167	GLU

Figure 4A. 6 The residues in subunit A which are involved in the inter-subunit interactions with subunit H around the Dps-like three-fold symmetry axis are colored red. The figure is generated using UCSF Chimera and the PDB ID of the protein is 1dps.

Table 4A. 5 The residues in subunit A which are involved in the inter-subunit interactions with subunit H around the Dps-like three-fold axis of DPS are listed as follows.

Residue number	Residue	Hydrogen bond and salt bridge
52	TRP	Hydrogen bond
54	MET	
55	ARG	
56	GLY	
57	ALA	Hydrogen bond
58	ASN	
59	PHE	
60	ILE	
61	ALA	Hydrogen bond
63	HIS	
64	GLU	