Mutagenesis studies to determine the key interfacial amino acid residues that govern the self-assembly of the ferritin protein cages

Zhang, Yu

2011

Zhang, Y. (2011). Mutagenesis studies to determine the key interfacial amino acid residues that govern the self-assembly of the ferritin protein cages. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/48015>

<https://doi.org/10.32657/10356/48015>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**Mutagenesis Studies to Determine the Key Interfacial Amino
Acid Residues that Govern the Self-assembly of the Ferritin
Protein Cages**

ZHANG Yu

**Division of Chemistry and Biological Chemistry
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

Foremost, I would like to address my sincerely appreciation to my supervisor Brendan Orner, one of the most respected persons in my life, for his four years of guidance, valuable advices and help. Brendan's rigorous attitude of scholarship, patience, supervision and encouragement impressed on my memory and set a model in my future career. He is modest, sincere and generous; he respects all his students, and always tries to help when they are in difficulty. I greatly appreciate the opportunity he provided for me to conduct research in his warm, united and positive group.

I would like to thank all the group members that I have worked with in the past four years for the help they have offered and the joyfulness they have brought. I would like to give special thanks to Rongli Fan, Maziar Soleymani Ardejani for their valuable discussions and help, and also Thomas Andrew Cornell, Limin Yu, Yueqin Fang, Xiaoming Sun. I'd also like to thank Fahriye Ceyda Dudak for the excellent work and helpful discussion in DSC experiments. I would also like to express my gratitude to all the undergraduates in CBC who have contributed in the project: Jing Fu, Emmiline Ang Xiao Wei, Wenlin Zhang, Sze Yin Chee, Hui Hian Teo, Gwenda Wei Shan Teo, Siti Raudah Binte Muhamed Lazim, and two hard-working and talented girls from Singapore Ngee Ann Polytechnic, Nur Fazlina Binte Aris and Pang Yi Teng. I would also like to give special acknowledgement to Wenlin and Emmiline for reading the manuscript.

Additionally, I am grateful to Miss Ai Hua Seow for her kind help in using various instruments in the CBC teaching lab. Many thanks are also extended to the graduate student Wang Yong in Prof. Hongyu Chen's lab for training me and maintaining

TEM. I would like to acknowledge Prof. Julien Lescar and Prof. Anh Tuan Phan for the use of their FPLCs.

I would also like to give special thanks to a very kind couple, my landlord Tan Tian Sze and Lilis, whom I have lived with for three years in a foreign country. I am very grateful for staying with such a warm family and being taken cared of like their young sister.

Last but not least, I would like to express my deepest love and appreciation to my parents for their everlasting support and to my husband, Steven Hao Yuxiang for his unconditionally contribution, support and encouragement to my study.

TABLE OF CONTENTS

ACKNOWLEDGEMENT..... i

TABLE OF CONTENTS.....iii

ABSTRACTvii

CHAPTER 1 Introduction1

1.1 Protein structure.....11.2 Protein self-assembly and folding2

1.2	Protein self-assembly and folding.....	2
1.3	Characteristics of protein-protein interactions.....	3

1.3.1 Amino acid composition at protein-protein interfaces31.3.2 Hydrophobicity.....5

1.3.2	Hydrophobicity.....	5
1.3.3	Hydrogen bonds.....	6

1.3.3	Hydrogen bonds.....	6
1.3.4	Electrostatic interactions.....	6

1.3.4	Electrostatic interactions	6
1.3.5	Interface size and shape	6

1.3.5	Interface size and shape	6
1.3.6	Other properties of protein-protein interfaces	7

1.3.6.1 Geometric complementarity between surfaces.....7

1.3.6.1	Geometric complementarity between surfaces	7
1.3.6.2	Secondary structures of interface residues.....	8

1.3.6.2	Secondary structures of interface residues	8
1.3.6.3	Conformational changes on complex formation.....	8

1.3.6.4 The role of bound water molecules on protein-protein interactions.....8

1.3.5.0.4	The role of bound water molecules on protein-protein interactions.....	8
1.4	Protein cages.....	9

1.4	Protein cages.....	9
1.4.1	Ferritins.....	11

1.4.1	Ferritins	11
1.4.1.1	Structures of maxi-ferritin and mini-ferritin	13

1.4.1.1	Structures of maxi-ferritin and mini-ferritin.....	13
1.4.1.2	Function of ferritins	15

1.4.1.2	Function of ferritins	15
1.5	Self-assembly of ferritin superfamily	16

1.5.1	Maxi-ferritin self-assembly.....	16
-------	----------------------------------	----

1.5.1	Maxi-ferritin self-assembly.....	16
1.5.1.1	Proposed pathways of maxi-ferritin assembly.....	16

1.5.1.1 Proposed pathways of maxi-ferritin assembly.....16

1.5.1.2 Mutation related studies on the effect of maxi-ferritins self-assembly...18

1.5.2 Mini-ferritin self-assembly 23

1.5.2	Mini-ferritin self-assembly	23
1.5.2.1	Proposed pathways of mini-ferritin assembly	23

1.5.2.2 Mutation related studies on the effect of mini-ferritins self-assembly ...23

1.5.2.3 pH effects on the self-assembly of mini-ferritin 24

1.5.2.3	pH effects on the self-assembly of mini-tertiary.....	24
1.6	Research perspectives.....	25

1.6	Research perspectives.....	23
1.7	References.....	25

CHAPTER 2 Alanine shaving mutagenesis studies on a maxi-ferritin,
bacterioferritin from *E. coli* (BFR)32

2.1 Introduction.....322.2 Bacterioferritin.....332.2.1 The structure of *E. coli* bacterioferritin (BFR)342.2.1.1 The monomeric structure of BFR.....352.2.1.2 The quaternary structure and assembly of subunits to octahedral symmetry352.2.2 Sequence and structural alignment of bacterioferritins362.3 Aims and objectives.....39

2.3	Plants and objectives.....	39
2.4	Materials and methods.....	40

2.4.1	Virtual alanine scanning of BFR.....	40
-------	--------------------------------------	----

2.4.2	Analysis of the protein-protein interactions at the symmetry related interfaces of BFR using PISA software	40
2.4.3	Synthesis of BFR and its mutants	40
2.4.3.1	Oligonucleotides design	40
2.4.3.2	Gene preparation and cloning	42
2.4.3.3	Plasmid preparation using site directed mutagenesis	46
2.4.4	Protein expression and purification.....	47
2.4.5	Transmission electron microscopy (TEM).....	48
2.4.6	Size exclusion chromatography (SEC).....	48
2.4.7	Dynamic light scattering (DLS).....	48
2.4.8	Temperature-dependent circular dichroism (CD).....	49
2.4.9	Native PAGE.....	49
2.4.10	Chemical denaturation.....	49
2.5	Results and discussion.....	49
2.5.1	Selection of key interfacial residues for mutagenesis.....	49
2.5.2	The protein-protein interactions at the symmetry related interfaces of BFR	54
2.5.2.1	Comparison of the buried surface area and exposed surface area between different bacterioferritins	54
2.5.2.2	The individual residues involved in protein-protein interactions at symmetry related interfaces in <i>E. coli</i> BFR	55
2.5.3	Verification of gene synthesis products	60
2.5.4	Protein over-expression and purification	62
2.5.5	Transmission electron microscopy	63
2.5.6	Size exclusion chromatographic analysis of assembly in solution	65
2.5.7	Dynamic light scattering (DLS) analysis of BFR R30A, R61A, Y114A, and E128A assembly in solution	67
2.5.8	Native gel electrophoretic (native PAGE) analysis of BFR R30A, R61A, Y114A, and E128A assembly in solution.....	68
2.5.9	The solution secondary structure as monitored by circular dichroism.....	68
2.6	Conclusion	74
2.7	References	79

CHAPTER 3 Additional insight into maxi-ferritin self-assembly and stability:		
Further mutagenesis at the BFR two-fold symmetry interface, wild-type BFR pH stability and thermodynamics with DSC.....		
3.1	Introduction.....	83
3.1.1.1	The influence of pH on protein structure.....	83
3.1.1.2	Differential scanning calorimetry (DSC) to study protein folding and stability	84
3.1.1.3	Introduction to FoldX	87
3.1.2	Previous studies on maxi-ferritin pH and thermodynamic stability.....	88
3.1.2.1	pH stability of maxi-ferritins	88
3.1.2.2	Calorimetric studies of maxi-ferritins	89
3.2	Aims and objectives	90
3.3	Materials and methods	91
3.3.1	Additional mutagenesis studies focusing on the BFR two-fold symmetry interface.....	92
3.3.1.1	Computed unfolding free energy using FoldX	92
3.3.1.2	Gene construction, expression and protein purification and analysis using CD, TEM, DLS, SEC and native PAGE.....	92

3.3.2	pH stability of BFR 24-mer	93
3.3.2.1	Sample preparation for acid denaturation.....	93
3.3.2.2	Native PAGE	93
3.3.2.3	Size exclusion chromatography at different pH.....	93
3.3.2.4	Circular dichroism analysis	94
3.3.3	Thermodynamic analysis using differential scanning calorimetry.....	94
3.4	Results and discussion.....	95
3.4.1	Mutagenesis studies on BFR two-fold interface.....	95
3.4.1.1	Computational assisted mutation determination by FoldX	95
3.4.1.2	Gene construction, expression and protein purification	96
3.4.1.3	Transmission electron microscopy.....	97
3.4.1.4	Size exclusion chromatographic analysis of assembly in solution.....	98
3.4.1.5	Dynamic light scattering (DLS) and native PAGE	99
3.4.1.6	The secondary structure and thermal stability as monitored by circular dichroism	100
3.4.2	pH stability analysis of wild-type BFR.....	103
3.4.2.1	pH-dependent behavior of wild-type BFR.....	103
3.4.2.2	Secondary structure and thermostability of wild-type BFR at different pH.....	106
3.4.3	Thermodynamic studies of BFR and its mutants R30A, Y114A and E128A by DSC.....	110
3.4.3.1	The effect of different concentrations of dimer on the DSC thermogram.....	110
3.4.3.2	Unfolding of the impure BFR dimer is scan rate dependent.....	111
3.4.3.3	The effect of different concentrations and scan rate on DSC thermal unfolding of isolated BFR 24-mer.....	112
3.4.3.4	Comparison of BFR 24-mer and dimer	113
3.4.3.5	Rescan of isolated BFR dimer and 24-mer results in similar thermal profiles	115
3.4.3.6	Comparison of DSC thermograms for mutants R30A, E128A, Y114A with the isolated dimer of wild-type BFR	117
3.5	Conclusion	119
3.6	References.....	120

CHAPTER 4 Alanine shaving of mini-ferritin, *E. coli* DPS test..... 123

4.1	Introduction.....	123
4.2	Structure features of Dps proteins.....	123
4.2.1	Tetrahedral symmetry of Dps and the structural comparison between Bfr and Dps.....	124
4.2.2	The ferroxidase center.....	127
4.2.3	DNA binding.....	129
4.3	Aims and objectives.....	130
4.4	Materials and methods	131
4.4.1	Virtual alanine scanning of DPS	131
4.4.2	Gene synthesis of DPS and its mutants.....	132
4.4.3	Protein expression and purification.....	132
4.4.4	Transmission electron microscopy (TEM).....	134
4.4.5	Size exclusion chromatography (SEC).....	135
4.4.6	Temperature-dependent circular dichroism (CD).....	135
4.4.7	Native PAGE.....	136
4.5	Results and discussion.....	136

4.5.1	Selection of key interfacial residues for mutagenesis.....	137
4.5.2	Verification of cloning.....	143
4.5.3	Protein over-expression and purification	144
4.5.4	Transmission electron microscopy	146
4.5.5	Size exclusion chromatography (SEC).....	147
4.5.6	Temperature-dependent circular dichroism analysis	149
4.6	Conclusion	154
4.7	References.....	158

Appendixes

Appendix 1.	The gene design of BFR synthesis.....	162
Appendix 2.	Sequencing results of BFR and BFR derived genes.....	163
Appendix 3.	The gene design of DPS synthesis.....	170
Appendix 4.	Sequencing results of DPS and DPS derived genes.....	171
Appendix 5.	The primers used for four charged mutants at BFR two-fold axis	179
Appendix 6.	Sequencing results of four charged mutants at BFR two-fold axis	179
Appendix 7.	The calibration for the size exclusion chromatography column Sephadex 200 10/300 at different pH.....	183

ABSTRACT

Investigations into protein quaternary structure can lead to deeper insight into the fundamentals governing both protein folding and protein-protein interactions. In addition they can provide a foundation for the eventual rational design of novel complex protein architectures. A maxi-ferritin, bacterioferritin from *E. coli* (BFR), and a mini-ferritin, DNA-binding protein from starved cells (DPS), despite their similar four-helix bundle tertiary structure, assemble into quaternary structure with different symmetries, octahedral and tetrahedral and oligomerization states, 24-mer and 12-mer, respectively. To understand how these two structurally analogous proteins assemble into nano-structures with different sizes and shapes, both proteins were chosen as the basis for a mutagenesis study to investigate the importance of key amino acid residues, located at symmetry-related protein-protein interfaces, in controlling protein stability and self-assembly. Several mutants were designed for each protein through simple inspection and computational analysis, synthesized and subjected to different chemical and biophysical methods to determine their thermal stability, self-assemble ability and structure. The data indicate that many of these residues may be hot spot residues. Several mutants were observed to completely shut down detectable solution formation of 24-mer, favoring a cooperatively folded dimer, suggesting that they may be oligomerization “switch residues”. This investigation into the structure and energetics of these self-assembling nano-cage proteins not only can act as a jumping off point for the eventual design of novel protein nano-structures and their applications, but it can also help to understand the role that structure plays in the function of these important classes of proteins.

CHAPTER 1 Introduction*

1.1 Protein structure

Protein, a word coined by Jons J. Berzelius in 1838, is derived from the Greek word “proteios” which means “the first rank” to emphasize the importance of this class of molecules¹. Proteins play crucial roles in almost all biological processes.

Amino acids are the basic structural units of proteins. Proteins in every species, ranging from bacteria to humans, are constructed from the same set of twenty canonical amino acids. These twenty amino acids have different side chains which vary in size, shape, charge, hydrogen-bonding capacity and chemical reactivity. Amino acids that are linked by peptide bonds to form a sequence which is the primary structure of a protein. Studying the relationship between the amino acid sequence and the three-dimensional structure of proteins plays a critical role in revealing the rules governing the polypeptide chain folding process. Secondary structure refers to the three-dimensional spatial arrangement of amino acid residues that are adjacent or nearly adjacent in the linear sequence. Some of these steric relationships are of a regular kind, giving rise to a periodic structure. These tend to be controlled by steric requirements about the rotatable bonds in the individual amino acid residues. The α helix and β strand are elements of secondary structure. Tertiary structure refers to the spatial arrangement of amino acid residues that are relatively nonadjacent in the linear sequence. The folding process is primarily driven by relatively non-specific hydrophobic interactions and the structure is stable when parts of the protein domain are locked into place by specific tertiary interactions, such as hydrogen bonds, salt bridges and disulfide bonds. Quaternary structure is the

* Parts of this chapter have been published previously with the author of this thesis as the first author.
Int. J. Mol. Sci. **2011**, 12

assembly of two or more polypeptide chains or protein subunits. The non-covalent interactions and disulfide bonds involved in stabilizing and rendering specificity to the tertiary structure are also used in generating the quaternary structure.

1.2 Protein self-assembly and folding

Self-assembly is the process where a disordered system forms an organized structure or pattern as a consequence of local interactions among the components themselves without external intervention. Self-assembly processes are common throughout nature and technology.

One of the most important and ubiquitous examples of biological self-assembly is protein folding. The amino acid sequence of a protein determines its structure, which determines its mechanism of action. The study of how biomacromolecular structure relates to activity directly or indirectly accounts for nearly one in four Nobel Prizes in Chemistry since 1956 ². Despite the enormous amount of effort devoted to understanding “the protein folding problem”, however, a complete understanding of the folding/unfolding mechanisms has not yet been established.

Mistakes can occur in the folding of proteins. These errors, in turn, can cause cellular disorders due either to the loss of function of the misfolded protein or to new and pathogenic activities attributed to the altered fold. Diseases associated with protein misfolding are wide ranging (Alzheimer’s Disease, Bovine Spongiform Encephalopathy, Huntington’s Disease, Parkinson’s Disease, Type II Diabetes etc) and have been discovered in many different tissues and organs ³. Thus, how proteins fold is not only one of the “big unanswered scientific questions” ⁴, but understanding the fundamentals of protein folding could also identify potential errors in the processes leading to misfolding, and could therefore shed significant light on all of these diseases in general.

Protein-protein interactions have grown in interest within the pharmaceutical industry following the discovery that many, though not all, while possessing interfaces with large surface areas, have binding energy that is focused onto a small number of amino acid residues, or “hot spots,”⁵⁻⁷ thereby suggesting that it could be possible to inhibit these medicinally attractive interactions with small molecule drugs⁸. An understanding of the fundamentals of protein-protein interactions would therefore provide insight into drug design in addition to facilitating a multitude of other applications such as the engineering of interfaces for the development of antibodies to new or non-immunogenic antigens, or the engineering of single-celled organisms for bio-remediation.

The understanding of protein folding has advanced to the stage where it is now possible to design proteins *de novo*⁹. However, the majority of these designs have either involved small, monomeric proteins, or are based on α -helix bundles¹⁰. Only recently has the rational design of multi-component protein structures begun to be studied¹¹⁻¹³. Engineering the quaternary structure of a protein is the ultimate test of our understanding of the processes of folding and self-assembly.

1.3 Characteristics of protein-protein interactions

Protein-protein interactions at protein interfaces are governed by both chemical and geometrical properties. Chemical properties include types of chemical functional groups at the interface which influence the hydrophobicity, hydrogen bonds and electrostatic interactions, while the geometrical properties involved interface size and shape, planarity and complementarity and spatial positioning of the functional groups.

1.3.1 Amino acid composition at protein-protein interfaces

Amino acid residues can be described as non-polar, polar or charged. Different types of complexes possess different propensities for different types of residues.

Studies on the residue propensities at protein-protein interfaces by Jones *et al* ¹⁴ showed that hydrophobic residues, with the exception of methionine, have a greater preference for the interfaces of homodimers than for hetero-complexes.

Alanine shaving mutagenesis allows an exploration of the energetic contributions of individual side-chains in protein complex formation ¹⁵. Wells and colleagues have discovered that free energies are not uniformly distributed across the interfaces and certain critical residues can contribute toward a large fraction of the binding free energy ^{15,16}. To analyze the contributions of individual amino acid residues in protein-protein binding, Bogan *et al* ⁶ compiled a database of 2325 alanine mutants and measured the change in free energy of binding upon mutation of respective amino acids to alanine. According to their definition, a hot spot is a residue that when mutated to alanine, gives rise to a distinct change in the absolute binding energy of more than 2 kcal/mol. These energetically important residues are often tryptophan, tyrosine and arginine, and are surrounded by the energetically less important residues (valine, leucine and serine) that probably serve to occlude bulk solvent from the hot spot.

Ma *et al* ¹⁷ studied structurally conserved residues between binding sites and exposed protein surfaces. They discovered that the conservation of tryptophan on the protein surface indicates a highly likely binding site. The conservation of phenylalanine and methionine, to a lesser extent, also imply a binding site. However, for all three residues (tryptophan, phenylalanine and methionine), there is no conservation on the exposed surface.

Generally, hydrophobic residues (except alanine) show higher preference in protein-protein interfaces ¹⁸. Furthermore, aromatic residues are reported to be more abundant at the interfaces than aliphatic residues ¹⁹. Tryptophan, tyrosine and arginine

showed the highest propensity in protein-protein interface ^{19,20}. In addition, hydrophilic and charged residues (except arginine) are found to be depleted at the interfaces ²¹.

1.3.2 Hydrophobicity

The “hydrophobic effect” has been shown to be the main driving force in protein folding ²². In an aqueous environment, this effect is driven by the strong tendency of hydrophobic residues to be excluded from water. Water is highly cohesive and hydrophobic groups are thermodynamically more stable when clustered in the interior of the protein as compared to extending into the aqueous surroundings. Hence, the polypeptide chain tends to fold spontaneously so that its hydrophobic side chains are buried and its polar, charged chains are exposed to the aqueous environment.

Generally, proteins have a hydrophobic core and a hydrophilic surface ²³. The homodimer interfaces are observed to be more hydrophobic when the interface hydrophobicity of a homodimer is compared to hetero-complexes. These observations may be attributed to the differing roles of the two types of complexes ²⁴. The homodimers usually function as the form of dimer in the cell. However, the hetero-complexes usually function as monomers in solution. It is therefore energetically unfavorable to have a large exposed hydrophobic area.

It has been assumed that proteins associate with each other through their hydrophobic patches ¹⁴. Some interfaces have recognizable hydrophobic cores surrounded by a large hydrophilic patch, which can be involved in inter-subunit polar interactions. On the other hand, other interfaces have many small hydrophobic patches consisting of 1 ~ 3 amino acids instead of a single hydrophobic core. These small patches are linked by inter-subunit hydrogen bonds and water molecules across the interface.

1.3.3 Hydrogen bonds

Hydrogen bonds are the attractive interactions between a hydrogen atom, an acceptor and an electronegative atom which acts as a donor. Generally, there are three main types of hydrogen bonds in protein structures. Apart from the solvent and hetero-atom mediated hydrogen bonds^{25,26}, hydrogen bonds can be formed between main-chain and main-chain, main-chain and side-chain, side-chain and side-chain. Although the inter-molecular hydrogen bonding energy is quite small (2 ~ 10 kcal/mol) as compared to the covalent bond, the high number of hydrogen bonds in a protein structure means they can play an important role in protein-protein interactions²⁷.

1.3.4 Electrostatic interactions

Electrostatic interactions are known to play a crucial role in protein-protein interactions^{28,29}. Opposite charged amino acids tend to form salt bridges between protein complexes. There is a range of 0 ~ 12 charged groups²⁸ and 0 ~ 5 salt bridges³⁰ per protein-protein interface. However, there is no clear correlation between salt bridges and the interface size.

The electrostatic contribution towards protein-protein interaction energetics is more favorable when the hydrophilic pair is transferred from the protein surface in the folded but unbound state to the complex interface in the bound state, as compared to the change from the unfolded state to the folded state²⁸. The hydrophobic effect and electrostatic interaction appeared to contradict each other in that as more hydrophobic atoms at the interface result in less hydrophilic pairs and vice versa. However, the hypothesis may be invalid and in fact, a favorable electrostatic interaction probably requires hydrophobic residues nearby to modulate solvation²⁸.

1.3.5 Interface size and shape

Interface size, an important geometrical property has been used to describe protein-protein interfaces. It is usually characterized by the interface area or the number of interface residues, and the number of interface residues has shown to correlate linearly to the interface area ^{21,31}. In addition, there is correlation between the hydrophobic free energy of transfer from polar to a hydrophobic environment and the solvent accessible surface area (ASA) ³². The size of the interface can be measured simply in terms of the changes in solvent accessible area (Δ ASA) on complexation and thus Δ ASA provides an evaluation of the binding strength.

Different types of protein structures possess different interfacial topologies. Protein-protein associations may form a relatively flat or even a twisted interface. Measuring the deviation of interface residues from a plane (termed planarity) is applied to determine whether the protein-protein interfaces are flat or twisted. The interface planarity is analyzed by calculating the RMSD of all the interface atoms from the best-fit plane through the atoms ¹⁴, and a smaller RMSD value represented flatter interfaces.

1.3.6 Other properties of protein-protein interfaces

Others properties characterizing protein-protein interactions include: (1) geometric complementarity; (2) secondary structures of interface residues; (3) conformational changes on complex formation; (4) the presence of discrete water molecules at the interface. These properties will be discussed in the following sections.

1.3.6.1 Geometric complementarity between surfaces

Geometric complementarity of two protein surfaces that are in close proximity is a major element of the recognition process between the two proteins. Parameters such as gap volume and gap index ¹⁴ are used to evaluate the geometric complementarity. The gap volume estimates the volume enclosed between any two molecules,

delimiting the boundary by defining a maximum allowed distance between both interfaces.

$$\text{Gap index (\AA)} = \text{gap volume between molecules (\AA}^3\text{)} / \text{interface ASA (\AA}^2\text{)} \text{ (per complex)}$$

1.3.6.2 Secondary structures of interface residues

There is usually more than one type of secondary structure in most protein-protein interfaces. It was found that there is an approximately equal involvement of helical, strand and coil residues. Although some interfaces may contain a single type of structure (helices, strand or loops), most interfaces consist a mixture of secondary structures¹⁴ and no generalities are immediately relevant.

1.3.6.3 Conformational changes on complex formation

Conformational changes between subunits can sometimes occur upon protein complex formation. The changes can be distinguished on different levels, ranging from solely side chain movement, segment movement involving the main chain, to changes associated over entire domains³³.

The magnitude of binding-induced conformational change upon binding can be predicted from the relative solvent accessible surface area. It is observed that conformational change upon binding is a common phenomenon³⁴. Additionally, subunits of protein complexes with large interfaces ($> 2000 \text{ \AA}^2$) have a higher tendency to undertake greater conformational changes upon binding than those with smaller interfaces³⁵.

1.3.6.4 The role of bound water molecules on protein-protein interactions

Bound water molecules are very often found at protein-protein interfaces. Water molecules contribute to the close packing of atoms that ensure complementarity

between the two protein surfaces, as well as provide polar interactions between two proteins³⁶. Lo Conte *et al* examined the number of interface water molecules in thirty-six protein structures. The number varies from three to fifty but has an average around eighteen. It was also shown that there is about one water molecule per 100 Å² interface area³⁶.

Most interface water molecules are involved in forming hydrogen bonds, and the number of water molecules has a weaker correlation with the interface area than does the number of hydrogen bonds³⁷. These water molecules form hydrogen bonds with protein residues or other water molecules and contribute to hydrogen bond networks²⁸. However, the effect of water molecules on protein-protein interactions has not been fully understood, and most studies on protein-protein interactions have placed little emphasis on the effect of bound waters.

1.4 Protein cages

Protein cages are structures that are roughly spherical in shape, and enclose a central, hollow space. There are two surfaces in the spherical cages: the inner surface and the outer surface which projects into the central cavity. Many proteins self-assemble from monomers in solution and vary in size. Some examples that have been widely studied include viral capsids, small heat shock protein (sHSP), ferritins and DNA-binding proteins from starved cells (DPS) (Figure 1.1).

Viral capsids are another major class of protein cages. The diameters range from 18 to 500 nm for icosahedral viruses. The variety in shape size, and self-assembly characteristics make viruses very suitable for the controlled synthesis of inorganic materials or the encapsulation of other materials³⁸. Cowpea chlorotic mottle virus (CCMV) and cowpea mosaic virus (CPMV) are particularly widely utilized capsids.

Heat shock proteins are produced in high levels in response to cellular stress and

assist in protein folding. The sHSPs can form large multimeric structures with the function of providing cells with thermotolerance by functioning as molecular chaperones. The twenty-four subunits of the sHSP from the hyperthermophilic archaeon *Methanococcus jannaschii* self assemble into a 16.5 kDa cage with octahedral symmetry ³⁹. The protein cage is stable at a pH range of 5 - 11 and can withstand temperatures up to 70 °C. The sphere has an outer diameter of 12 nm and an inner diameter of 6.5 nm and has large pores (3 nm in diameter) ⁴⁰.

Vaults are large barrel-shaped ribonucleoprotein capsules with an inner cavity large enough to encapsulate hundreds of proteins ⁴¹. Vaults are highly conserved in a wide variety of eukaryotes and mostly found in the cytoplasm. Although vaults have been proposed to play a role in multidrug resistance, cell signaling and innate immunity, since their discovery in 1986 ⁴²⁻⁴⁴, their cellular function still remains unclear.

Ferritins, one of the most widely studied and utilized cage proteins, have a primary role of *in vivo* storage of ferric oxide, and are made up of twenty-four identical, or nearly identical, protein subunits. Highly conserved, ferritins are found in all domains of life. Dps proteins are members of the ferritin superfamily. These proteins prevent DNA damage by condensing DNA and accumulating iron atoms within their central cavity. The subunit of the DPS proteins have structural homology to that of the ferritins however they self-assemble into a smaller 12-mer cage. Detailed properties of these proteins will be discussed in the following sections.

Figure 1.1. Examples of the protein cages emphasizing a range in size. **(A)** Rat liver vault (adapted from the molecule of the month of RCSB protein data bank, <http://www.pdb.org/pdb/101/motm.do?momID=114>). **(B)** Cowpea chlorotic mottle virus (CCMV) (adapted from Bode *et al*, 2011⁴⁵). **(C)** Small heat shock protein **(D)** ferritin **(E)** Dps, (the latter three figures were prepared using UCSF Chimera with PDB entries of 1shs, 2q3y and 1dps respectively).

1.4.1 Ferritins

The ferritins, a family of protein cages, play a key role in iron sequestration and are highly evolutionary ubiquitous⁴⁶⁻⁴⁸. The first ferritin was isolated from horse spleen in 1937⁴⁹ and the crystal structure was determined in 1991, which revealed the protein to be a 24-meric cage with octahedral symmetry⁵⁰. Subsequently, ferritins from diverse organisms such as animals^{51,52}, plants^{53,54} and bacteria⁵⁵⁻⁵⁷ have been isolated and crystallized, and these possess structures related to that of horse spleen ferritin. Further studies have demonstrated that the protein cages store excess cellular iron as mineralized hydrous ferric oxide in their cavities. The ability to sequester iron grants ferritins dual functions in iron detoxification and in establishing a cellular iron

reserve. Although the DNA and amino acid sequences for the ferritins vary considerably ⁵⁸ (up to 80%), their well-conserved three-dimensional tertiary structures indicate identical or similar folding⁴⁶. Their self-assembled cage structures, however, can differ considerably.

The ferritin superfamily can be broken into three sub-families: the classical ferritins (Ftn), the bacterioferritins (Bfr), and the DNA-binding proteins from starved cells (Dps). The Ftn and Bfr proteins are considered maxi-ferritins, whereas Dps proteins are mini-ferritins. These three sub-families share the same characteristic four-helix bundle fold ^{59,60}. The Ftn proteins are found in all three domains of life (eukarya, archaea and bacteria) and are typical members of the ferritin family. The Bfr proteins have identical quaternary structure as does the Ftn proteins, however they are restricted to bacteria and archaea. The presence of twelve heme moieties in Bfr proteins is the most significant difference between the Ftn and Bfr proteins (Figure 1.2). The Dps proteins form a smaller molecule with a lower iron storage capacity than the Ftn and Bfr proteins and utilize unique ferroxidase sites.

Figure 1.2 A bacterioferritin and heme. **(Left)** A bacterioferritin viewed from the three-fold symmetry axis with the heme indicated in red. **(Right)** The heme is located at the interface between protein monomers at the two-fold symmetry axis. Figures were generated using Chimera ⁶¹, PDB ID: 1bfr.

1.4.1.1 Structures of maxi-ferritin and mini-ferritin

Maxi-ferritins are composed of twenty-four identical or similar subunits (~ 19 kDa) that assemble into a large spherical cage (outer diameter ~ 120 Å) with a hollow cavity (inner diameter ~ 80 Å). The cavity can accommodate up to ~ 4500 Fe atoms in the form of a hydrous ferric oxide mineral core with variable amounts of phosphate⁶². Each monomer is made up of a four-helix bundle (the A, B, C and D helices) with a short fifth helix (the E helix) at the C-terminus (Figure 1.3 (a)).

A number of intra- and inter-subunit hydrogen bonds and salt bridges contribute to the stability of the ferritin shell to a wide range of pH (2.0 ~ 12.0)⁶³ and temperature (up to 80 °C)⁶⁴ conditions. Maxi-ferritin assembles to form a cage with eight and six inter-subunit channels along the three and four-fold symmetry axes respectively. The three-fold channels are lined by several hydrophilic Asp and Glu carboxyl side chains which are thought to provide the main transfer pathway for Fe (II), Fe (III) and other small molecules into and out of the ferritin shell⁶⁵. The four-fold inter-subunit channels are lined by hydrophobic amino acid side chains, and are therefore not responsible for ion transport.

The mini-ferritin Dps proteins, form similar cage-like oligomers as the maxi-ferritins but self assemble from only twelve monomers. The first Dps protein family member was isolated in 1992 from *E. coli* grown in starvation conditions. Along with protecting bacteria from oxidative damage it also forms an extremely stable complex *in vitro* with DNA without apparent sequence specificity (see Chapter 4 DNA binding section)⁶⁶. The crystal structure of dodecameric *E. coli* Dps reveals a hollow protein with 32 (tetrahedral, D_{3h}) point group symmetry demonstrating that Dps is a structural analogue of the maxi-ferritins⁶⁷, although the main function of Dps proteins is iron detoxification as opposed to iron storage. The Dps dodecamer

measures ~ 9 nm in diameter and has a central cavity of ~ 4.5 nm which can hold an iron core of up to ~ 500 Fe^{3+} iron ions ⁶⁷. Similar to the maxi-ferritins, the Dps monomer folds into a four-helix bundle (the A, B, C and D helices). However, unlike the maxi-ferritin, the loop between the B and C helices forms a short helix (Figure 1.3 (b)). The BC helix runs nearly orthogonal to the four-helix bundle axis and is exposed on the outside of the assembled protein cage. Moreover the Dps monomer contains no E helix as is found in the maxi-ferritins.

Both the maxi- and mini-ferritins utilize the electrostatic potential generated by the negatively charged residues lining the pores at the N-terminal, three-fold symmetric, ferritin-like, interface to help iron enter and exit the protein inner cavity ^{68,69}. The pore formed at the unique C-terminal Dps-like interface of the mini-ferritins is smaller, due to hydrophobic constriction, and less acidic ⁶⁷. Therefore, it most likely plays no role in iron transport.

Figure 1.3 Typical structures of octahedral maxi-ferritin (**a**, PDB ID: 1bfr) and tetrahedral mini-ferritin (**b**, PDB ID: 1dps). Their four-helix bundle monomers are shown in ribbons.

The A helix is colored orange, the B helix yellow, the BC loop in the maxi-ferritin (a) or BC in the mini-ferritin helix (b) cyan, the C helix blue, the D helix green and the E helix red. The figure is generated using UCSF Chimera ⁶¹.

1.4.1.2 Function of ferritins

The ferritins of animals and plants and the bacterioferritins have a common iron-storage function that oxidizes toxic Fe (II) ions to crystalline ferrihydrite in the protein cavity using either oxygen or hydrogen peroxide ⁷⁰. Despite common functionality and structural similarity, differences in primary structure, cytological location and biosynthetic regulation are exhibited among members of the ferritin superfamily. For example, animal ferritins are usually found in the cytosol whereas plant ferritins are located in plastids ⁶². Based on Theil ⁶², the common feature of ferritin function, iron storage, is categorized into three different types: iron storage for other cells (specialized-cell ferritin), iron storage for intracellular needs (normal housekeeping ferritin), and iron storage for intracellular protection from iron overload (stress housekeeping ferritin) ⁷¹. Specialized-cell ferritin, the most abundant type of ferritin, e. g. in hepatocytes, performs long term storage for the organism. In contrast, housekeeping ferritin are present in most cell types at low levels to maintain the iron reserve required for intracellular metabolism.

Iron oxidation and mineralization is a multi-step process. Deposition of iron into the ferritin cavity requires binding of Fe (II) to the ferritin at the ferroxidase site, followed by oxidation from Fe (II) to Fe (III). Subsequently, the Fe (III) undergoes hydrolysis, migration and nucleation to form the mineral core ^{46,72}. Eight hydrophilic three-fold symmetry axes associated channels that traverse the protein shell are proposed to be the pathways taken by Fe (II) to enter the interior of eukaryotic ferritins. Mammalian ferritins usually consist of two types of subunits, heavy (H) chain and light (L) chain which have complementary functions in iron uptake. Within

the H-chain monomer a dinuclear ferroxidase site catalyzes the oxidation of Fe (II) ion by oxygen to produce hydrogen peroxide. Although the L-chain lacks a ferroxidase site, it contains additional glutamate residues on the interior surface of the protein shell, which facilitates mineralization and turnover of Fe (III) at the H-chain ferroxidase center.

Another, often neglected, function of the ferritins is the release of stored iron from their mineral core for iron-requiring metabolic processes. Two models have been proposed for recovering the stored iron from ferritin. First, the ferritin-degradation model proposes that cytoplasmic ferritins that are incorporated into the lysosome are degraded to release the mineral core. Subsequently, the ferritin proteins are further digested and the resulting Fe (III) is exported to the cytoplasm. The second model invokes iron chelating agents to transport Fe (II) which has been produced from the reduction of the Fe (III) mineral core⁷³.

1.5 Self-assembly of ferritin superfamily

1.5.1 Maxi-ferritin self-assembly

1.5.1.1 Proposed pathways of maxi-ferritin assembly

Resolving the self-assembly mechanism of protein supramolecular complexes is difficult especially if the structures are highly symmetrical and homo-oligomeric. Moreover, the characterization of the various thermodynamically and kinetically accessible intermediates for the association and disassembly pathways can help achieve mechanistic insight; however, these studies can be challenging due to difficulties with the determination of folding and assembly rate-limiting steps and coupling between the two⁷⁴. Thus, only a few studies have explored the self-assembly mechanism of maxi-ferritins.

X-ray diffraction analysis into the assembly mechanism of horse spleen apoferritin suggests that stable dimers act as assembly intermediates ⁷⁵. Furthermore, sedimentation velocity analytical ultracentrifugation of this protein at various pHs demonstrates the stability of dimers in solution ⁷⁶ and also suggests that assembly proceeds from dimers to tetramers and octamers. The overall assembly mechanism of horse spleen apoferritin was first proposed by Gerl *et al* ⁷⁷ who interpreted data obtained through intrinsic fluorescence, far-UV circular dichroism and glutaraldehyde cross-linking experiments. It was observed that the completely self-assembled product was formed through a series of concentration dependent association reactions involving a mixture of partially assembled subunits. These subunits include the “structured monomer”, dimer, which was the most highly populated species, trimers, hexamers, in small amounts, and dodecamers. The overall proposed mechanism is described by the following scheme:

where M_1^* is the unfolded monomer, M_i are intermediates with i folded monomers and M_{24} is the completely self-assembled ferritin. In this model, to initiate assembly, the unfolded monomers must first acquire a native-like conformation to generate “structured monomers” which provide complementary interfaces and subsequently dimerize at high rate. This mechanism was supported by further crosslinking experiments ⁷⁸ where dimers, trimers and tetramers were isolated to investigate their capacity for reassociation. It was shown that two hexamers could be used to form a dodecamer, and two dodecamers could assemble into a 24-mer. These results led to a refined model where the 24-meric cage assembles from dimer (M_2) via tetramers (M_4) and hexamers (M_6). This mechanism is supported by Banyard *et al* ⁷⁵, who proposed that the monomer is expected to be unstable, and the stable tetramer and

hexamer intermediates can be thought of as dimers of dimers and trimers of dimers respectively. However, the cross-linking experiments suggested that both dimer and monomer are involved in the formation of a stable trimer, hexamer is a transient intermediate that could be detected only in small amounts and tetrameric and octameric intermediates are undetectable. While these results don't support the mechanism, they don't completely rule it out in that it is possible to imagine a model where some of the isolatable oligomerization states are unproductive assembly dead-ends that need to completely or partially disassemble before forming a productive, albeit short lived, intermediate.

Horse spleen apoferritin is highly resistant to chemical denaturation, pH changes and heat ^{63,64}. Recently, though, it was shown through quantitative data analysis of SAXS measurements that apoferritin can undergo stepwise disassembly through several structural intermediates below pH 3.40 ⁷⁹. The dissociation process starts with hollow spherical structures with two holes, followed by “headphone”-shaped structures, and ultimately, rod-like oligomers (mainly trimers) or monomers. The structural recovery of the intermediates during the pH-induced reassembly process is dependent on the history of the disassembly process; for example the hollow sphere with the double-hole defect could never be recovered back to the intact hollow sphere. How this data relates to proposed mechanisms of assembly has not been fully explored.

1.5.1.2 Mutation related studies on the effect of maxi-ferritins self-assembly

Many mutations and modifications of maxi-ferritins have been studied to evaluate the role that various residues and regions play in protein assembly and function. Through this process, many mutations have been discovered that have little to no effect on the self-assembly of the cage, demonstrating their robustness, which bodes

well for the utility of these proteins in engineering and bio-conjugation applications. Conversely, mutations have been discovered that radically affect the self-assembly though they are relatively conservative and in some cases are even single point mutations. These cases dramatically emphasize the possibility of rationally manipulating the properties of these proteins with precision.

One early study involved engineering of *E. coli* BFR to replace the C-terminal residues (REEG) with the eighteen residue λ peptide (RLPFTSCAVCLQDSMRSR)⁸⁰. The C-terminus in this maxi-ferritin is positioned at the end of the E helix and thus points inside the cavity, suggesting the peptide extension would also would be present inside the assembled ferritin. Gel filtration demonstrated that the mutant exists solely as 24-mer, whereas the wild type is a mixture of 24-mer and dimer. The greater assembly stability of the variant with respect to wild type BFR implies that the C-terminal extensions at the C4 interface are forming additional contacts. However, crystallization and *in vitro* iron uptake experiments suggest that the peptide extension is possibly blocking access of iron to the central cavity in the three-fold interfaces.

The E helix from the C-terminus of *E. coli* BFR was found to exhibit great power in controlling the assembly. Removal of the E helix from BFR resulted in a destabilized protein that could only assemble into a dimer⁸¹. However, Luzzago and coworkers⁸² reported that the E helices along the four-fold symmetry axis are not essential for human ferritin H-chain assembly and proposed that the ferritin can assemble into a cage with the E helix either flipped out or flipped inside the central cavity. These conflicting results suggest that the role the E helix plays in ferritin self-assembly may be protein specific.

Arosio and coworkers⁸³ deleted twenty-two residues at the C-terminus in human H-chain ferritin including the E helix along with an unstructured C-terminal tail. This mutant assembled and maintained the ability to catalyze iron oxidation. However, if the mutation involved six additional C-terminal residues to include a total of twenty-eight amino acids, the mutant failed to assemble,⁸⁴ indicating that six amino acids at the end of the C-terminal tip of the D helix are essential for human H-chain ferritin self-assembly. The unstructured C-terminal tail of human ferritin was further explored by Ingrassia *et al*⁸⁵ who found that modification of the last six-residues had no major effect on the physical properties of ferritin, however the solubility and assembly decreased progressively with the extension of the tail suggesting that this part of the protein plays some, although minor, role in assembly.

Modifications at the N-terminus and the loop between the B and C helices of maxi-ferritins have also been investigated to examine the role that these regions play in protein assembly. Arosio and coworkers⁸³ engineered human H-chain ferritin by deleting the first thirteen residues at the N-terminus and found that the protein could still assemble and also catalyze iron oxidation. This is not particularly surprising as the N-terminus points outside of the protein cage and has little protein-protein overlap in most family members. Yoshizawa *et al*⁸⁶ engineered two mutants of horse L-chain apoferritin by deleting four and eight residues at the N-terminus. They found that one mutant dissociated easily to subunit dimers at or below pH 2.0 and proposed that the inter- and intra-subunit hydrogen bonds of the N-terminus are responsible for the lower endurance against acidic denaturation. Another study investigated the role of the BC loop by deleting two residues (L82 and I85) from human H-chain ferritin. These simple mutations resulted in disassembly of the cage probably resulting from disruption of the dimer assembly intermediate (see above).

Further modification of bacterial proteins were performed by Fan *et al*⁸¹, where several mutants were constructed by swapping the E helix from *E. coli* BFR and the BC helix from DPS to determine how these two structural elements affect the stability and oligomerization of the two proteins with respect to each other. The E helix from the C-terminus of BFR was found to exhibit great control in the assembly. Fusion of the BFR E helix to DPS resulted in a protein that discretely assembles into a 12-mer with a size uniquely intermediate between the two parents. The BC helix from DPS, on the other hand, plays less of a role in stabilizing oligomerization. Removal of the BC helix from DPS displayed insignificant changes in the oligomerization state and BFR remained as a mixture of dimer and 24-mer despite addition of the BC helix.

Mutation-based studies can help to understand the role that certain residues, and hence specific interactions play in ferritin protein self-assembly, or its underlying folding, in addition to identifying potentially relevant assembly intermediates. Santambrogio *et al*⁸⁷ reported that the renaturation of human H-chain ferritins mutated around the four-fold, three-fold and two-fold axes (L169R at the four-fold axis, D131I/E134F at the three-fold axis, I85C at the two-fold axis) yielded assembly intermediates ranging from monomers and dimers to all types of higher oligomers. Any of these intermediates could be induced to assemble into 24-mers either through concentration or by co-renaturing them with wild-type H-ferritin. However, as all pathways required initial assembly of dimers, these were proposed to be the essential intermediates for assembly.

Kilic *et al*⁸⁸ modified the bacterioferritin from *R. capsulatus* by site-directed mutagenesis to result in proteins that assembled into discrete dimers. The amino acids Glu128 and Glu135, located at the three-fold interface were mutated to either alanine

or arginine to determine their role in stabilizing the 24-mer. It was found that E128A/E135A and E128R/E135R double mutants formed stable dimers, strongly suggesting that the interactions involving Glu128 and Glu135 contribute significantly to the stabilization of the assembled 24-meric cage. The X-ray structures of *E. coli* and *R. capsulatus* BFR ^{56,89} reveal that the conserved Glu128 and Glu135 on one monomer are most likely interacting with Arg61 and the N-terminal amine of another monomer. The importance of salt bridge formation between Glu128 and Arg61, and its significance in self-assembly of the cage in *E. coli* bacterioferritin was also reported by Zhang *et al* ⁹⁰. It was observed that E128A and R61A single mutants formed stable dimers in solution. Interestingly, R61A possessed higher thermostability than the wild type bacterioferritin. This report also disclosed that two other key amino acids; Y114 at the three-fold axis and R30 located at the two-fold axis play significant roles in self-assembly.

The importance of water pockets at the protein-protein interfaces defined by the two-fold symmetry axes was demonstrated by Ardejani *et al* ⁹¹. Assisted by computational analysis, it was proposed that stabilizing point mutations could be achieved by bridging the water pocket associated with N23. All three predicted stabilizing mutations, N23L, N23F and N23W, assembled into cages and possessed a higher thermal stability than wild type BFR. One of the mutants, N23F, appeared to push the oligomerization state toward the 24-mer at the expense of the dimer, whereas the remaining two mutants had an opposite effect. These results, taken together, suggested that mutation at the two-fold interface could possibly result in a more stable but “unproductive” dimer which forms a geometry incompatible with cage formation. This proposal complicates the mechanism of assembly, but potentially explains previously unresolved experimental data (see above).

1.5.2 Mini-ferritin self-assembly

1.5.2.1 Proposed pathways of mini-ferritin assembly

Compared to that of the maxi-ferritins, the self-assembly of dodecameric mini-ferritins has not been as widely explored. One of the most extensively studied mini-ferritins is the *Mycobacterium smegmatis* Dps protein. Vijayan and coworkers proposed that the protein cage can be assembled in three different ways. First, the monomers could form two-fold symmetric dimers, which would later assemble into a dodecamer with 32 symmetry. The other two self-assembly mechanisms include forming either of the two types of three-fold symmetric trimers followed by assembly into the tetrahedral dodecamer⁹².

Supporting a trimer-dependent assembly mechanism, the presence of these intermediates was experimentally observed. Two stable oligomeric forms of *M. smegmatis* Dps, a trimer and a dodecamer were detected by Gupta *et al*⁹³. The conversion between the trimeric and dodecameric intermediates could be achieved by incubation at 37 °C for 12 h. Interestingly, it was found that these two oligomeric forms have different DNA binding affinities which may be due to a simple multivalent effect⁹⁴.

1.5.2.2 Mutation related studies on the effect of mini-ferritins self-assembly

The role of the N- and C-termini of *M. smegmatis* Dps in the self-assembly of the dodecameric protein cage was examined by Vijayan and coworkers⁹⁵. A protein with sixteen C-terminal residues deleted could assemble whereas if twenty-six residues of the C-terminus were deleted, no cage was observed, emphasizing the importance of the C-terminus in self-assembly. Deletion of the short N-terminal tail resulted in a protein that formed trimer instead of dodecamer in solution but could fully assemble

when crystallized, emphasizing that this region also is important. Recalling that the N-terminus and C-terminus of DPS lie at the ferritin-like and Dps-like three-fold symmetry axes, respectively, and that the trimer has been implicated as a DPS folding intermediate (see above), these results can shed light on the assembly mechanism when taken together. The fact that disruption of both these interfaces results in proteins with impaired cage formation implies a near-equal role for these two types of trimers in the mechanism of self-assembly.

The same research group identified clusters of amino acids on various interfaces of MsDps and predicted key residues that could disrupt the assembly upon mutation^{96,97}. Two mutations, E146A at the ferritin-like trimer interface and F47E at the Dps-like trimer interface, were constructed in an attempt to reinforce the predictions. Despite predicting that E146A would impair trimer formation, the mutant could form either trimer or dodecamer at different temperatures. On the other hand, F47 was found to be crucial for dodecamerization as the single mutant F47E and double mutant E146AF47E were observed as trimer and monomer respectively in solution with no observed cage. These results, although with some caveats, suggest that the ferritin-like trimer may play a more important role in the assembly mechanism.

Unlike maxi-ferritins, the loop between the B and C helices in mini-ferritins forms a short helix. Fan *et al*⁸¹ reported that the deletion of BC helix in *E. coli* Dps results in a mutant that can still form a cage, indicating that the BC helix is not essential to self-assembly.

1.5.2.3 pH effects on the self-assembly of mini-ferritin

The stability of *L. innocua* Dps protein in acidic conditions was investigated by Chiaraluce *et al*⁹⁸ and the protein was found to be extremely pH stable. Dissociation into dimers was observed below pH 2.0 and further dissociation to monomers with

significant secondary structure loss was observed at pH 1.0⁹⁸. The study suggested that the basic structural unit is dimer which, because it is resistant to pH, is held together mainly by hydrophobic interactions.

Examination of the pH stability of *M. smegmatis* Dps by Ceci *et al*⁹⁹ revealed that MsDps dissociated reversibly into dimers at conditions above pH 7.5 and below 6.0. Furthermore, dimers dissociate to monomers at pH 4.0. The two dissociation steps were attributed to the destruction of salt bridges between Glu 157 and Arg 99 and Asp 66 and Lys36 located at the three fold symmetry axes and across the dimer interface respectively.

1.6 Research perspectives

The structures, functions and self-assembly of ferritins and characteristics of protein-protein interactions have been discussed in this chapter. Determination of the key interfacial amino acid residues that play the key role in the protein-protein interactions that govern the formation of 24-mer bacterioferritin and 12-mer Dps protein nano-cages has been studied in this thesis.

1.7 References

1. Stryer L. Biochemistry. New York: W. H. Freeman.; 1995
2. Seringhaus M, Gerstein M. Chemistry Nobel rich in structure. Science 2007;315(5808):40-41.
3. Dobson CM. Protein-misfolding diseases: Getting out of shape. Nature 2002;418(6899):729-730.
4. [Anon]. So much more to know. Science 2005;309(5731):78-102.
5. Jin L, Wells JA. Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting. Protein Sci 1994;3(12):2351-2357.
6. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. Journal of Molecular Biology 1998;280(1):1-9.
7. Moreira IS, Fernandes PA, Ramos MJ. Hot spots--a review of the protein-protein interface determinant amino-acid residues. Proteins 2007;68(4):803-812.
8. Fletcher S, Hamilton AD. Protein-protein interaction inhibitors: small molecules from screening techniques. Curr Top Med Chem 2007;7(10):922-927.

9. Calhoun JR, Bell CB, 3rd, Smith TJ, Thamann TJ, DeGrado WF, Solomon EI. Oxygen reactivity of the biferrous site in the de novo designed four helix bundle peptide DFsc: nature of the "intermediate" and reaction mechanism. *J Am Chem Soc* 2008;130(29):9188-9189.
10. Hill RB, Raleigh DP, Lombardi A, DeGrado WF. De novo design of helical bundles as models for understanding protein folding and function. *Acc Chem Res* 2000;33(11):745-754.
11. Papapostolou D, Smith AM, Atkins EDT, Oliver SJ, Ryadnov MG, Serpell LC, Woolfson DN. Engineering nanoscale order into a designed protein fiber. *P Natl Acad Sci USA* 2007;104(26):10853-10858.
12. Grueninger D, Treiber N, Ziegler MO, Koetter JW, Schulze MS, Schulz GE. Designed protein-protein association. *Science* 2008;319(5860):206-209.
13. Bromley EHC, Channon K, Moutevelis E, Woolfson DN. Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *Acs Chem Biol* 2008;3(1):38-50.
14. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 1996;93(1):13-20.
15. Cunningham BC, Wells JA. Rational Design of Receptor-Specific Variants of Human Growth-Hormone. *P Natl Acad Sci USA* 1991;88(8):3407-3411.
16. Clackson T, Wells JA. A Hot-Spot of Binding-Energy in a Hormone-Receptor Interface. *Science* 1995;267(5196):383-386.
17. Ma B, Elkayam T, Wolfson H, Nussinov R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A* 2003;100(10):5772-5777.
18. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins-Structure Function and Genetics* 2001;43(2):89-102.
19. Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins* 1997;28(4):494-514.
20. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. *Proteins-Structure Function and Genetics* 2002;47(3):334-343.
21. Bahadur RP, Chakrabarti P, Rodier F, Janin J. Dissecting subunit interfaces in homodimeric proteins. *Proteins* 2003;53(3):708-719.
22. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci* 1997;6(1):53-64.
23. Hirakawa H, Muta S, Kuhara S. The hydrophobic cores of proteins predicted by wavelet analysis. *Bioinformatics* 1999;15(2):141-148.
24. Janin J, Miller S, Chothia C. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol* 1988;204(1):155-164.
25. Janin J, Chothia C. The structure of protein-protein recognition sites. *J Biol Chem* 1990;265(27):16027-16030.
26. Ippolito JA, Alexander RS, Christianson DW. Hydrogen bond stereochemistry in protein structure and function. *J Mol Biol* 1990;215(3):457-471.
27. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 1994;238(5):777-793.
28. Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* 1997;10(9):999-1012.

29. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* 2000;10(2):153-159.
30. Jones S, Thornton JM. Protein-Protein Interactions - a Review of Protein Dimer Structures. *Prog Biophys Mol Bio* 1995;63(1):31-&.
31. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. *Proteins* 2002;47(3):334-343.
32. Chothia C. Hydrophobic Bonding and Accessible Surface-Area in Proteins. *Nature* 1974;248(5446):338-339.
33. Betts MJ, Sternberg MJ. An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Eng* 1999;12(4):271-283.
34. Marsh JA, Teichmann SA. Relative Solvent Accessible Surface Area Predicts Protein Conformational Changes upon Binding. *Structure* 2011;19(6):859-867.
35. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol* 1999;285(5):2177-2198.
36. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* 1999;285(5):2177-2198.
37. Larsen TA, Olson AJ, Goodsell DS. Morphology of protein-protein interfaces. *Structure* 1998;6(4):421-427.
38. Uchida M, Klem MT, Allen M, Suci P, Flenniken M, Gillitzer E, Varpness Z, Liepold LO, Young M, Douglas T. Biological containers: Protein cages as multifunctional nanoplateforms. *Adv Mater* 2007;19(8):1025-1042.
39. Kim KK, Kim R, Kim SH. Crystal structure of a small heat-shock protein. *Nature* 1998;394(6693):595-599.
40. Liepold LO, Revis J, Allen M, Oltrogge L, Young M, Douglas T. Structural transitions in Cowpea chlorotic mottle virus (CCMV). *Phys Biol* 2005;2(4):S166-172.
41. Kong LB, Siva AC, Rome LH, Stewart PL. Structure of the vault, a ubiquitous cellular component. *Structure* 1999;7(4):371-379.
42. Kowalski MP, Dubouix-Bourandy A, Bajmoczy M, Golan DE, Zaidi T, Coutinho-Sledge YS, Gygi MP, Gygi SP, Wiemer EA, Pier GB. Host resistance to lung infection mediated by major vault protein in epithelial cells. *Science* 2007;317(5834):130-132.
43. Scheffer GL, Schroeijers AB, Izquierdo MA, Wiemer EA, Scheper RJ. Lung resistance-related protein/major vault protein and vaults in multidrug-resistant cancer. *Curr Opin Oncol* 2000;12(6):550-556.
44. Yu Z, Fotouhi-Ardakani N, Wu L, Maoui M, Wang S, Banville D, Shen SH. PTEN associates with the vault particles in HeLa cells. *J Biol Chem* 2002;277(43):40247-40252.
45. Bode SA, Minten IJ, Nolte RJ, Cornelissen JJ. Reactions inside nanoscale protein cages. *Nanoscale* 2011;3(6):2376-2389.
46. Andrews SC, Arosio P, Bottke W, Briat JF, Vondarl M, Harrison PM, Laulhere JP, Levi S, Lobreaux S, Yewdall SJ. Structure, Function, and Evolution of Ferritins. *J Inorg Biochem* 1992;47(3-4):161-174.
47. Harrison PM, Arosio P. Ferritins: Molecular properties, iron storage function and cellular regulation. *Bba-Bioenergetics* 1996;1275(3):161-203.
48. Andrews SC. The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim Biophys Acta* 2010;1800(8):691-705.

49. Laufberger V. Contribution to the technique of renin seclusion. *Cr Soc Biol* 1937;126:107-109.
50. Lawson DM, Artymiuk PJ, Yewdall SJ, Smith JMA, Livingstone JC, Treffry A, Luzzago A, Levi S, Arosio P, Cesareni G, Thomas CD, Shaw WV, Harrison PM. Solving the Structure of Human H-Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature* 1991;349(6309):541-544.
51. Hempstead PD, Yewdall SJ, Fernie AR, Lawson DM, Artymiuk PJ, Rice DW, Ford GC, Harrison PM. Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *Journal of Molecular Biology* 1997;268(2):424-448.
52. Granier T, Gallois B, d'Estaintot AD, Dautant A, Chevelier JM, Mellado JM, Beaumont C, Santambrogio P, Arosio P, Precigoux G. Structure of mouse L-chain ferritin at 1.6 angstrom resolution. *Acta Crystallogr D* 2001;57:1491-1497.
53. Vanwuytswinkel O, Briat JF. Conformational-Changes and in-Vitro Core-Formation Modifications Induced by Site-Directed Mutagenesis of the Specific N-Terminus of Pea Seed Ferritin. *Biochem J* 1995;305:959-965.
54. Wardrop AJ, Wicks RE, Entsch B. Occurrence and expression of members of the ferritin gene family in cowpeas. *Biochem J* 1999;337:523-530.
55. Dautant A, Meyer JB, Yariv J, Precigoux G, Sweet RM, Kalb AJ, Frolow F. Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallogr D* 1998;54:16-24.
56. Cobessi D, Huang LS, Ban M, Pon NG, Daldal F, Berry EA. 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 D* 2002;58:29-38.
57. Macedo S, Romao CV, Mitchell E, Matias PM, Liu MY, Xavier AV, LeGall J, Teixeira M, Lindley P, Carrondo MA. The nature of the di-iron site in the bacterioferritin from *Desulfovibrio desulfuricans*. *Nat Struct Biol* 2003;10(4):285-290.
58. Theil EC. Ferritin protein nanocages use ion channels, catalytic sites, and nucleation channels to manage iron/oxygen chemistry. *Curr Opin Chem Biol* 2011;15(2):304-311.
59. Andrews SC. Iron storage in bacteria. *Adv Microb Physiol* 1998;40:281-351.
60. Andrews SC, Robinson AK, Rodriguez-Quinones F. Bacterial iron homeostasis. *Fems Microbiol Rev* 2003;27(2-3):215-237.
61. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004;25(13):1605-1612.
62. Theil EC. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annu Rev Biochem* 1987;56:289-315.
63. Crichton RR, Bryce CFA. Subunit Interactions in Horse Spleen Apoferritin - Dissociation by Extremes of Ph. *Biochem J* 1973;133(2):289-299.
64. Stefanini S, Cavallo S, Wang CQ, Tataseo P, Vecchini P, Giartosio A, Chiancone E. Thermal stability of horse spleen apoferritin and human recombinant H apoferritin. *Arch Biochem Biophys* 1996;325(1):58-64.
65. Treffry A, Bauminger ER, Hechel D, Hodson NW, Nowik I, Yewdall SJ, Harrison PM. Defining the roles of the threefold channels in iron uptake, iron

- oxidation and iron-core formation in ferritin: a study aided by site-directed mutagenesis. *Biochem J* 1993;296 (Pt 3):721-728.
66. Almiron M, Link AJ, Furlong D, Kolter R. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes & Development* 1992;6(12b):2646-2654.
 67. Grant RA, Filman DJ, Finkel SE, Kolter R, Hogle JM. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* 1998;5(4):294-303.
 68. Romão CV, Mitchell EP, McSweeney S. The crystal structure of *Deinococcus radiodurans* Dps protein (DR2263) reveals the presence of a novel metal centre in the N terminus. *JBIC Journal of Biological Inorganic Chemistry* 2006;11(7):891-902.
 69. Bellapadrona G, Stefanini S, Zamparelli C, Theil EC, Chiancone E. Iron Translocation into and out of *Listeria innocua* Dps and Size Distribution of the Protein-enclosed Nanomineral Are Modulated by the Electrostatic Gradient at the 3-fold "Ferritin-like" Pores. *Journal of Biological Chemistry* 2009;284(28):19101-19109.
 70. Crichton RR. Ferritin: structure, synthesis and function. *N Engl J Med* 1971;284(25):1413-1422.
 71. Munro HN, Linder MC. Ferritin: structure, biosynthesis, and role in iron metabolism. *Physiol Rev* 1978;58(2):317-396.
 72. Theil EC. The ferritin family of iron storage proteins. *Adv Enzymol Relat Areas Mol Biol* 1990;63:421-449.
 73. Takagi H, Shi D, Ha Y, Allewell NM, Theil EC. Localized unfolding at the junction of three ferritin subunits. A mechanism for iron release? *J Biol Chem* 1998;273(30):18685-18688.
 74. Jaenicke R, Rudolph R. Refolding and association of oligomeric proteins. *Methods Enzymol* 1986;131:218-250.
 75. Banyard SH, Stammers DK, Harrison PM. Electron-Density Map of Apoferritin at 2.8-Å Resolution. *Nature* 1978;271(5642):282-284.
 76. Stefanini S, Vecchini P, Chiancone E. On the Mechanism of Horse Spleen Apoferritin Assembly - a Sedimentation-Velocity and Circular-Dichroism Study. *Biochemistry* 1987;26(7):1831-1837.
 77. Gerl M, Jaenicke R. Mechanism of the self-assembly of apoferritin from horse spleen. Cross-linking and spectroscopic analysis. *Eur Biophys J* 1987;15(2):103-109.
 78. Gerl M, Jaenicke R, Smith JMA, Harrison PM. Self-Assembly of Apoferritin from Horse Spleen after Reversible Chemical Modification with 2,3-Dimethylmaleic Anhydride. *Biochemistry* 1988;27(11):4089-4096.
 79. Jin KS, Kim M, Rho Y, Ahn B, Jung S, Kim H, Ree M. pH-Dependent Structures of Ferritin and Apoferritin in Solution: Disassembly and Reassembly. *Biomacromolecules* 2011;12(5):1629-1640.
 80. Andrews SC, Smith JMA, Hawkins C, Williams JM, Harrison PM, Guest JR. Overproduction, Purification and Characterization of the Bacterioferritin of *Escherichia-Coli* and a C-Terminally Extended Variant. *Eur J Biochem* 1993;213(1):329-338.
 81. Fan R, Boyle AL, Cheong VV, Ng SL, Orner BP. A Helix Swapping Study of Two Protein Cages. *Biochemistry* 2009;48(24):5623-5630.

82. Luzzago A, Cesareni G. Isolation of Point Mutations That Affect the Folding of the H-Chain of Human Ferritin in *Escherichia-Coli*. *Embo J* 1989;8(2):569-576.
83. Levi S, Luzzago A, Cesareni G, Cozzi A, Franceschinelli F, Albertini A, Arosio P. Mechanism of Ferritin Iron Uptake - Activity of the H-Chain and Deletion Mapping of the Ferro-Oxidase Site - a Study of Iron Uptake and Ferro-Oxidase Activity of Human-Liver, Recombinant H-Chain Ferritins, and of 2 H-Chain Deletion Mutants. *Journal of Biological Chemistry* 1988;263(34):18086-18092.
84. Levi S, Luzzago A, Franceschinelli F, Santambrogio P, Cesareni G, Arosio P. Mutational Analysis of the Channel and Loop Sequences of Human Ferritin H-Chain. *Biochem J* 1989;264(2):381-388.
85. Ingrassia R, Gerardi G, Biasiotto G, Arosio P. Mutations of ferritin H chain C-terminus produced by nucleotide insertions have altered stability and functional properties. *J Biochem* 2006;139(5):881-885.
86. Yoshizawa K, Mishima Y, Park SY, Heddle JG, Tame JRH, Iwahori K, Kobayashi M, Yamashita I. Effect of N-terminal residues on the structural stability of recombinant horse L-chain apoferritin in an acidic environment. *J Biochem* 2007;142(6):707-713.
87. Santambrogio P, Pinto P, Levi S, Cozzi A, Rovida E, Albertini A, Artymiuk P, Harrison PM, Arosio P. Effects of modifications near the 2-, 3- and 4-fold symmetry axes on human ferritin renaturation. *Biochem J* 1997;322:461-468.
88. Kilic MA, Spiro S, Moore GR. Stability of a 24-meric homopolymer: Comparative studies of assembly-defective mutants of *Rhodobacter capsulatus* bacterioferritin and the native protein. *Protein Sci* 2003;12(8):1663-1674.
89. Frolov F, Kalb AJ, Yariv J. Structure of a Unique Twofold Symmetrical Heme-Binding Site. *Nat Struct Biol* 1994;1(7):453-460.
90. Zhang Y, Raudah S, Teo H, Teo GWS, Fan RL, Sun XM, Orner BP. Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* 2010;285(16):12078-12086.
91. Ardejani MS, Li NX, Orner BP. Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* 2011;50(19):4029-4037.
92. Roy S, Gupta S, Das S, Sekar K, Chatterji D, Vijayan M. X-ray Analysis of *Mycobacterium smegmatis* Dps and a Comparative Study Involving Other Dps and Dps-like Molecules. *J Mol Bio* 2004;339(5):1103-1113.
93. Gupta S, Chatterji D. Bimodal protection of DNA by *Mycobacterium smegmatis* DNA-binding protein from stationary phase cells. *Journal of Biological Chemistry* 2003;278(7):5235-5241.
94. Kiessling LL, Gestwicki JE, Strong LE. Synthetic multivalent ligands as probes of signal transduction. *Angew Chem Int Ed Engl* 2006;45(15):2348-2368.
95. Roy S, Saraswathi R, Gupta S, Sekar K, Chatterji D, Vijayan M. 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* 2007;370(4):752-767.

96. Chowdhury RP, Saraswathi R, Chatterji D. Mycobacterial Stress Regulation: The Dps "Twin Sister" Defense Mechanism and Structure-Function Relationship. *Iubmb Life* 2010;62(1):67-77.
97. Chowdhury RP, Vijayabaskar MS, Vishveshwara S, Chatterji D. Molecular mechanism of in vitro oligomerization of Dps from *Mycobacterium smegmatis*: mutations of the residues identified by "interface cluster" analysis. *Biochemistry* 2008;47(42):11110-11117.
98. Chiaraluce R, Consalvi V, Cavallo S, Ilari A, Stefanini S, Chiancone E. The unusual dodecameric ferritin from *Listeria innocua* dissociates below pH 2.0. *Eur J Biochem* 2000;267(18):5733-5741.
99. Ceci P, Ilari A, Falvo E, Giangiacomo L, Chiancone E. Reassessment of protein stability, DNA binding, and protection of *Mycobacterium smegmatis* Dps. *Journal of Biological Chemistry* 2005;280(41):34776-34785.

CHAPTER 2 Alanine shaving mutagenesis studies on a maxi-ferritin, bacterioferritin from *E. coli* (BFR) *

2.1 Introduction

Many of the ferritin proteins and their structural homologues assemble into multi-subunit, hollow, nano-scale cages ¹. Ferritins and other cage proteins have been the focus of much recent attention as part of bioorthogonal chemical reaction methodology development ², drug delivery studies ^{3,4} as platforms to generate nano-structured materials⁵, and in basic biophysical studies as they are composed of monomers folded into a four-helix bundle motif, a fold well-studied by the protein *de novo* design community. Thus, ferritins could act as important model systems for developing the fundamentals of how to expand the dimensionality of rational protein engineering into the realm of self-assembly.

Genetic and biochemical evidence suggests a direct role for ferritins in human diseases including those associated with aging and protein misfolding. One of the roles of ferritins in the cell is to sequester iron in a mineralized form thereby protecting the cell from iron-induced oxidative damage. Human ferritin is made up a heavy chain (H) and a light chain (L). Although it is possible for homo-oligomeric nano-cages of these proteins to form, the activity of the two monomers is complementary. The H-chain possesses ferrioxidase activity, facilitating rapid iron uptake, whereas the L-chain stabilizes the mineralized iron core ⁶. Iron homeostasis is important for normal cellular function and iron usage is especially enhanced in the brain due to increased respiration demands, biosynthesis of neurotransmitters, and myelogenesis ⁷. Ferritin upregulation has been shown to play a role in a number of

* Parts of this chapter have been published previously with the author of this thesis as the first author. *J. Biol. Chem.*, **2010**, 285, 12078-12086.

autoimmune diseases such as lupus and rheumatoid arthritis. In addition, the symptoms of multiple sclerosis (MS), as disease caused by the demyelination of neurons, can be reversed in mice by the injection of apoferritin⁸. Neuroferritinopathy generates motor defects and tremors, possibly causing it to be misdiagnosed as Huntington's or Parkinson's Disease. The condition seems to be caused by single nucleotide missense mutations in the ferritin L-chain gene resulting in a mistranslated and extended C-terminus. These mutations destabilize the folded protein and result in the formation of brain inclusions made up of iron and misfolded, aggregated protein^{7,9,10}.

Structural studies have revealed that of the ferritins that naturally self-assemble into nano-capsules there are two types. Mini-ferritins form hollow spheres composed of twelve monomers and maxi-ferritins are composed of twenty-four monomers. Although their monomers share similar four-helix bundle tertiary structure, mini-ferritins like DNA-binding protein from starved cells (DPS)^{11,12} have tetrahedral symmetry while maxi-ferritins like *E. coli* bacterioferritin (BFR)¹³ have octahedral symmetry (Figure 2.1). Structural comparisons have been made, and yet it is still an open question as to what controls the differing nano-architecture of these proteins. Little modern research has been undertaken to understand how the “structural energetics” of these systems define their quaternary structure¹⁴. Mutagenesis and design studies targeting the ferritins will not only help to understand the fundamentals of protein folding, quaternary structure formation, self-assembly, and misfolded disease states, they may also provide insight into helping to establish dynamic drug delivery systems¹⁵ or the development of new nanomaterials using these proteins¹⁶. In this chapter, we focus on the study of maxi-ferritin, bacterioferritin from *E. coli*.

2.2 Bacterioferritin

There are two types of iron-storage protein in bacteria, the heme-free ferritins and heme-containing bacterioferritins. Bacterioferritins and ferritins share only slight sequence similarity but retain similar structural and functional properties. They both are composed of twenty-four identical or similar subunits (~ 19 kDa) that assemble to form approximately spherical proteins (outer diameter ~ 120 Å) with a hollow cavity (inner diameter ~ 80 Å) (Figure 2. 1). The cavity can accommodate up to ~ 4500 Fe atoms in the form of a ferric-hydroxyphosphate core. Each subunit folds into a four-helix bundle which carries the active site (ferroxidase center) of the protein. Each bacterioferritin contains twelve protoporphyrin IX heme groups residing at the two-fold interfaces between the two-fold related subunits. The function of the heme is still unknown¹⁷.

Figure 2.1 The space-filled structure of human H ferritin (*left*) and a view with half of the cage is removed so as to show the inner cavity (*right*) (PDB ID: 1fha). This figure is adopted from PDB-101: molecule of the month (November 2002 by David Goodsell).

2.2.1 The structure of *E. coli* bacterioferritin (BFR)

The nucleotide sequence of BFR gene of *E. coli* K-12 was reported by S. C. Andrew in 1989¹⁸. The gene is comprised of 158 amino acid residues with a subunit MW of ~ 18.5 kDa (Figure 2.2 c).

2.2.1.1 The monomeric structure of BFR

The 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 and helix E (Leu146-Gln151) at the C-terminus which lies almost perpendicular (78°) to the bundle axis (Figure 2. 2). The AB helix pair is anti-parallel to the CD pair with loop (Leu65-Asp82) connecting helix B and C.

Figure 2.2 The monomeric structure and amino acid sequence of *E. coli* bacterioferritin (BFR). **(a)** The four-helix bundle subunit is displayed as gray pipe helices and black loops view perpendicular to the bundle axis. **(b)** The four-helix bundle subunit is displayed as gray ribbon helices and black loops viewed down the bundle axis. **(c)** The primary and secondary structure of BFR, α -helices (bars) and loops (dot lines).

2.2.1.2 The quaternary structure and assembly of subunits to octahedral symmetry

Crystallographic studies have shown that BFR forms a well-packed, mostly spherical cage and is composed of twenty-four structurally equivalent subunits arranged with 432 (O_h point group) symmetry¹⁹ (Figure 2.3). When the subunits form a cage-like structure, the E helix at the C-terminus points toward the core and the N-terminus is oriented towards the outside of the cage. The A and C helices are positioned on the outer surface, whereas the B and D helices face the inner cavity.

The BC loop adopts a random coil conformation and is on the outside of the cage. In the octahedral cage structure, each BFR subunit interacts with six surrounding monomers through three types of symmetry-related interfaces. There are twelve dimerization interaction sites at two-fold axes, eight trimerization interaction sites at three-fold axes and six tetramerization sites at the four-fold axes (Figure 2. 3).

Figure 2.3 Symmetry related protein-protein interfaces in the packing of the *E. coli* BFR protein cage. Subunits are represented in ribbon. Views are from the outside of the cage towards the inside along the (A) two-fold, (B) three-fold and (C) four-fold symmetry axes. Note that subunits packing around the two-fold axes are in anti-parallel pairs (A). The position of subunits in relation to the symmetry axes is shown in the center as outlined rectangles.

2.2.2 Sequence and structural alignment of bacterioferritins

The sequences of several members of the bacterioferritin protein family are aligned (Figure 2.4). From this analysis, it is clear that each of ferrioxidase center residues (E18, E51, H54, E94, E127, H130) is identical in all sequences in the alignment. Residues residing at the symmetry related interfaces that are subjected to mutation to alanine described in this chapter are highly but not absolutely conserved (see below).

Figure 2.4 Alignment of amino acid sequences of several bacterioferritins. The residues mutated in this study are underlined and indicated with their numeric position in *E. coli* BFR. The ferroxidase centre residues are indicated with "o". An "*" indicates that the residues are identical in all sequences in the alignment, ":" indicates conserved substitutions, and "." indicates semiconserved substitutions. Different colors indicate the physiochemical characteristics of the amino acids. The sequences were aligned with the ClustalW2 program²⁰ that is available online (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>)

Structural alignment of all the bacterioferritins with available crystal structures revealed that they all fold into a four-helix bundle with highly superimposable backbones (average RMSD of 0.8 Å, z-score above 20) (Figure 2. 5) even for proteins with a low sequence identity (Table 2.1). *D. desulfuricans* bacterioferritin (cyan), with the least sequence identity, has slightly longer N-terminus with no loop extension but extremely longer C-terminus than other proteins. Sequence alignment also shows C-terminal extensions after helices E to occur in almost all the bacterioferritins (Figure 2.5).

Figure 2.5 The sequence-based superimposition of monomers of bacterioferritins. *Black*, *E. coli*; *red*, *P. aeruginosa*B; *magenta*, *A. vinelandii*; *yellow*, *R. capsulatus*; *green*, *M. tuberculosis*; *blue*, *M. smegmatis*; *purple*, *B. Melitensis*; *brown*, *R. Sphaeroides*; *cyan*, *D. desulfuricans*. The figure is generated using the DaliLite online program²¹ (<http://www.ebi.ac.uk/Tools/dalilite/index.html>)

Proteins (PDB ID)	Z-Score	RMSD (Å)	Seq. Identity [%]
<i>P. aeruginosa</i> B (3isf)	28.0	0.5	70

<i>A. vinelandii</i> (2fkz)	27.8	0.4	66
<i>R. capsulatus</i> (ljgc)	26.0	0.8	48
<i>M. tuberculosis</i> (2wtl)	24.4	0.7	43
<i>M. smegmatis</i> (3bkn)	26.0	1.0	42
<i>B. Melitensis</i> (3fvy)	25.7	0.9	50
<i>R. Sphaeroides</i> (3gvy)	25.9	0.8	46
<i>D. desulfuricans</i> (Infv)	2.50	1.5	27
Average	23.3	0.8	49

Table 2.1 Analysis of structural alignment of *E. coli* BFR (2Y3Q) monomer to the other bacterioferritin momomers with available crystal structures.

2.3 Aims and objectives

As protein-protein interactions are essential for the self-assembly of ferritin cages, we sought to determine the role of specific key residues residing at interfaces between monomers. We employed alanine shaving mutagenesis ^{22,23} to explore whether “hot spots” exist at the protein-protein interfaces and how these affect self-assembly. These interfacial residues can be classified as being related to different symmetry elements within the assembled form of the protein. Through inspection of the crystal structure of BFR ^{12,13} we focused our attention on interfacial salt bridges, due to their straightforward identification, and aromatic amino acids, because of their high hot spot propensity ^{24,25}. In addition, we performed virtual alanine scanning ^{26,27} to confirm the predicted importance of these residues. The selected amino acids were then subjected to mutagenesis to alanine and the resulting mutants were probed using transmission electron microscopy (TEM), size exclusion chromatography (SEC), dynamic light scattering (DLS), native polyacrylamide gel electrophoresis (native PAGE), and temperature dependent circular dichroism (CD) to determine the role that these individual interfacial interactions play in the stability and structure of the ferritin.

2.4 Materials and methods

2.4.1 Virtual alanine scanning of BFR

Computational prediction of energetically important amino acid residues on the protein-protein interfaces followed standard instructions using web tools developed by the Baker laboratory ²⁶ (available at <http://robetta.bakerlab.org/alaninescan>). The input consisted of the symmetry-related protein-protein interfaces described by the coordinates for the crystal structure of BFR (Protein Data Bank (PDB) ID: 1bfr). From these coordinates, two monomers on the two-fold axis, three monomers on the three-fold axis, and two monomers that are 90 degrees adjacent at the four-fold axis were used. The program defined hot spot residues if the $\Delta\Delta G$ for mutagenesis was greater than 1 kcal/mol.

2.4.2 Analysis of the protein-protein interactions at the symmetry related interfaces of BFR using PISA software

The analysis of the protein-protein interfaces of BFR was performed using the web-based software “Protein interfaces, surfaces and assemblies service” (PISA, European Bioinformatics Institute)²⁸. The PDB structure used as input file is 1BFR. The symmetry related interfaces are defined in Figure 2.12.

2.4.3 Synthesis of BFR and its mutants

2.4.3.1 Oligonucleotides design

The protein sequence of bacterioferritin from *E. coli* (PDB ID: 1bfr) was obtained from the Protein Data Bank (PDB). The protein sequence of wild type and mutants were converted to gene sequences by utilizing optimal codons for maximizing the expression level in *E. coli*. The optimized codons are listed below (Table 2.2). The full length gene was divided into short oligonucleotides which are composed of 27

bases with a 15-bases overlap at their 5' end and a 12-bases overlap at their 3' end with the adjacent oligonucleotides. The oligonucleotides were also designed to ensure that the GC content were within the range of 40-60%. The sense and antisense primers were designed for annealing with the pET32 Ek/LIC vector using ligation independent cloning.

The forward and reverse primers of the mutagenic oligonucleotide for each mutation were designed according to their positions and mutated to the GCA alanine codon. The mutant constructs for BFR R30A and BFR R61A were generated using the QuikChange site-directed mutagenesis kit (Stratagene) with the BFR construct as the template. The mutagenic primers using Quick Change method were designed according to the following considerations: (i) primers were between 25-45 bases in length with a melting temperature (T_m) of ≥ 78 °C; (ii) the desired mutation should be in the middle of the primer with 10 - 15 bases on both sides; (iii) the GC content of the primers should be within the range of 40-60% and should terminate with at least one C or G bases; (iv) the primers should be purified by fast protein liquid chromatography (FPLC) or polyacrylamide gel electrophoresis (PAGE) to ensure higher mutation efficiency.

All the oligonucleotides were synthesized by 1st BASE Pte. Ltd. (Singapore) using solid phase phosphoramidite technology. The sequences of all the oligonucleotides are listed in Appendix 1.

<i>amino acid</i>	<i>codon</i>	<i>amino acid</i>	<i>codon</i>	<i>amino acid</i>	<i>codon</i>	<i>amino acid</i>	<i>codon</i>
Ala	GCA	Gly	GGT	Leu	CTG	Ser	AGC
Arg	CGT	Gln	GAA	Lys	AAA	Thr	ACC
Asp	GAT	Glu	TAC	Met	ATG	Trp	TGG
Asn	AAC	His	GAT	Phe	TTC	Tyr	TAC
Cys	TGC	Ile	ATC	Pro	CCG	Val	GTG

Table 2.2 Codons, optimized for *E. coli* expression, used to translate protein sequences to gene sequences.

2.4.3.2 Gene preparation and cloning

Assembly PCR- The recombinant genes were synthesized in a two-step, total gene synthesis method combining dual asymmetrical PCR (DA-PCR) and overlap extension PCR (OE-PCR) as per Young and Dong²⁹ (Figure 2.6).

In the DA-PCR, every four consecutive oligonucleotides were used as templates, with the outer two oligonucleotides in four times molar excess to the inner ones. The gene synthesis reactions were performed by adding each of the outer two (2.0 µl of a 10 µM solution) and each of the inner oligonucleotides (0.5 µl of a 10 µM solution) to a solution made up of a dNTP mix (Fermentas, 2.0 µl of a 2 mM solution), pfu buffer (Fermentas, 2.0 µl, 10×), pfu DNA polymerase (Fermentas, 0.2 µl, 2.5 U/µl) and deionized water to bring the total volume to 20 µl. An initial denaturation (95 °C, 5 min) was followed by 28 cycles of denaturation (95 °C, 0.5 min), annealing (53 °C, 0.75 min), and extension (72 °C, 0.75 min) which was followed by a final extension (72 °C, 7 min).

In the subsequent OE-PCR, the gene synthesis reactions were performed by adding the reaction product solutions (1.0 µl) from each DA-PCR reaction to a solution made up of dNTP mix (Fermentas, 2.0 µl of a 2 mM solution), pfu buffer (Fermentas, 3.0 µl, 10×), pfu DNA polymerase (Fermentas, 0.3 µl, 2.5 U/µl) and deionized water to bring the total volume to 30 µl. An initial denaturation (95 °C, 5 min) was followed by 10 cycles of denaturation (95 °C, 1 min), annealing (43 °C, 1 min), and extension (72 °C, 1 min) which was followed by a final extension (72 °C, 5 min).

Figure 2.6 Schematic diagram of the two-step assembly PCR method using six oligonucleotides as an example. This figure is adapted from Young *et al*²⁹.

Amplification PCR- The assembled gene was amplified by PCR. The reaction was performed by mixing pfu buffer (Fermentas, 2.0 μ l, 10 $\times$), pfu polymerase (Fermentas, 0.2 μ l, 2.5 U/ μ l) and dNTP mix (Fermentas, 2.0 μ l of a 2 mM solution), the reaction product from the OE-PCR (2.0 μ l), forward and reverse primers (i.e. the end primers from gene synthesis, 1.0 μ l of a 10 μ M solution of each) and deionized water to bring the total reaction volume to 20 μ l. An initial denaturation (95 $^{\circ}$ C, 5 min) was followed by 28 cycles of denaturation (95 $^{\circ}$ C, 5 min), annealing (53 $^{\circ}$ C, 0.75 min), and extension (72 $^{\circ}$ C, 0.75 min) followed by a final extension (72 $^{\circ}$ C, 7 min). Products were visualized and analyzed by 2% agarose gel electrophoresis to ensure that the assembled gene was of the correct size.

Extension PCR- The product of amplification PCR (3.0 μ l) was added to a solution of sense and antisense primers (2.5 μ l of a 10 μ M solution of each), pfu buffer

(Fermentas, 5.0 μ l of 10 $\times$), pfu polymerase (Fermentas, 0.5 μ l, 2.5 U/ μ l) and dNTP mix (Fermentas, 5.0 μ l of a 2 mM solution), and deionized water to bring the total reaction volume to 50 μ l. An initial denaturation (95 $^{\circ}$ C, 5 min) was followed by 28 cycles of denaturation (95 $^{\circ}$ C, 5 min), annealing (53 $^{\circ}$ C, 0.75 min), and extension (72 $^{\circ}$ C, 0.75 min) followed by a final extension (72 $^{\circ}$ C, 7 min). Products were gel-purified with extraction from agarose using Perfectprep Gel Cleanup kit (Eppendorf).

Figure 2.7 Schematic diagram of amplification and extension PCR.

Preparation of Target Insert and Annealing to pET32 Ek/LIC Vector- The purified product from extension PCR (17.5 μ l) was treated with T4 DNA polymerase (Novagen, 0.5 μ l, 2.5 U/ μ l) in a solution of dATP (Novagen, 2.5 μ l of a 25 mM solution), NEB Buffer 2 (Novagen, 2.5 μ l, 10 $\times$), BSA (Novagen, 0.25 μ l, 100 $\times$) and deionized water to bring the total reaction volume to 25 μ l. The solution was incubated (22 $^{\circ}$ C, 30 min) and the enzyme was subsequently inactivated (75 $^{\circ}$ C, 20 min). Annealing was performed by incubating the polymerase treated insert (5.0 μ l) with pET-32 Ek/LIC vector (Novagen, 2.0 μ l, 22 $^{\circ}$ C, 30 min). EDTA (Novagen Chemicals, 2.3 μ l of a 25 mM solution) was subsequently added and the solution was further incubated (22 $^{\circ}$ C, 1 h).

Transformation reaction- The plasmids were then transformed by electroporation (1.8 kV, 1 mm pathlength) into Novablue electro-competent cells (Novagen, 50 μ l) and incubated in LB media (1 mL) immediately with shaking (37 $^{\circ}$ C, 150 rpm) for

one hour. An appropriate volume of cells was plated on LB agar plates containing antibiotics (50 µg/mL carbenicillin and 12.5 µg/mL tetracycline) and incubated at 37 °C overnight.

Colony PCR screening- Ten colonies were randomly picked from the antibiotic-containing plates and transferred to sterile water (50 µl). The bacteria cells were dispersed by vortex and boiled (99 °C, 5 min) to lyse the cells and denature DNases. The supernatant after the cell debris was cleared by centrifugation was used as the template solution for the colony PCR. The supernatant (5.0 µl) was added to a solution of Taq polymerase buffer (Fermentas, 1.0 µl, 10×), Taq DNA polymerase (Fermentas, 0.05 µl, 3.0 U/µl), forward and reverse primers (i.e. the end primers from gene synthesis, 1.0 µl of a 10 µM solution of each), dNTP mix (Fermentas, 1.0 µl of a 2 mM solution) and deionized water to bring volume to 10 µl. The PCR was performed in a thermal cycler program, consisting of 95 °C for 5 min, 28 cycles of 95 °C for 30 s, 55 °C for 45 s, 72 °C for 45 s and a final extension step at 72 °C for 7 min. Products were visualized and analyzed by 2% agarose gel to detect the positive colonies.

The positive colonies were grown in LB and the plasmids were extracted using Miniprep (QIAprep Spin Miniprep Kit). Samples were sent to 1st Base Pte. Ltd. (Singapore) for sequencing to confirm the correctness of the genes.

Figure 2.8 Schematic diagram of preparation of extended single stranded complementary ends and subsequent ligation independent cloning.

2.4.3.3 Plasmid preparation using site directed mutagenesis

Mutant strand synthesis reaction-The basic procedure for preparation of plasmid with target mutation utilizes a supercoiled dsDNA vector with two oligonucleotide primers which contain the desired mutation. The reaction was performed by adding 10x pfu reaction buffer (2.0 ul), pfu Turbo DNA polymerase (0.5 ul, 2.5U/ul), dsDNA template of wild type BFR (2.0 ul, 60 ng/ul), forward and reverse oligonucleotide primer containing the target mutation (1.0 ul, 10 mM), dNTP mix (Fermentas, 2.0 µl of a 2 mM solution), and deionized water to bring the total reaction volume to 20.0 ul. An initial denaturation (95 °C, 1 min) was followed by 18 cycles of denaturation (95 °C, 0,5 min), annealing (55 °C, 1 min), and extension (68 °C, 6 min) followed by a final extension (72 °C, 10 min).

DpnI digestion of the amplification product- The amplification PCR product (5.0 ul) was then treated with DpnI restriction enzyme (0.6 ul, 20 U/ul) together with NEBuffer 4 (New England Biolabs, 1.0 ul) and deionized water to bring the total reaction volume to 10.0 ul. The solution was gently and thoroughly mixed by

pipetting up and down several times and was immediately incubated at 37 °C for 2 hours to digest the parental nonmutated supercoiled dsDNA.

The plasmid was then transformed by electroporation into Novablue electro-competent cells (Novagen). The resulting antibiotic resistant colonies were screened by colony PCR and the plasmids were verified by sequencing (1st Base) after miniprep (QIAprep Spin Miniprep Kit).

Figure 2.9 Schematic diagram of site-directed mutagenesis. This figure is adapted from the instruction manual of QuikChange II XL Site-Directed Mutagenesis kit.

2.4.4 Protein expression and purification

Plasmids harboring the desired genes were transformed into *E. coli* BL21(DE3) (Novagen) cells. Expression culture in 500 ml of LB was induced with isopropyl-D-thiogalactopyranoside (IPTG) (Fermentas, 2 ml of a 0.1 M solution), and the culture was further incubated (3 h, 30 °C). The cells were isolated by centrifugation and lysed by sonication. The soluble protein was applied to nickel-nitrilotriacetic acid resin (Ni-NTA) (Qiagen) and eluted by affinity tag cleavage following incubation with enterokinase (New England Biolabs, 8 μ l, 2 μ g ml⁻¹, 36 h, 4 °C). The protein was concentrated via ultrafiltration (Sartorius Vivaspin 6), and SDS-PAGE was carried out to verify the purity of the protein. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy of the monomers confirmed the identity of the purified proteins. The matrix solution is 20 mg/ml sinapinic acid in ACN:

water: TFA (50:50:0.1).

2.4.5 Transmission electron microscopy (TEM)

A drop of a solution of purified protein (5 μl of a $\sim 200 \mu\text{g ml}^{-1}$ solution) was placed on a piece of parafilm and a circular carbon-coated copper grid was placed on top of the drop (1 min). Excess sample was removed by wicking with a piece of filter paper, and the grid was placed on a drop of either uranyl acetate or phosphotungstic acid (5 μl of a 1% w/v solution, 1 min) and excess solution was again removed by wicking with filter paper. TEM data was obtained using a JOEL JEM-1400 microscope operating at 120 keV. TEM micrographs were analyzed using Image J (Freeware, National Institutes of Health, USA) to measure the longest diameter of the protein cages. For each protein, 100 particles were measured.

2.4.6 Size exclusion chromatography (SEC)

Chromatography experiments were performed on a $\ddot{\text{A}}\text{KTAFPLC}^{\text{TM}}$ (GE Healthcare Life Sciences) system equipped with a UV detector ($\lambda = 280 \text{ nm}$) using a Superset 200 10/300 GL gel filtration column at a flow rate of 0.5 ml min^{-1} (running buffer: 50 mM phosphate, 0.15 M NaCl, pH 7.0) at 4°C . The column was previously calibrated using six well-characterized proteins as standards (GE Biosystems Calibration Kit). Oligomerization state was determined from the calculated molecular weight and the predicted monomer molecular weight.

2.4.7 Dynamic light scattering (DLS)

DLS analysis was carried out on a 90Plus particle size instrument (Brookhaven Instruments Ltd.) using a 0.5 cm path length cuvette. Samples were filtered ($0.2 \mu\text{m}$) prior to measurement. Sample concentration was determined by BCA kit (Novagen) and diluted to $200 \mu\text{g ml}^{-1}$ with phosphate buffer. For each protein, (0.6 ml, $\sim 200 \mu\text{g ml}^{-1}$) (with the exception of BFR R61A and BFR R30A (0.6 ml, $\sim 800 \mu\text{g ml}^{-1}$), three

scans were collected at 25 °C and the diameter was averaged from the three replicates. The polydispersity was less than 20% for all samples.

2.4.8 Temperature-dependent circular dichroism (CD)

CD spectra were collected on a Jasco J-710 spectropolarimeter fitted with a Peltier temperature controller in the wavelength range 200-250 nm using a 1 mm quartz cell. The concentration of the protein solutions was determined by BCA (Novagen) and diluted ($100\ \mu\text{g ml}^{-1}$) with phosphate buffer (50 mM NaH_2PO_4 , pH 7.2). The temperature was increased from 20 to 90 °C in 5 °C steps. Samples were allowed to equilibrate (8 min) at each temperature. After melting, the protein solutions were cooled slowly to 20 °C in ten minutes and the spectra were compared to those obtained before unfolding. At least three replicates were performed.

2.4.9 Native PAGE

A 7% gel was run using 20 μg of all the proteins (with the exception of BFR E128A where 10 μg was used) and was stained with Coomassie Blue. Tris-glycine electrophoresis buffer at pH 8.3 was used as gel running buffer.

2.4.10 Chemical denaturation

Samples for chemical denaturation were prepared with $100\ \mu\text{g ml}^{-1}$ protein and GdnHCl or urea (0 - 8 M) varying in phosphate buffer (50 mM NaH_2PO_4 , pH 7.2). The samples were mixed and incubated with no agitation (room temperature, 16 h). The tryptophan emission at 355 nm was recorded using a Varian Cary Eclipse spectrofluorometer with excitation at 295 nm.

2.5 Results and discussion

2.5.1 Selection of key interfacial residues for mutagenesis

Based on the criteria defined in Section 2.4 and with the aid of virtual alanine scanning, nine mutants of BFR were designed (Figure 2.10).

Figure 2.10 Position of mutated residues with respect to symmetry axes of the nano-cage protein. (*Left*) The BFR crystal structure with mutated residues highlighted with respect to the three axes of symmetry and schematized diagrams representing the protein-protein interactions defining these symmetries. (*Right*) The residues of BFR mutated in this study and their schematized position at the symmetry related protein-protein interfaces.

The only potential salt bridge identifiable by inspection between the BFR monomers around the four-fold axis involves E157 and K38 (Figure 2.11A). However, in the crystal structure their side chains are pointing slightly away from each other and virtual alanine scanning^{26, 27} (Table 2.3) suggests that this interaction is subtle at best. As there are no other, even partially promising interactions involving either salt bridges or aromatics at this interface, residue E157 was nevertheless selected for mutagenesis. This residue is part of a C-terminal domain adjacent to the E-helix which is outside of the monomer four-helix bundle, but defines a channel at the 24-mer four-fold axis. In a previous study³⁰ our lab has determined the necessity of the E-helix for oligomerization (see above), however, the importance of this C-terminal domain has not been established.

Along the three-fold symmetry axis, two possible interactions were clearly evident. One is a salt bridge between E128, R61 and R102 (Figure 2.11B); the other involves contacts between Y114 and a hydrophobic pocket in the other subunits. These residues were among those predicted to be energetically important by virtual alanine scanning (Table 2.3). The calculations indicated that R102 is less significant than R61 and E128; therefore R102 is not included among the mutants. It has been shown that in some ferritins the regions located at or near the pores along a three-fold symmetry axis are dynamic and fold independently from the rest of the protein. Moreover, they may be involved in not only iron uptake but also release ^{31,32}. Of these three-fold associated residues, Y114 is in the pore whereas the other two are nearby.

One potential interaction along the two-fold symmetry axis involves residue Y45 and a hydrophobic pocket so it was therefore selected for mutagenesis. The salt bridges on the two-fold symmetry axis are complicated by the fact that four residues appear to be interacting with each other. In the crystal structure, D56 and E60 of one subunit make contact with R30 and K33 of another (Figure 2.11C), and virtual alanine scanning reinforces our prediction of the energetic importance of these interactions (Table 2.3). Although the side chain of D56 is aligned with R30 and E60 with K33, all four are proximal in space and may form an extended hydrogen bonding “hot stretch” ³³ as opposed to a more canonical “hot spot”. Therefore conceptually isolating these into simple binary interactions is not possible. Hence, it was decided to mutate K33, D56, and R30 as single mutations and D56 and E60 as a double mutation. Although this is beyond the scope of this current study, we think clustered residues of this type are especially interesting as they exemplify the complexity of many real biological interactions ³⁴.

The eventual investigation into this complexity could broaden the understanding outside current reductionist views and help explain the non-additivity³⁵ that leads to cooperative folding.

Figure 2.11 The interaction between E157 and K38 at four-fold axis (A), the interactions between E128 and R61 at the three-fold axis (B) and the interactions between four charged residues (K33, R30, E60, and D56) at the two-fold axis (C). The heavy atom distances between the interacting groups are indicated in Å.

Residue	Chain	Int id	$\Delta\Delta G(\text{bind})$	$\Delta G(\text{partner})$	Interface
39	A	1	-0.01	2.66	C4
136	A	1	0.78	-0.06	C4
139	A	0	-0.06	-1.26	C4
140	A	1	1.45	0.88	C4
143	A	1	0.17	-0.42	C4
144	A	1	1.18	1.41	C4
148	A	1	2.34	-0.74	C4
149	A	0	0.06	3.2	C4
151	A	1	0.22	-0.82	C4
157	A	1	0.82	-0.43	C4

1	A	1	0.4	0.46	C3
61	A	1	4.36	5.03	C3
64	A	1	1.84	0.34	C3
102	A	1	1.7	0.87	C3
105	A	1	0.67	2.26	C3
109	A	0	-0.29	-1.14	C3
112	A	1	0.53	0.15	C3
114	A	1	2.43	0.74	C3
115	A	1	1.21	-0.14	C3
117	A	1	0.14	1.6	C3
118	A	1	-0.41	-0.72	C3
121	A	1	1.25	0.45	C3
122	A	0	-0.09	1.09	C3
124	A	0	0.84	2.06	C3
125	A	1	0.26	-0.36	C3
128	A	1	5.71	0.32	C3
132	A	1	-0.1	-1.22	C3
135	A	1	-0.16	-0.18	C3
23	A	1	1.26	-0.48	C2
26	A	1	1.04	0.64	C2
27	A	1	1.81	0.88	C2
30	A	1	9.14	1.41	C2
31	A	0	0.3	1.19	C2
33	A	1	0.8	-0.18	C2
34	A	1	0.32	-0.73	C2
56	A	1	4.54	-1.24	C2
59	A	1	1.2	1.45	C2
60	A	1	2.07	-0.26	C2
63	A	1	1.59	1.25	C2
68	A	0	0.2	0.54	C2
70	A	0	0	2.82	C2
71	A	1	1.52	1.72	C2
72	A	1	2.94	-0.66	C2
74	A	1	1.11	1.91	C2
76	A	1	0.04	-0.28	C2
77	A	1	1.24	2.18	C2
79	A	0	0.19	1.44	C2

Table 2.3 Results from “virtual alanine scanning” of the BFR different protein-protein interfaces (C2, C3 and C4). Black shaded indicates residues involved in this study. The website that was used was: <http://robetta.bakerlab.org/alaninescan>.

The abbreviations used in Table 2.3 are as follows: “residue” is the residue number that was mutated to alanine using the numbering system in the PDB coordinates; “chain” is the pdb chain identifier; “int_id” describes the nature of the interfacial residue. A value of 1 indicates that, in the complex, at least one atom from the residue is within 4 Å of an atom in a residue from the other protein. A value of 0 indicates that the number of beta carbon atoms surrounding the beta carbon atom of the residue in question increases and therefore this residue is one that is buried upon formation of the protein-protein interaction although it may not be making a direct

contact; “ $\Delta\Delta G(\text{bind})$ ” is the change in the binding energy between the partners due to the mutation of the residue in question to alanine. A positive value indicates the mutation has a destabilizing affect whereas a residue causing a negative value upon mutation is stabilizing; “ $\Delta G(\text{partner})$ ” is an indication of how a mutation affects the stability of the interacting partner.

2.5.2 The protein-protein interactions at the symmetry related interfaces of BFR

The stability of a macromolecular complex in a biological system is by the overall free energy of formation including its associated solvation energy. These interactions can be driven by hydrogen bonds, salt bridges and the burial of hydrophobic surface area.²⁸

2.5.2.1 Comparison of the buried surface area and exposed surface area between different bacterioferritins

The buried surface area at three different symmetry-related interfaces (C2, C3, and C4) compared to total buried and exposed surface areas in the homo-oligomeric nanocage assemblies (A24) of bacterioferritin family were calculated using PISA (Protein Interfaces, Surfaces and Assemblies) software²⁸. The protein-protein interaction around the four-fold symmetry axis is made up of side by side (AL or AX in Figure 2.12) and head to head (AM in Figure 2.12) interfaces which are analyzed separately. The analysis defines each monomer as having one C2-related surface, two C3-related surfaces, two side-by-side C4-related surfaces and one head to tail C4-related surface. The analysis of the protein-protein interfaces of the bacterioferritins with available crystal structures indicate that the area buried upon cage formation for *E. coli* BFR is slightly smaller than the average. The total buried surface area of *E. coli* BFR (72190 Å²) is slightly smaller than the average (73004.0 Å²) due to its smaller than average

two-fold and three-fold interfaces even though the four-fold interface is larger than average.³⁶

Interface type	C2	C3	C4 [†]		Total in A ₂₄	Total for A ₂₄
			Side- by- side	Head-to- head		
Proteins (PDB ID)	Buried surface area, sq. Å [‡]					Exposed Surface area, sq. Å [‡]
<i>EcBfr</i> (2vxi,1bfr,1bcf,2htn,3e1n)	887.9	510.4	535.0	29.4	72190	137144
<i>PaBfrB</i> (3isf,3ise,3is8,3is7)	820.9	448.5	460.4	19.4	63553	140838
<i>AvBfr</i> (2fl0,2fkz,1sof)	826.5	468.4	415.5	17.8	62690	140130
<i>BaBfr</i> (3fvb)	959.3	464.6	622.8	31.9	75980	138380
<i>RcBfr</i> (1jgc)	793.5	385.6	694.3	14.1	71170	136690
<i>RsBfr</i> (3gvy)	954.0	514.6	561.9	18.2	75100	134160
<i>MtBfrA</i> (2wtl)	1017.9	387.2	749.1	88.2	80980	129670
<i>MsBfr</i> (3bkn)	904.9	513.7	668.7	75.1	80510	126360
<i>DdBfr</i> (1nfv,1nf6,1nf4)	1564.4	432.6	338.9	12.6	74863	150120
Average	969.9	458.4	560.7	34.1	73004.0	137055

Table 2.4 Buried surface area (Å²) at the three symmetry-related interfaces (C2, C3, and C4) compared to total buried and exposed surface areas in the homo-oligomeric octahedral nanocage assemblies (A₂₄) of bacterioferritin family. The table is taken from *Ardejani et al*³⁶.

2.5.2.2 The individual residues involved in protein-protein interactions at symmetry related interfaces in *E. coli* BFR

The protein-protein interactions were further analyzed using PISA software to understand the assembly of *E. coli* BFR. The number of residues, solvent accessible area and solvation energy at these six inter-subunit interfaces were calculated to analyze their physical properties (Figure 2.12, Table 2.5). Furthermore, the types of interactions (hydrogen bond, salt bridges, disulfide bond etc.), accessible surface area, buried surface area and solvation energy effect (ΔG^i) for each residue that residing at the six different interfaces are quantified and listed too (Tables 2.6).

Figure 2.12 Each monomer (for example monomer A) in BFR is involved in six different protein-protein interactions at the respective symmetry related interfaces (C2, C3, C4 which are highlighted with red circles). The other monomers (B, H, W, X, M and L) that monomer A interacts with are all indicated. The residues in monomer A that are involved in the protein-protein interactions are shown in red spheres. The figure is generated using UCSF Chimera (PDB ID: 1bfr).

Types of protein-protein interactions	AB (C2)	AH (C3)	AW (C3)	AX (C4)	AM (C4)	AL (C4)
Number of residues						
Interface	24 (15.2%)	13 (8.2%)	16 (10.1%)	18 (11.4%)	2 (1.3%)	12 (7.6%)
Total Interface	85 (53.8%)					
Total Surface	144 (91.1%)					
Total	158 (100.0%)					
Solvent inaccessible area, Å²						
Interface (buried surface area)	908.5 (10.6%)	511.0 (5.9%)	460.8 (5.4%)	533.1 (6.2%)	21.6 (0.3%)	552.8 (6.4%)
Total surface area	8600.3 (100.0%)					
Solvation energy, kcal/mol						
Isolated monomer	-131.2 (100.0%)					
Gain at complexation (ΔG^S)	-2.8 (2.1%)	-1.4 (1.1%)	-1.7 (1.3%)	-1.7 (1.3%)	-0.1 (0.1%)	-1.6 (1.2%)
Interface CSS	1.00	1.00	1.00	0.14	0.00	0.14

Table 2.5 Calculated properties of the six monomer-monomer interactions defined in Figure 2.12. “Solvation energy of isolated monomer” is the solvation energy of the isolated chain folding ΔG^f . “Solvation energy gain at complexation” is the change of the solvation energy of the structure due to the interface formation. “CSS” stands for the complexation significance score. CSS ranges from 0 to 1 as interface relevance to complexation increases. Note: some of the protein-protein interfaces share residues; the calculated properties displayed are all about monomer A. The data is generated with the PDB entry 1bfr using the PISA online service: http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html

The analysis demonstrates that 8.9% (11 residues) of the 158 total residues in the monomer are buried inside the core of the folded four-helix bundle; the remaining 91% (144 residues) are therefore considered “surface residues” (Table 2.5). Of these, about 50% are located at protein-protein interfaces which can be categorized into six types as described in Figure 2.12 (Note: some of the protein-protein interfaces share residues and therefore adding them up will result in a number larger than the total that are buried from solvent; for example Asn148 and Gln151 are involved both in AM and AL, Table 2.6). The total solvent accessible area of monomer A is (8600.3 Å²), where roughly C2, C3 and C4 interfaces each contribute one third of the total.

For solvation to occur, energy is required to release individual ions from the crystal lattices. The energy comes from the energy released when ions of the lattice associate with molecules of the solvent. In other words, solvation energy is the change in Gibbs energy when an ion or molecule is transferred from the gas phase to a solvent. It can be calculated in the simplest ASA (Accessible Surface Area)-based method:

$$\Delta G_{solv} = \sum_i \sigma_i ASA_i$$

where ASA_i is accessible surface area of atom i , and σ_i is solvation parameter of atom i , i.e. a contribution to the free energy of solvation of the particular atom i per surface unit area. Upon complex formation, residues at the interface are buried and the accessible surface area of the complex is reduced compared to the sum of ASA of the dissociated structures. The change of the solvation energy of the structure due to the interface formation is the free energy gain upon the complexation. The calculation showed that the gain at two-fold interface is the highest (2.1%) whereas the donation from AH, AW, AX and AL type are close. In addition, the contribution of the AM

type of inter-subunit interaction at four-fold interface to the solvent accessible area (0.3%) and solvation energy gain is negligible (Table 2.5).

“CSS” stands for the complexation significance score. The higher CSS value indicates the more important the interface in the complex formation. The calculation indicates that two-fold (AB) and three-fold (AH and AW) interfaces play an essential role in complex formation (CSS=1.00). Among the three types of inter-subunit interactions in four-fold interface, types of AX and AL play an auxiliary role in complexation (CSS=0.14), whereas interface type AM does not play any role in complexation and seems to be a result of crystal packing only (CSS=0.00).

Residues involved in the symmetry related inter-subunit interactions are listed in Tables 2.6. In addition to the salt bridge and hydrogen bonding interactions between the residues, the buried and accessible surface areas and solvation energies of the interfacing residues were analyzed. The solvation energy gain of the interface is calculated as difference in solvation energies of all residues between dissociated and associated (interfacing) structures. Hence, the positive solvation energy ΔG^i of the residue makes a negative contribution to the solvation energy gain of the interface, which corresponds to hydrophobic effect. Except for the aromatic residue Tyr114, residues that subjected to mutation in this study are all involved in the hydrogen bonding or salt bridging interactions with negative solvation gain upon the respective interface formation. It is showed that solvent-accessible surface area of Tyr114 is almost completely buried upon the interface formation with a high positive solvation energy effect (1.05 kcal/mol), which is attributed to hydrophobic effect. Residue E128 which makes contacts with R61 and R102 through hydrogen bonding and salt bridging interactions is buried upon the three-fold interface formation.

Residue	Interactions	ASA, Å ²	BSA, Å ²	ΔG^i , Solvation energy effect, kcal/mol	Types of Inter-subunit interaction
---------	--------------	---------------------	---------------------	--	------------------------------------

Leu19		28.80	12.55	IIII	0.20	AB (C2)
Asn23	H	63.69	39.95	IIIIII	-0.15	AB (C2)
Phe26		70.68	35.87	IIII	0.57	AB (C2)
Leu27		77.80	76.12	IIIIIIII	1.19	AB (C2)
Arg30	H & S	86.15	63.49	IIIIII	-0.22	AB (C2)
Met31		44.98	24.70	IIII	0.51	AB (C2)
Lys33	S	104.14	22.91	III	0.07	AB (C2)
Asn34	H	126.09	47.82	III	-0.31	AB (C2)
Met52		66.10	16.21	III	0.66	AB (C2)
Asp56	H & S	72.65	44.58	IIIIII	-0.13	AB (C2)
Ile59		41.97	31.78	IIIIII	0.51	AB (C2)
Glu60	H & S	107.09	46.33	IIII	-0.08	AB (C2)
Leu63	H	93.10	76.54	IIIIIIII	1.00	AB (C2)
Phe64		101.23	11.86	II	0.16	AB (C2)
Leu68		149.40	25.03	II	0.40	AB (C2)
Pro69		26.63	19.26	IIIIII	0.10	AB (C2)
Leu71	H	101.04	94.09	IIIIII	0.82	AB (C2)
Gln72	H	151.75	92.61	IIIIII	-0.90	AB (C2)
Asp73		93.61	0.48	I	0.01	AB (C2)
Leu74		92.75	60.18	IIIIII	0.84	AB (C2)
Gly75	H	42.26	6.42	II	-0.05	AB (C2)
Lys76		168.71	11.23	I	0.18	AB (C2)
Leu77	H	53.69	38.47	IIIIII	0.47	AB (C2)
Ile79		46.53	10.04	III	0.16	AB (C2)
Met1	H & S	123.95	96.51	IIIIII	2.07	AH (C3)
Arg61	H & S	32.91	9.61	III	-0.24	AH (C3)
Phe64		101.23	70.99	IIIIII	1.14	AH (C3)
Glu66		92.62	4.15	I	-0.05	AH (C3)
Asp109		52.31	3.19	I	-0.04	AH (C3)
Val111		50.17	0.31	I	0.00	AH (C3)
His112	H	127.13	54.06	IIII	0.08	AH (C3)
Asp113		23.25	9.42	IIII	0.02	AH (C3)
Tyr114		101.54	92.19	IIIIIIII	1.05	AH (C3)
Val115		99.67	84.51	IIIIIIII	1.32	AH (C3)
Arg117		60.80	34.58	IIII	-0.60	AH (C3)
Asp118	H & S	70.00	42.42	IIIIII	-0.16	AH (C3)
Glu122		71.89	9.09	II	-0.15	AH (C3)
Leu95		86.37	13.73	II	0.22	AW (C3)
Ala98		2.76	0.33	II	0.01	AW (C3)
Arg102	H	127.56	77.80	IIIIII	-0.48	AW (C3)
Glu103		93.66	1.68	I	0.03	AW (C3)
Ile105		19.93	19.59	IIIIIIII	0.31	AW (C3)
Gly106		42.22	20.88	IIII	0.31	AW (C3)
Asp109		52.31	1.15	I	-0.01	AW (C3)
Arg117		60.80	26.22	IIII	-0.46	AW (C3)
Asp118		70.00	2.24	I	0.03	AW (C3)
Ile121		59.30	57.45	IIIIIIII	0.92	AW (C3)
Leu124		18.41	18.41	IIIIIIII	0.29	AW (C3)
Arg125	H & S	156.98	59.76	III	-0.26	AW (C3)
Glu128	H & S	79.72	74.87	IIIIIIII	-0.34	AW (C3)
Ile131		18.08	18.08	IIIIIIII	0.29	AW (C3)
Asp132		88.50	26.78	III	0.15	AW (C3)
Glu135	H & S	80.92	41.83	IIII	-0.45	AW (C3)
Arg39		46.41	19.12	IIII	-0.18	AX (C4)
Asp132		88.50	12.30	II	-0.14	AX (C4)
Thr136	H	64.59	42.15	IIIIII	0.39	AX (C4)
Asp139		71.85	2.03	I	0.00	AX (C4)
Leu140		57.49	50.91	IIIIII	0.81	AX (C4)
Gln142		106.49	2.09	I	-0.02	AX (C4)
Lys143	H	154.27	82.39	IIII	-1.16	AX (C4)
Met144		79.39	78.17	IIIIIIII	1.26	AX (C4)
Gly145		30.25	10.93	III	0.16	AX (C4)

Asn148	H	86.50	59.03	IIIIII	-0.53	AX (C4)
Tyr149		3.33	1.41	IIII	0.02	AX (C4)
Gln151	H	123.23	16.91	II	-0.27	AX (C4)
Ala152	H	56.01	43.84	IIIIIIII	0.41	AX (C4)
Gln153		3.35	2.47	IIIIIIII	-0.03	AX (C4)
Arg155	H	191.08	12.09	I	0.10	AX (C4)
Glu156		100.37	5.47	I	-0.02	AX (C4)
Glu157	H & S	135.58	81.96	IIIIII	-0.15	AX (C4)
Gly158		122.01	9.83	I	-0.11	AX (C4)
Asn148		86.50	7.30	I	-0.09	AM (C4)
Gln151		123.23	14.35	II	-0.20	AM (C4)
Asn34	H	126.09	24.84	II	-0.33	AL (C4)
Trp35		106.76	47.46	IIII	0.15	AL (C4)
Gly36		36.35	22.80	IIIIIIII	0.33	AL (C4)
Leu37		13.89	12.89	IIIIIIIIII	0.21	AL (C4)
Lys38	H & S	93.39	21.78	III	-0.73	AL (C4)
Gln147	H	160.41	67.65	IIII	0.03	AL (C4)
Asn148	H	86.50	14.48	II	-0.25	AL (C4)
Leu150		57.02	48.65	IIIIIIIIII	0.78	AL (C4)
Gln151	H	123.23	89.57	IIIIIIII	-0.41	AL (C4)
Ile154		79.79	61.71	IIIIIIII	0.99	AL (C4)
Arg155	H & S	191.08	124.27	IIIIII	-1.89	AL (C4)
Glu156		100.37	16.67	II	-0.16	AL (C4)

Table 2.6 The residues in monomer A which are involved in the inter-subunit interactions with other monomers (B, H, W, X, M, L). The types of symmetry related interfaces that monomer A involves are also indicated (C2, C3 and C4). Residues mutated to alanine in this study are shaded black. “H” and “S” indicate hydrogen bond and salt bridge respectively. “ASA” is the solvent-accessible surface area of the corresponding residue. “BSA” indicates the solvent-accessible surface area of the corresponding residue that is buried upon interface formation. The vertical bars give a mnemonic representation of the buried area fraction: one bar corresponds to 10% of the total solvent-accessible surface area buried. ΔG^i indicates the solvation energy of the corresponding residue.

2.5.3 Verification of gene synthesis products

Short oligonucleotides were designed based on the amino acid sequences of the native ferritins with codons changed in preference for those that have high usage in *E. coli*. The full-length sequence is generated progressively in a single PCR reaction by overlap extension by means of a two-step PCR procedure ²⁹, followed by PCR amplification with two outermost primers (Figure 2.13A). The assembly PCR product showed smeared bands at around 100 ~ 400 base pairs which is common in this technique ³⁰. A clean band at ~ 500 bp was achieved by amplification using a PCR with the two end primers.

This PCR product was then used as template for extension with designed primers derived from the pET 32 EK/LIC expression vector. The extension PCR product was

analyzed by agarose gel electrophoresis and the band at ~ 500 bp was excised and purified by gel purification (Figure 2.13B).

T4 DNA polymerase has 3'→5' exonuclease activity. In our strategy, the insert target genes are treated with T4 DNA polymerase and dATP to generate overhangs which will be complementary to the pET-32 vector that is also treated with T4 DNA polymerase and dTTP. These two complementary overhangs are long enough (14bp) to anneal by themselves to form a plasmid. The plasmid which confers a resistance to carbenicillin was then transformed by electroporation into Novablue (Novagen) cells, and the resulting resistant colonies were screened by colony PCR which was then analyzed by agarose gel electrophoresis (Figure 2.13C). The bands corresponding to the correct molecular weight (~ 500 bp) was observed.

Figure 2.13 DNA electrophoretic analysis of typical gene synthesis (BFR K33A is presented as an example): gene assembly and amplification (A), Amplification with LIC primers (B) and colony PCR (C).

After colony PCR, the plasmids harboring the respective gene inserts were obtained by miniprep and sent for sequencing. The resulting sequences were aligned with the designed gene sequences using the Blast web tool that is available online:

<http://blast.ncbi.nlm.nih.gov/>. The alignment results are listed in Appendix 1 and 2.

All the mutations were successfully introduced into the expression vector.

Mutants BFR R30A and R61A were constructed using QuikChange site-directed mutagenesis kit (Stratagene) with the BFR vector as the template using the primers in Appendix 1.

2.5.4 Protein over-expression and purification

The pET-32 Ek/LIC expression vector (Novagen) provides proteins fused to a series of affinity tags to simplify purification and these tags can be removed through cleavage at an enterokinase protease site which abuts the natural start site of the protein, leaving no additional residues-an essential requirement for self-assembly studies. The protein was purified through immobilization on affinity resin followed by on-resin enterokinase cleavage and protein release. The advantage of this strategy is that because the monomer is bound to the resin it only self-assembles upon release (the large affinity tags presumably prevent self-assembly), thus it would conceivably be relatively simple to load for drug delivery studies. This purification method proved general and robust (Figures 2.14 and 2.15).

A series of SDS-PAGE gels were carried out to verify each step of the expression and purification procedure. The gels demonstrate that (Figure 2.14) wild type BFR and the BFR derived mutants were all expressed as soluble proteins. After they were bound to Ni-NTA affinity resin, enterokinase was added to cleave off the His-tag to release the proteins. Pure protein bands of ~19 kD were observed by SDS-PAGE and molecular weights were determined by MALDI-TOF mass spectrometry (Figure 2.15). These results verified that the target proteins were expressed and purified successfully.

Figure 2.14 Electrophoretic monitoring of the purification of the BFR-derived proteins. Lane 1: ladder. Lane 2: before induction. Lane 3: post induction. Lane 4: soluble fraction. Lane 5: insoluble fraction. Lane 6: flow through. Lane 7: wash. Lane 8: cleavage. The schematics follow the convention defined in Figure 2.7.

Protein	MW (calculated)	MW (measured)
BFR	18495.0	18501.8
BFR E157A	18437.0	18442.7
BFR E128A	18437.0	18427.7
BFR Y114A	18402.9	18400.7
BFR R61A	18409.9	18417.1
BFR Y45A	18402.9	18397.8
BFR K33A	18437.9	18445.3
BFR D56A	18451.0	18452.8
BFR D56A E60A	18392.9	18389.1
BFR R30A	18409.9	18412.0

Figure 2.15 MALDI-TOF analysis of expressed proteins.

2.5.5 Transmission electron microscopy

The proteins were analyzed to determine whether each possessed the ability to assemble into a nano-cage. Samples of the proteins were prepared and treated with uranyl acetate negative stain. Electron micrographs indicated that all proteins form nano-cages under these conditions (Figures 2.16). Image analysis of these micrographs indicates that the native proteins form structures of sizes on par with those described in the literature ¹². Interestingly, the mutants assemble into cages having the same or slightly smaller size than the native protein; none are larger (Figure 2.17).

Figure 2.16 All BFR mutants can form nano-cages under TEM conditions. Negatively stained TEM micrographs indicate that all of the BFR-derived proteins can form nano-cages. Scale bars indicate 50 nm. Schematics follow the convention in Figure 2.10.

Figure 2.17 Quantification of nano-cage sizes measured by TEM. Schematics follow the convention defined in Figure 2.10.

2.5.6 Size exclusion chromatographic analysis of assembly in solution

The proteins were analyzed by size exclusion chromatography, which separates different structures by size, to determine the effect the mutations have on the ability of the proteins to assemble in solution (Figure 2.19). The retention volumes of each species were correlated to the molecular weights of protein standards, which were then used to calculate approximate oligomerization states.

The wild-type protein BFR, consistent with the literature³⁰, forms a mixture of 24-mer and dimer. The mutants were also found to form only these states. The mutant BFR E157A exists in a similar ratio to that of BFR while the other mutants form altered ratios of dimer and 24-mer, all of which favor the dimer. Strikingly, BFR R30A, R61A, Y114A, and E128A only assemble into dimer. No 24-mer is observable for all of these four mutants. This is somewhat surprising because these mutants form hollow nano-structures under TEM conditions (*see above*). As SEC experiments were conducted in neutral condition and uranyl acetate arguably is acidic, 1% PTA, which

is neutral, was used as a negative stain in additional TEM experiments (Figure 2.18). These results agree with those using uranyl acetate suggesting that the difference in observed nano-cage formation in SEC and TEM is unrelated to pH. In addition, the 24mer:dimer ratio was further probed with SEC and no concentration dependence was determined although it should be noted that the concentration range is limited due to UV detector limits. We are exploring further the fundamentals of this observation with other techniques; however, to resolve these inconsistencies and to confirm these solution data, two other solution techniques, dynamic light scattering and native PAGE were employed (*see below*).

Figure 2.18. BFR E128A can form nano-cages under TEM conditions using neutral negative stain 1% PTA. Scale bar indicates 50 nm.

Figure 2.19. All BFR mutants form 24-mer and/or dimer in solution and four residues control the oligomerization state upon single point mutation. *(Top Left)* SEC chromatograms of all the proteins showing mixtures of two states which are identified as 24-mer and dimer. *(Bottom Left)* Dynamic light scattering *(Top Right)* and native gel electrophoresis *(Bottom Right)* of the four mutants that exhibit no detectable 24-mer agree with SEC solution results confirming that these amino acids act as oligomerization “switch residues”. Schematics follow the convention in Figure 2.10.

2.5.7 Dynamic light scattering (DLS) analysis of BFR R30A, R61A, Y114A, and E128A assembly in solution

To resolve the inconsistencies between the SEC and TEM results, DLS was used to determine whether the four mutants, BFR R30A, R61A, Y114A, and E128A, form

24-mer nano-structures in solution (Figure 2.19). Comparison of the DLS data for the four mutants and the native protein, BFR, reinforces the SEC experiments. All four mutants form particles in solution that are much smaller than BFR.

2.5.8 Native gel electrophoretic (native PAGE) analysis of BFR R30A, R61A, Y114A, and E128A assembly in solution

To further resolve the inconsistencies between the solution (SEC and DLS) and the TEM data, native PAGE was performed (Figure 2.19). Consistent with the SEC chromatogram, the native protein separates into two bands corresponding to 24-mer and dimer oligomerization states. The lanes corresponding to the four mutants display only single bands that have the same mobility as the dimer state in BFR. With the strength of three solution techniques all supporting each other, we conclude that in solution these mutants only form dimer with no detectable 24-mer and that the TEM results do not correspond to what is happening in solution.

It is probable that the TEM results are an artifact of the drying step during sample preparation. It is possible that as drying occurs, the concentration of protein reaches a threshold level, driving the subunits together into a cage. Although the TEM does not correlate with the solution data, the micrographs suggest that while these mutants do not form nano-cages in solution, mutation does not completely destroy the potential of the mutants to assemble under “forcing conditions”-it is simply disfavored in solution. In addition, these results emphasize the importance of confirming electron microscopy-based conclusions with data derived from solution experiments.

2.5.9 The solution secondary structure as monitored by circular dichroism

The wild-type protein and all of the mutants fold into helical secondary structure as evidenced by the presence of CD minima at 208 and 222 nm (Figure 2.20). Nearly all of the mutants exhibit reduced secondary structure as compared to the parent. Not

unexpectedly, the lone exception is BFR R30A which is the most thermally stable (*see below*). The double mutant BFR D56A E60A, which is predictably less stable than the two proteins that eliminate single interactions within the cluster (BFR K33A and BFR D56A), also exhibits the largest reduction in helicity.

Increasing the temperature results in the melting of the secondary structure (Figures 2.20, 2.21 and 2.22). By monitoring the 222 nm band, it is evident that, for all the proteins, this unfolding is cooperative, suggesting that they all adopt tight, stable folds. Most of the mutants melt at a lower temperature than native BFR, suggesting that even simple point mutations can affect the protein thermal stability. The exceptions to this trend are BFR E157A, which melts at nearly the same temperature as BFR, and BFR R61A and R30A, both of which exhibit increased thermal stability.

Figure 2.20 Full CD scans for melts used in Figure 2.21. Schematics follow the conventions defined in Figure 2.10. Data is the average of at least three replicates.

Slowly reducing the temperature back to starting conditions after the melting

experiments allows for the exploration of folding reversibility (Figure 2.19). The most thermally stable protein, BFR R30A, exhibited the least folding reversibility.

Figure 2.21 The role mutation plays in BFR-derived protein thermal stability and ability to refold after denaturation. (**Top**) CD spectra of the BFR derivatives before denaturation (solid lines) and after slow cooling post-thermal denaturation (dashed lines). (**Bottom**) Thermal transitions of the BFR derivatives as monitored by CD at 222 nm. Data is the average of at least three replicates. Schematics follow the convention in Figure 2.10 and the coloring follows that of Figure 2.19.

Figure 2.22 Quantified melting temperatures from data presented in Figures 2.20 and 2.21. These values are the average of at least three replicates and the error bars represent standard deviation. Schematics follow the convention defined in Figure 2.10.

As it was surprising that the removal of salt bridges would result in an increase in protein stability for R61A and R30A, these latter results were confirmed by denaturation with both guanidine and urea (Figure 2.23) and monitoring endogenous tryptophan fluorescence. In both cases, the concentration of denaturant required to unfold the protein is greater than that for the parent protein. These data confirm that a single point mutation to alanine is possible to generate a mutant that is more stable than wild-type BFR.

Figure 2.23 Chemical denaturation of BFR (black) and BFR R30A (red) with urea (diamonds) or GdnHCl (circles) following intrinsic tryptophan fluorescence.

Comparison of the role mutation plays on the thermal stability of the helical structure of these proteins results in a wide range of melting temperatures (Figures 2.20 and 2.21). The only residue about the four-fold axis that was mutated was E157. This residue, which is located on a C-terminal strand adjacent to the structurally significant E helix, appears to be making an hydrogen bond with K38, albeit poorly aligned (Figure 2.11A). However “virtual alanine scanning” predicts that this interaction makes a negligible contribution to the overall stability (*see above*). The data agree with the computation in that the E157A mutant, which destroys the possibility of forming this hydrogen bond, has a thermal denaturation transition at nearly the same temperature as the native protein. Moreover, this mutation has no effect on the oligomerization state. Interestingly, our previous studies showed that deletion of the entire E helix resulted in thermal destabilization by more than 10 °C and complete disruption of the 24-mer in favor of the dimer³⁰. Taken together these data suggest that E157, and possibly the entire C-terminal strand, is not contributing to the structural energetic importance of the E-helix³⁷.

The point mutants, E128A and Y114A, both located along the three-fold axis and the double mutant D56A E60A at the two-fold axis have melting temperatures depressed by more than 15 °C from the native protein. These data suggest that thermodynamic “hot spots” do indeed exist in this system.

In contrast are the thermal stabilities of the three-fold axis-associated mutant, R61A and R30A, which is located at the two-fold axis. These single point mutants exhibit melting temperatures 4 °C and 8 °C higher than the native protein, and the increased stability over that BFR was confirmed using the chemical denaturants, urea and guanidine (Figure 2.23). The R61A result is even more surprising when compared to the destabilization observed upon mutation of E128 with which it interacts. An explanation of these somewhat surprising observations could be reflected in the complexity of interactions that exist in natural systems. The basic amino acids, arginine and lysine, due to the long aliphatic linker between the alpha carbon and charged head group often play steric roles in concert with that as a component in a salt bridge^{38,39}. Therefore, it may not be a coincidence that these two residues are arginine and it could be possible that removal of these large side chains are permitting rearrangement of the interface to maximize stability. Whatever their origin, these results suggest that BFR has not evolved solely to maximize structural stability. This is reasonable considering that BFR exists naturally as a mixture of oligomerization states^{30,40} and has a complex cellular function that is intrinsically linked to its structure.

2.6 Conclusion

Residue	Interface	$\Delta\Delta G^*$ (kcal/mole)	T _m (°C)	ΔT_m (°C)	DLS Hydrodynamic diameter (nm)	Oligomerization state from SEC
Wild type	-	-	64.8±1.1	-	13.3±0.2	24-mer (45%) + Dimer (55%)
E157A	C4	0.82	64.9±0.7	-0.1±1.8	-	24-mer (41%) + Dimer (59%)

R61A	C3	4.36	68.8±0.8	-4.0±1.9	6.1±0.3	Dimer (100%)
Y114A	C3	2.43	48.4±0.8	16.4±1.9	6.8±0.3	Dimer (100%)
E128A	C3	5.71	46.9±0.2	17.9±1.3	6.7±0.6	Dimer (100%)
R30A	C2	9.22	72.2±0.6	-7.4±1.7	6.0±0.2	Dimer (100%)
K33A	C2	0.8	56.9±1.5	7.9±2.6	-	24-mer (4%) + Dimer (96%)
Y45A	C2	1.48	58.6±0.4	6.2±1.5	-	24-mer (22%) + Dimer (78%)
D56A	C2	4.54	51.8±0.3	13.0±1.4	-	24-mer (16%) + Dimer (84%)
E60A	C2	2.07	-	-	-	-
D56A/ E60A	C2	-	48.6±0.6	16.2±1.7	-	24-mer (24%) + Dimer (76%)

Table 2.7 The properties of BFR proteins determined by virtual alanine shaving computational calculation, CD, DLS and SEC analysis. $\Delta\Delta G^*$ is obtained from the virtual alanine scanning studies.

Many ferritin nano-cage proteins are formed through the homo-oligomerization of relatively small and simple monomers that fold into helix bundles, a motif that has been extensively studied. Therefore these proteins may develop into a next generation of model systems for the protein engineering and design communities, which are turning their attention from tertiary structure to more complex quaternary structure.

The octahedrally symmetric maxi-ferritin, *E. coli* bacterioferritin, BFR, was used in this study to help establish an understanding of the role that protein-protein interactions play in the self-assembly of nanostructure. By conceptually shaving to alanine the side chains of nine amino acids, identified by inspection of the crystal structures in concert with “virtual alanine scanning”, the role of specific interactions at symmetry-related protein-protein interfaces was probed. While the residues that were chosen for mutagenesis are only a fraction of those located at the various protein-protein interfaces, we predicted that they would be among the most significant.

These interfacial amino acids were organized based on their relationship with different symmetry elements within the octahedral BFR nano-cage with the hope that they could help shed light on the importance, if any, these symmetries have on the stability or assembly of the protein. The selected nine amino acids residues were either aromatic or involved in salt bridging interactions. One of the mutants disrupted interactions at the four-fold axis, three mutants were at the three-fold axis, and five mutants were focused on residues along the two-fold axis (Figure 2.24).

After cloning, expression, and purification, we determined that all nine of the mutants formed cooperatively folded α -helical secondary structures and had the ability to form nano-cages as evidenced by temperature dependent CD and TEM. SEC, DLS and native PAGE were also performed to characterize these mutants. This evidence has proved that many of the selected amino acids act as hot spot residues (Figure 2.22 and 2.23).

Figure 2.24 The role protein-protein interface residues play in the thermodynamic stability of the BFR nano-cage protein. The change in thermal stability upon mutation to alanine is mapped onto the BFR crystal structure. Schematics follow the convention in Figure 2.10. The legend in the lower left reflects the degree to which each residue influences the stability upon mutation.

Figure 2.25 The change in thermal stability mapped onto the BFR crystal structure for interfacial residues at the two-fold (A), three-fold (B), and four-fold axes (C). Each panel displays schematics indicating the direction each subunit has been pulled apart and rotated to facilitate viewing, and the approximate position and name of the residues that were mutated following the convention in Figure 2.10. The legend in the lower right reflects the degree to which each residue influences the stability.

Electron microscopy indicates that mutagenesis does not disrupt the ability of the proteins to form nano-cage structures. In solution, however, all the mutants, with the exception of E157A, demonstrate a shift in the ratio of oligomerization to favor the dimer over the 24-mer suggesting destabilization, to some degree of the cage form. Four residues, E128, Y114, R61, and R30, when mutated to alanine, result in proteins that form no detectable 24-mer in solution, suggesting that they are “switch residues” that control the formation of nano-assembly while having little affect on the lower order structure. While these data present a caveat concerning oligomerization state, conclusions based on TEM of semi-dried samples, when taken together suggest that mutagenesis may not completely destroy the ability to form specific assembled structure under extreme conditions even if it completely shifts the oligomerization state in solution. The fact that R30A is among these four mutants is especially intriguing as the mutated residue is located at the two-fold axis. Conventional

wisdom would dictate that alanine shaving at the dimer interface would result in destruction of interactions that favor dimerization. A possible explanation is that the conformational rearrangement upon removal of the large arginine side chain is transferred to other surfaces of the protein thereby resulting in disruption of other symmetry related interfaces. This remodeling could result in an unpacking of the protein in general which may explain the decreased cooperativity in the thermal denaturation profile (Figure 2.20). We are currently exploring this hypothesis by determine the change in solvent exposed hydrophobic surface area upon mutation. Another explanation could be the “lock and keystone” hypothesis³⁶ where mutation alters the overall geometry of dimer quaternary structure affecting the presentation of other surfaces and thereby influencing the ability to oligomerize further beyond the dimer. Another possibility is that this dimer is unique and unrelated to the dimer centered around the two-fold axis in the 24-mer. A dimer is a proposed intermediate in the mechanism of assembly of ferritins^{41,42} and the potential for the existence of alternative dimers warrants further investigation. Together these four mutants further emphasize the strength of the model system in its sensitivity to even subtle mutagenesis. In addition, these results are especially intriguing in that they demonstrate that self-assembling nano-structure can be controlled by single point mutation.

The residues in this study were selected by simple inspection and thus an arguably arbitrary selection strategy (although it has worked with even greater success in the related protein DPS. See Chapter 4 in this thesis). Confidence in our approach was enhanced by the fact that many of these residues agreed with “virtual alanine scanning” calculations. It is a testament to the tractability of this system that of the nine mutants predicted through such a low level algorithm, only one, E157A

had no effect on either the stability or the oligomerization state. That the residues selected for mutation clearly are energetically and structurally significant is evidenced by the fact that three mutants, E128A, Y114A and D56A E60A have depressed melting temperatures by at least 15 °C, and two mutants have melting temperatures elevated by at least four degrees. Moreover four of the mutants, E128A, Y114A, R61A, R30A, have completely lost the ability to form the nano-cage structure and only form cooperatively folded dimer in solution. Therefore these are oligomerization “switch residues”. It is remarkable that multiple properties of the system can be tuned so easily. We are currently exploring the possibility that other computationally predicted residues are significant either structurally or energetically. Moreover, we are actively working to disentangle the relationship between thermal stability and nano-structure, which this current study indicates are linked in perhaps quite complicated ways. In addition we are applying similar strategies to other nanocages with similar and yet structurally distinct oligomerization states and symmetries so that the knowledge gleaned can be used to rationally design oligomers with unique structures which in turn can be used for the templation of unique inorganic nanomaterials ^{5,16} and to aid the fundamental analysis of the evolutionary relationships of key interfacial interactions that control assembly in this class of proteins.

2.7 References

1. Liu XF, Theil EC. Ferritins: Dynamic management of biological iron and oxygen chemistry. *Accounts Chem Res* 2005;38(3):167-175.
2. Wang Q, Lin T, Tang L, Johnson JE, Finn MG. Icosahedral virus particles as addressable nanoscale building blocks. *Angew Chem Int Ed Engl* 2002;41(3):459-462.
3. Destito G, Yeh R, Rae CS, Finn MG, Manchester M. Folic acid-mediated targeting of cowpea mosaic virus particles to tumor cells. *Chemistry & Biology* 2007;14(10):1152-1162.
4. Ren Y, Wong SM, Lim LY. Folic acid-conjugated protein cages of a plant virus: A novel delivery platform for doxorubicin. *Bioconjugate Chem* 2007;18(3):836-843.

5. Uchida M, Klem MT, Allen M, Suci P, Flenniken M, Gillitzer E, Varpness Z, Liepold LO, Young M, Douglas T. Biological containers: Protein cages as multifunctional nanoplatfoms. *Adv Mater* 2007;19(8):1025-1042.
6. Curtis AR, Fey C, Morris CM, Bindoff LA, Ince PG, Chinnery PF, Coulthard A, Jackson MJ, Jackson AP, McHale DP, Hay D, Barker WA, Markham AF, Bates D, Curtis A, Burn J. Mutation in the gene encoding ferritin light polypeptide causes dominant adult-onset basal ganglia disease. *Nat Genet* 2001;28(4):350-354.
7. Levi S, Cozzi A, Arosio P. Neuroferritinopathy: a neurodegenerative disorder associated with L-ferritin mutation. *Best Pract Res Clin Haematol* 2005;18(2):265-276.
8. Zandman-Goddard G, Peeva E, Shoenfeld Y. Gender and autoimmunity. *Autoimmun Rev* 2007;6(6):366-372.
9. Vidal R, Delisle MB, Rascol O, Ghetti B. Hereditary ferritinopathy. *J Neurol Sci* 2003;207(1-2):110-111.
10. Ingrassia R, Gerardi G, Biasiotto G, Arosio P. Mutations of ferritin H chain C-terminus produced by nucleotide insertions have altered stability and functional properties. *J Biochem* 2006;139(5):881-885.
11. Almiron M, Link AJ, Furlong D, Kolter R. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* 1992;6(12B):2646-2654.
12. Grant RA, Filman DJ, Finkel SE, Kolter R, Hogle JM. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* 1998;5(4):294-303.
13. Dautant A, Meyer JB, Yariv J, Precigoux G, Sweet RM, Kalb AJ, Frolow F. Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from *E-coli*. *Acta Crystallogr D* 1998;54:16-24.
14. Luzzago A, Cesareni G. Isolation of Point Mutations That Affect the Folding of the H-Chain of Human Ferritin in *Escherichia-Coli*. *Embo J* 1989;8(2):569-576.
15. Tong GJ, Hsiao SC, Carrico ZM, Francis MB. Viral Capsid DNA Aptamer Conjugates as Multivalent Cell-Targeting Vehicles. *J Am Chem Soc* 2009;131(31):11174-11178.
16. Fan RL, Chew SW, Cheong VV, Orner BP. Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* 2010;6(14):1483-1487.
17. Carrondo MA. Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *Embo J* 2003;22(9):1959-1968.
18. Andrews SC, Smith JM, Guest JR, Harrison PM. Amino acid sequence of the bacterioferritin (cytochrome b1) of *Escherichia coli*-K12. *Biochem Biophys Res Commun* 1989;158(2):489-496.
19. Smith JMA, Ford GC, Harrison PM, Yariv J, Kalb AJ. Molecular-Size and Symmetry of the Bacterioferritin of *Escherichia-Coli* - X-Ray Crystallographic Characterization of 4 Crystal Forms. *Journal of Molecular Biology* 1989;205(2):465-467.
20. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and clustal X version 2.0. *Bioinformatics* 2007;23(21):2947-2948.
21. Holm L, Park J. DaliLite workbench for protein structure comparison. *Bioinformatics* 2000;16(6):566-567.

22. Jin L, Wells JA. Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting. *Protein Sci* 1994;3(12):2351-2357.
23. Cunningham BC, Wells JA. High-Resolution Epitope Mapping of High-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science* 1989;244(4908):1081-1085.
24. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* 1998;280(1):1-9.
25. Hu Z, Ma B, Wolfson H, Nussinov R. Conservation of polar residues as hot spots at protein interfaces. *Proteins* 2000;39(4):331-342.
26. Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein-protein interfaces. *Sci STKE* 2004;2004(219):pl2.
27. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *P Natl Acad Sci USA* 2002;99(22):14116-14121.
28. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 2007;372(3):774-797.
29. Young L, Dong QH. Two-step total gene synthesis method. *Nucleic Acids Research* 2004;32(7).
30. Fan R, Boyle AL, Cheong VV, Ng SL, Orner BP. A Helix Swapping Study of Two Protein Cages. *Biochemistry* 2009;48(24):5623-5630.
31. Liu X, Jin W, Theil EC. Opening protein pores with chaotropes enhances Fe reduction and chelation of Fe from the ferritin biomineral. *Proc Natl Acad Sci U S A* 2003;100(7):3653-3658.
32. Bellapadrón G, Stefanini S, Zamparelli C, Theil EC, Chiancone E. Iron Translocation into and out of *Listeria innocua* Dps and Size Distribution of the Protein-enclosed Nanomineral Are Modulated by the Electrostatic Gradient at the 3-fold "Ferritin-like" Pores. *Journal of Biological Chemistry* 2009;284(28):19101-19109.
33. Pieraccini S, Saladino G, Cappelletti G, Cartelli D, Francescato P, Speranza G, Manitto P, Sironi M. In silico design of tubulin-targeted antimetabolic peptides. *Nat Chem* 2009;1(8):642-648.
34. Turk JA, Smithrud DB. Synthesis and physical properties of protein core mimetics. *J Org Chem* 2001;66(25):8328-8335.
35. Jencks WP. On the Attribution and Additivity of Binding-Energies. *P Natl Acad Sci-Biol* 1981;78(7):4046-4050.
36. Ardejani MS, Li NX, Orner BP. Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* 2011;50(19):4029-4037.
37. Levi S, Luzzago A, Franceschinelli F, Santambrogio P, Cesareni G, Arosio P. Mutational Analysis of the Channel and Loop Sequences of Human Ferritin H-Chain. *Biochem J* 1989;264(2):381-388.
38. Chamberlain AK, Lee Y, Kim S, Bowie JU. Snorkeling preferences foster an amino acid composition bias in transmembrane helices. *Journal of Molecular Biology* 2004;339(2):471-479.
39. Segrest JP, Jones MK, Deloof H, Brouillette CG, Venkatachalapathi YV, Anantharamaiah GM. The Amphipathic Helix in the Exchangeable Apolipoproteins - a Review of Secondary Structure and Function. *J Lipid Res* 1992;33(2):141-166.
40. Andrews SC, Smith JMA, Hawkins C, Williams JM, Harrison PM, Guest JR. Overproduction, Purification and Characterization of the Bacterioferritin of

- Escherichia-Coli and a C-Terminally Extended Variant. Eur J Biochem 1993;213(1):329-338.
41. Gerl M, Jaenicke R, Smith JMA, Harrison PM. Self-Assembly of Apoferritin from Horse Spleen after Reversible Chemical Modification with 2,3-Dimethylmaleic Anhydride. Biochemistry 1988;27(11):4089-4096.
 42. Stefanini S, Vecchini P, Chiancone E. On the Mechanism of Horse Spleen Apoferritin Assembly - a Sedimentation-Velocity and Circular-Dichroism Study. Biochemistry 1987;26(7):1831-1837.

CHAPTER 3 Additional insight into maxi-ferritin self-assembly and stability: Further mutagenesis at the BFR two-fold symmetry interface, wild-type BFR pH stability and thermodynamics with DSC

3.1 Introduction

3.1.1.1 The influence of pH on protein structure

Proteins contain acidic and basic functionality and therefore possess net charges that can be altered when solution conditions change in pH. In addition the effect of pH may affect protein stability when the charges of individual ionizable groups on the protein change their protonation state. Although solvation and the associated hydrophobic effect contribute significantly to a folded protein's stability, Coulombic interactions operate at longer distances. Therefore the effects of pH on denaturation equilibria can be explained in terms of these long-range electrostatic interactions. The long-range electrostatic interactions are nearly zero for any denaturation processes when the pH of the solution is near the isoelectric pH of the protein. However, the effect is significant at extreme pH, where most of the charged sites on a protein molecule carry the same charge. The mutual repulsion between these charges in a compact conformation will favor transition to a denatured state to keep the same charges apart.

Changing the pH can help to identify key electrostatic forces controlling the protein structure as changes in proton concentration can change the net protein charge. In oligomeric proteins, the protonation and deprotonation of amino acid residues as a result of pH change can give rise to repulsive interactions that lead to loss of quaternary, tertiary and even secondary structure. Therefore, protein structural changes induced by changes in pH may lend insights into the electrostatic forces

involved in subunit recognition and association, and also on the possible mechanisms for the assembly and disassembly processes.

3.1.1.2 Differential scanning calorimetry (DSC) to study protein folding and stability

Differential scanning calorimetry (DSC) is an experimental technique which measures the differences in heat required to maintain equal temperature between a sample and reference cell at various temperatures which relates to differences in apparent heat capacity. The data from DSC studies will thereby provide direct information about the energetics of thermally-induced processes in the sample. DSC can not only be used to determine thermal transition (melting) temperatures for samples in solution, solid, or mixed phases, but it also reveals absolute thermodynamic data for these thermally induced transitions. Typical data from a representative DSC experiment is shown in Figure 3.1.

Figure 3.1 Raw DSC data for thermal unfolding of hen egg white lysozyme using a Microcal-MCS system. The inset shows the same data after subtraction of the buffer baseline and concentration normalization. The figure is taken from Cooper *et al*¹.

The output from a DSC experiment is a thermogram which shows the excess heat capacity (C_p) as a function of temperature. There are three thermogram regions for globular proteins with simple transitions: the pre-transition baseline, the endothermic unfolding transition and the post-transition baseline (Figure 3.2). The C_p reflects the difference in heat capacity between the protein and the solvent (mainly water). The apparent C_p in the pre-transition state usually appears negative as protein often has a lower heat capacity compared to that of water. The pre-transition baseline with a positive slope indicates a gradual increase in heat capacity with temperature. Unfolding of the protein is reflected by an increase in C_p as more heat energy is taken up to denature the protein without a corresponding change in temperature. The C_p will reach a maximum at approximately the mid-point (T_m) temperature of the process, and will subsequently decrease to establish the high temperature baseline. The post-transition baseline represents the relative heat capacity of the unfolded polypeptide, and is usually found at a higher level (positive ΔC_p) than that for the folded protein and has a smaller slope^{2,3}.

Figure 3.2 A normalized DSC trace for the reversible unfolding of a monomeric protein. The maximum of the heat capacity curve is the T_m (60°C). The dashed lines are the linear extrapolations of the pre- and post-transition baselines into the transition region. The difference between these values at the T_m is the ΔC_p . The thick black line is the theoretical

“progress” baseline. The area between the DSC trace and the progress baseline is the calorimetric enthalpy, ΔH_{cal} . The figure is taken from Ladbury *et al*².

The calorimetric enthalpy (ΔH_{cal}) is the total integrated area under the thermogram peak which represents the total heat uptake by the sample undergoing the transition. This enthalpy is a “model-free” measure of the absolute enthalpy of the process involved and can shed light on the nature of the transition (see below).

Another independent estimate of the unfolding enthalpy may be made from the van't Hoff analysis of the shape of the peak in the DSC thermogram⁴⁻⁷. The estimate is independent of the sample concentration or absolute calorimetric enthalpy. The validity of the two-state assumption can be derived by determining the size of the “cooperative unfolding unit” through comparing the heat uptake per mole of cooperative unit in the transition and the calorimetric enthalpy (ΔH_{cal}). The van't Hoff enthalpy (ΔH_{vH}) depends on the shape of the endotherm and assumes a true two-state reversible equilibrium. Ideally, in a cooperative two-state transition, $\Delta H_{\text{vH}} = \Delta H_{\text{cal}}$ and the unfolding transition is reversible. However, in many cases such as higher oligomeric state protein complexes, this is not a valid assumption and the ratio, $\Delta H_{\text{cal}} : \Delta H_{\text{vH}}$ is used to indicate the number of molecules involved in the cooperative unfolding unit⁸. If a protein is composed of two identical domains for instance, which unfold independently with the same T_m and ΔH_{cal} , the ratio of $\Delta H_{\text{cal}} : \Delta H_{\text{vH}}$ will be 2.0. If the protein had only a single domain, then the ratio would be 1.0. On the other hand, if the protein dimerized and the dimer went through a single coupled transition, the ratio would be 0.5. Hence, it is clear that the ΔH_{cal} refers to heat change per mole while ΔH_{vH} is the heat change per unfolding unit (cooperative unfolding unit) and the ratio of $\Delta H_{\text{cal}} : \Delta H_{\text{vH}}$ can be regarded as the number of cooperative units per mole.

3.1.1.3 Introduction to FoldX

The computer program FoldX employs a semi empirical force field and has been developed for fast evaluation of the effect of mutation on the stability of proteins or protein complexes. It calculates the free energy of unfolding (ΔG) of a desired protein using the following equation:

$$\Delta G = W_{vdw} \cdot \Delta G_{vdw} + W_{solvH} \cdot \Delta G_{solvH} + W_{solvP} \cdot \Delta G_{solvP} + \Delta G_{wb} + \Delta G_{hbond} + \Delta G_{el} + \Delta G_{Kon} \\ + W_{mc} \cdot T \cdot \Delta S_{mc} + W_{sc} \cdot T \cdot \Delta S_{sc}$$

In this expression, ΔG_{vdw} is the sum of the van der Waals contributions of all atoms in the folded state with respect to those with the solvent (i.e. in the unfolded state). ΔG_{solvH} and ΔG_{solvP} are the differences in solvation energy for apolar and polar side chains respectively for the unfolding transition. ΔG_{wb} is the additional stabilizing free energy provided by a water molecule to establish more than one hydrogen bond to the protein (water bridges). Hence, ΔG_{wb} cannot be included in non-explicit solvent approximations. ΔG_{hbond} is the free energy difference between the formation of an intramolecular hydrogen bond in the folded state and intermolecular hydrogen bond formation with water in the unfolded state. ΔG_{el} is the contribution of electrostatic interactions between the charged groups, including the helix dipole. When oligomeric proteins or protein complexes are analyzed, ΔG_{Kon} is applied to determine the effect of electrostatic interactions on the association constant K_{on} (this applies only to the subunit binding energies). ΔS_{mc} is the entropy cost of fixing the backbone in the folded state, and ΔS_{sc} is the entropic cost of fixing a side chain in a particular conformation. The energy values of ΔG_{vdw} , ΔG_{solvH} , ΔG_{solvP} , and ΔG_{hbond} for each atom type are derived from a set of experimental data. The terms W_{vdw} , W_{solvH} , W_{solvP} , W_{mc} and W_{sc} are the weighting factors for each raw energy terms. Values for entropic terms of ΔS_{mc} and ΔS_{sc} are taken from theoretical estimates ⁹.

The stabilizing or destabilizing effect of mutations on the stability of a protein or a protein complex can be assessed by the change in the free energy of unfolding ($\Delta\Delta G$) using the following equation:

$$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wt}$$

where ΔG_{mutant} and ΔG_{wt} are free energies of unfolding for mutant and wild type protein respectively.

3.1.2 Previous studies on maxi-ferritin pH and thermodynamic stability

3.1.2.1 pH stability of maxi-ferritins

Only a few reports have been published on the stability and assembly of maxi-ferritin at various pHs. Horse spleen apoferritin is known to assemble spontaneously upon neutralization of its monomer subunits solubilized at acidic and capsule assembly starts around pH 3.5¹⁰.

Crichton *et al*¹¹ examined the dissociation of horse spleen apoferritin as a function of pH by sedimentation velocity. It was found that the oligomer is stable at the pH range of 2.8 ~ 10.6. Both oligomer and monomers are detected at pH 1.6 ~ 2.8 and 10.6 ~ 13.0, but monomer is the sole species detected at pHs ranging from 1.0 ~ 1.6. Monomer reassociation occurs at the pH range 3.1 ~ 4.3 when dialyzed back to high pH from low pH. The wild-type *R. capsulatus* BFR was discovered to maintain its 24-mer oligomeric form over the range of pH 3.0 ~ 10.0. Outside the pH range this range, some of the BFR are dissociated into the disassembled state, and BFR is completely unfolded at pH 12.0¹².

Similar results were obtained by Kim *et al*¹³, where the intact hollow spherical structure of the horse spleen apoferritin shell was stable over a wide pH range (3.4 ~ 10.0), and the subunits undergo aggregation below pH 0.8. The low pH dissociation process begins through the formation of hollow spherical structures with two holes,

followed by “headphone”-shaped structures, and ultimately, rod-like oligomers (mainly trimers) or monomers.

The denaturation of recombinant horse L-chain apoferritin in acidic solution was investigated by Yoshizawa¹⁴. Two mutants that lack four and eight N-terminal amino acid residues were studied to examine the effect of N-terminal residues on the protein’s stability. The mutant lacking eight amino acid residues dissociated easily into subunit dimers at or below pH 2.0. The intra-subunit hydrogen bonds of N-terminal residues were proposed to affect the oligomer formation of apoferritin subunits.

3.1.2.2 Calorimetric studies of maxi-ferritins

The thermal stability of recombinant H and L chain human ferritins, self-assembled from monomers, was analyzed within the pH range of 2.0 ~ 2.8 using DSC to obtain accurate measurement of the energies involved in the folding and unfolding processes¹⁵. At pH 2.0 ~ 2.8, the H subunit displayed properties of a molten globule-like state. It had a low transition enthalpy (~ 170 kJ/mol) with T_m of thermal unfolding at approximately 50 °C, and a wide transition range (20 °C). In contrast to the H chain, the L-ferritin subunit displayed two calorimetric transitions with (T_m of 35 °C and 40 °C) at pH 2.0. These domains showed similar unfolding enthalpies. At higher pH (2.2 ~ 2.8) a single transition with a higher T_m was observed. As there are significant differences in intra-subunit hydrogen bonds, which are about 50% more abundant in H chain and more salt bridges in L chain, it is proposed that the H- and L- subunits are stabilized by protein-protein interactions, and the presence of additional salt bridges contribute to the higher stability of L-chain as compared to the H-chain.

Comparison of the thermal stability of horse spleen apoferritin with human recombinant H-chain ferritin was also conducted by DSC ¹⁶. It was found that the thermal denaturation process displayed substantial reversibility for both proteins up to several degrees Celsius below the T_m (93 °C and 77 °C) however the thermograms were influenced by the exothermic contribution of aggregation or precipitation at temperatures higher than the T_m . The acid dissociation of apoferritin produces stable dimeric subunits that are reversible up to at least the T_m during thermal unfolding (at pH 3.5, T_m is 42 °C and 50 °C for horse spleen and human H ferritin respectively). The results proved that the stabilization of the oligomeric assembly arises mainly from interactions between dimers, and correlate to previous findings that dimers are the building blocks of the ferritin molecules.

3.2 Aims and objectives

This research focuses on the charged interactions along the two-fold axis which contribute significantly to the formation of the dimeric building blocks of ferritin self-assembly. To further understand the role that the key amino acids play in the two-fold axis discovered in Chapter 2, interfacial amino acids (R30, D56 and E60) were mutated to opposite charged residues on the hypothesis that such mutation would lead to the destruction of salt bridges, and the repulsive interactions at the interface would result in overall destabilization (Figure 3.3). These three residues were specifically chosen because these appear to interact with one another in the crystal structure by forming an extended salt-bridge network with K33 which was not included due to predicted insignificant binding energy in the interaction (see below). FoldX calculations ¹⁷ were employed to calculate the free energy of unfolding and determine the point mutations which would impose the highest destabilizing effects. The resulting mutants were examined using transmission electron microscopy (TEM), size

exclusion chromatography (SEC), dynamic light scattering (DLS), native polyacrylamide gel electrophoresis (native PAGE) and temperature dependent circular dichroism (CD) to determine the role the individual interfacial interactions at the two-fold axis play in the stability and structure of the ferritin.

Figure 3.3 (*Left*) BFR protein cage viewed along the two-fold axis. The positions of the interactions of interest are boxed. Note that they are related by the C₂ symmetry axis. (*Right*) The expansion of one of these regions shows the interactions between four charged residues (K33, R30, E60 and D56). The heavy atom distances between the interaction groups are indicated in Å.

The thermal stability for wild-type BFR, 24-mer and dimer, and the mutants (R30A, R61A, Y114A, E128A) which were found to be energetically important in the previous studies (see Chapter 2) were investigated using differential scanning calorimetry (DSC). We aim to reveal the unfolding pathway of the complicated 24-meric macromolecule dissociation.

In addition, protein complexes were subjected to dissociation under various pH conditions through salt bridge destruction. Therefore, our approach in this project focused on determining the effect of salt bridges on the overall self-assembly of BFR using a variety of techniques. The energetic stability upon acid or base dissociation to structural changes and will thereby help to shed light on the overall structural energetics governing the self-assembly of the protein.

3.3 Materials and methods

3.3.1 Additional mutagenesis studies focusing on the BFR two-fold symmetry interface

3.3.1.1 Computed unfolding free energy using FoldX

Computational prediction of virtual mutation experiments with BFR were carried out using FoldX 3.0 Beta 3 (2009)¹⁷. The server is publicly available at <http://foldx.embl.de/>. The unfolding free energy of a protein is calculated in FoldX based on the energy function below, which includes terms that have been found to be important for protein stability using the equation discussed above.

PDB entry 2vxi¹⁸ was chosen for computational mutagenesis experiments because of all the BFR crystal structures it has the highest resolution. The dimer was generated by deleting all the subunits except two chains, C and D, which are centered at a C2 symmetry axis. The RepairPDB protocol was used to minimize the energy of the structure. The side chains were rotated while the backbones were fixed and the buffer sulfate and heme ligands were removed. The repaired PDB file was then used with the BuildModel protocol to explore the changes in unfolding free energies (relative to that of the wild type) of the possible point mutants (at 298 K, pH 7.0, 50 mM ionic strength). The calculation for each mutant was repeated at least three times, the $\Delta\Delta G$ values were averaged and the standard deviation was calculated.

3.3.1.2 Gene construction, expression and protein purification and analysis using CD, TEM, DLS, SEC and native PAGE

The genes of the mutants were prepared using site directed mutagenesis. The sequences of primers used are listed in Appendix 5. The procedures of gene expression, protein purification, temperature dependent circular dichroism, dynamic

light scattering, size exclusion chromatography and native PAGE are as described in Chapter 2.

3.3.2 pH stability of BFR 24-mer

3.3.2.1 Sample preparation for acid denaturation

The 24-mer BFR isolated by SEC (3.5 mg/mL in 50 mM phosphate, 0.15 M NaCl, pH 7.0) was dialyzed (24 h, 4 °C) to various pH conditions (50 mM phosphate, 0.15 M NaCl, pH 1.0 ~ 12.0). The pH of the buffers was measured with an Schott pH electrode connected to a Schott Lab 859 ion meter. The samples were analyzed by native PAGE, size exclusion chromatography and circular dichroism.

3.3.2.2 Native PAGE

Protein samples (0.45 mg/mL in 50 mM phosphate, 0.15 M NaCl, pH 1.0 ~ 12.0) were mixed with an equal volume of native standard buffer, and loaded into wells of a 15 % polyacryamide gel. Gel electrophoresis was carried out on a Bio-Rad Laboratories electrode (S1, 80 V, 20 min; S2, 120 V, 120 min). Tris-glycine electrophoresis buffer at pH 8.3 was used as gel running buffer. The gels were stained (0.025 % commassie brilliant blue, 1h) and subsequently destained (1h) before the gels were visualized.

3.3.2.3 Size exclusion chromatography at different pH

Size exclusion of the protein in various pH conditions was performed on a GE ÄKTATMFPLC system with a Superdex 200 10/30 GL gel filtration column at a flow rate of 0.5 mL/min (running buffer, 50 mM NaH₂PO₄·H₂O, 150 mM NaCl, pH corresponding to the protein sample above). Calibration was performed by injection of several protein standards of various molecular weights: ferritin, aldolase, aprotinin, ribonuclease A, carbonic anhydrase, ovalbumin, conalbumin. Blue dextran was also

used to determine the column void volume. The calibration details are described in Chapter 4.

3.3.2.4 Circular dichroism analysis

CD spectra were recorded on a Jasco J-720 spectropolarimeter with a Peltier temperature controller. The UV absorbance of protein was determined (Nanodrop 2000c, Thermo Scientific) at 280 nm using molar extinction coefficient $21430 \text{ M}^{-1}\text{cm}^{-1}$. The protein was diluted (150 - 200 $\mu\text{g/mL}$). Secondary structures of protein were determined by far-UV (190 - 250 nm) using a 0.1 cm pathlength quartz cuvette. Each CD spectrum was collected as an average of two successive scans with buffer background scans subtracted.

3.3.3 Thermodynamic analysis using differential scanning calorimetry

The protein samples were dialyzed for 24 h in phosphate buffer (50 mM NaH_2PO_4 , pH 7.2) at 4 °C prior to DSC experiment. DSC data was monitored and collected on a VP-DSC microcalorimeter equipped with degassing equipment. Protein concentration was determined at 280 nm (Nanodrop 2000c, Thermo Scientific) using molar extinction coefficient $21430 \text{ M}^{-1}\text{cm}^{-1}$ (calculated based on the equation described in reference¹⁹). The concentration range is 100 - 250 $\mu\text{g/mL}$. Baseline and DSC scans were measured and collected using a temperature range of 15 °C - 110 °C and a scan rate of 30 °C/hr - 90 °C/hr.

All curves were corrected for the instrumental baseline obtained by filling both cells with the buffer. Integration and deconvolution were performed using the ORIGIN software provided by Microcal Incorporation after normalizing the data for protein concentration and subtraction. Non-2-State with zero ΔC_p was used to fit the curves in the thermodynamic analysis. The value of T_m (temperature at which excess

heat capacity reaches a maximum) is obtained by analyzing the resulting thermal profiles.

3.4 Results and discussion

3.4.1 Mutagenesis studies on BFR two-fold interface

3.4.1.1 Computational assisted mutation determination by FoldX

Mutations of amino acid residues (R30, D56 and E60) that form salt bridges at the dimer interface were designed so the opposite charge was achieved at that residue at neutral pH. For example, the positively charged amino acid R30 can be mutated to the negative charged amino acids, Asp (D) or Glu (E). The semi empirical force field FoldEF was used to calculate the free energy differences of all possible mutants. In addition, these amino acids were virtually mutated to alanine, so as to compare to the previous alanine shaving studies described in Chapter 2 using the same computational approach.

Mutating the residues to the opposite charge may destroy the salt bridges and hydrogen bonds at the two-fold interface and generate repulsion. Hence, it is proposed that the mutations would destabilize the protein complex. The pH is set to 7.0 in FoldX, so it will not affect the protonation of charged amino acid residues. As predicted, most calculations of $\Delta\Delta G$ using FoldX are positive, which are interpreted as destabilization of the structure upon mutation. Subsequently, the mutants with the highest positive $\Delta\Delta G$, R30D, D56R, E60H and D56R E60H were selected for mutagenesis and further characterization. Surprisingly, mutations for E60 showed stabilizing effect ($\Delta\Delta G < 0$) when mutating to the positive charged amino acid residues.

<i>Mutants</i>	$\Delta\Delta G$ (kcal/mol)	<i>SD*</i>
R30D	12.46	0.43

R30E	11.54	0.39
R30A	5.45	0.66
D56R	6.76	0.31
D56H	6.23	0.12
D56K	5.6	0.17
D56A	4	0.04
E60H	1.27	0.11
E60R	-1.23	0.45
E60K	-0.45	0.00
E60A	-0.15	0.03
D56RE60H	8.5	0.52
D56RE60R	6.6	0.13
D56KE60K	6.72	0.46
D56RE60K	7.29	0.31
D56KE60R	5.86	0.21
D56AE60A	4.38	0.16

*The standard deviation of three repeats of the same mutation.

Table 3.1 Results of virtual mutation experiments of the charged amino acids at the two-fold symmetry interface. The calculations were performed using FoldX¹⁷.

3.4.1.2 Gene construction, expression and protein purification

All four mutants were constructed using QuikChange site-directed mutagenesis kit (Stratagene) with the BFR vector as the template using the primers in Appendix 5. The plasmids harboring the respective gene inserts were obtained by miniprep and sent for sequencing. The resulting sequences were aligned with the designed gene sequences using the Blast web tool that is available online: <http://blast.ncbi.nlm.nih.gov/>. The alignment results are listed in Appendix 6. All the mutations were successfully introduced into the expression vector.

A series of SDS-PAGE gels were carried out to verify each step of the expression and purification procedure. The gels (Figure 3.4) demonstrate that all four BFR derived charged mutants are expressed as soluble proteins, and pure protein bands are ~ 19 kD in size. Therefore, SDS-PAGE has successfully verified that the target proteins were expressed and purified.

Figure 3.4 Electrophoretic monitoring of the purification of the BFR-derived proteins. Lane 1: ladder. Lane 2: before induction. Lane 3: post induction. Lane 4: soluble fraction. Lane 5: insoluble fraction. Lane 6: flow through. Lane 7: wash. Lane 8: cleavage.

3.4.1.3 *Transmission electron microscopy*

The proteins were analyzed to determine whether each possessed the ability to assemble into a nano-cage. Samples of the proteins were prepared and treated with uranyl acetate negative stain. Electron micrographs indicated that all proteins form nano-cages under TEM conditions (Figures 3.5). Image analysis of the micrographs indicates that the size of the native protein structures as comparable as those described in the literature²⁰ (Table 3.2).

Figure 3.5 All four BFR-derived mutants can form nano-cages under TEM conditions. Scale bars indicate 50 nm.

<i>Mutants</i>	<i>Particle Diameter</i>	
	<i>(nm)</i>	<i>S.D.</i>
R30D	12.1	0.8
D56R	12.1	0.7
E60H	13.0	1.1
D56RE60H	10.8	0.9

Table 3.2 Average particle diameters of 100 BFR-derived mutants measured using ImageJ. Scale bars indicate 50 nm.

3.4.1.4 Size exclusion chromatographic analysis of assembly in solution

The proteins were analyzed by size exclusion chromatography, which separates the different structures based on size, to determine the effect of mutation on the ability of the proteins to assemble in solution (Figure 3.6 A). The wild-type protein BFR, consistent with the literature^{21,22}, forms a mixture of 24-mer and dimer. The mutant BFR E60H exists in a similar ratio to as the wild-type protein BFR, while the other mutants (R30D, D56R, D56RE60H) were observed to form a homogenous population of the lower order complex corresponding to dimer. Mutation of negative charged glutamate to a histidine residue at position 60 does not tip the equilibrium mixture to

solely a dimeric state possibly because of the relatively high pKa of His with respect to Lys or Arg. However, these data taken together with the FoldX prediction of subtle changes in the binding free energy ($\Delta\Delta G$), suggests that the interactions between Glu60 and either Lys33 or Arg30 do not contribute significantly to the overall stability of protein structure as compared to the other two positions (Figure 3.3). However when viewed through a wide lens, these results confirm that even a single mutation can affect the wider quaternary structure.

Figure 3.6 All the mutants form dimer solely in solution except BFR E60H. (A) SEC chromatograms of BFR wild type and four mutants showing that three mutants form only dimer in these conditions. (B) Hydrodynamic diameters for proteins as measured by dynamic light scattering. Error bars represent the standard deviation. (C) Native PAGE of three mutants exhibiting sole dimers coincides with the SEC results.

3.4.1.5 Dynamic light scattering (DLS) and native PAGE

DLS was used to determine whether the three mutants, R30D, D56R, D56RE60H form 24-mer in solution. Comparison of the DLS data for the three mutants and BFR wild type supported the SEC results. These mutants form particles that are smaller than BFR (Figure 3.6 B).

Native PAGE was performed to further confirm the previous data. Consistent with the SEC experiment, only E60H forms 24-mer. It was observed that a greater migration distance for the BFR D56R dimer as compared to BFR dimer. The faster migration rate is not only due to changes (+2 per monomer) in net charge upon mutation but also possibly the size of the protein is smaller, because the double

mutant D56RE60H which forms dimer too was observed to have similar migration rate as wild-type dimer (Figure 3.6 C).

Although the TEM results do not coincide with the solution data, we should not eliminate the potential of mutants assembling into 24-mers under “forcing conditions”. The same phenomenon has been discovered and discussed in other mutation studies of BFR (see Chapter 2).

3.4.1.6 The secondary structure and thermal stability as monitored by circular dichroism

Circular dichroism demonstrated that all the mutants fold into α -helical structures as evidenced by minima at 208 nm and 222 nm. This suggests that mutations at the two-fold symmetry axis do not completely destroy the helicity and mostly likely the tertiary structure of the proteins is intact. With that said, all the mutants exhibited reduced secondary structure as compared to the parent BFR protein.

The thermal stability and folding reversibility of the BFR mutants was analyzed by heating the samples beyond denaturation, followed by gradual cooling to the starting temperature, and re-measuring the secondary structure. Despite having the lowest depression in stability, R30D exhibited a higher degree of thermal folding reversibility than wild-type BFR (Figure 3.7, Tabel 3.3) which may be due to a more simple oligomerization state. The largest reduction in reversibility is observed for D56RE60H, which may be due to aggregation kinetically blocking the mutant from returning to the folded state. Though mutant E60H showed the highest thermal stability, even higher than the wild-type BFR, the reversibility is much lower.

The positive $\Delta\Delta G$ values calculated from FoldX indicate changes in the binding energy between the wild type and the mutation; the higher the $\Delta\Delta G$ (bind) value, the greater the destabilizing effect on the dimer structures. Consistent with the

computational calculation, comparison of mutant R30D, D56R, D56RE60H with mutant E60H demonstrated that high $\Delta\Delta G$ value correlate to greater destabilizing effect. Interestingly, mutant E60H which is estimated to possess little destabilizing effect by virtual mutation experiments has a stable structure with higher thermostability as compared to wild-type BFR. This reflects the complexity of interactions that exist in natural systems. Further experiments, such as denaturation with guanidine or urea are necessary to understand and confirm the enhanced stability of E60H.

Figure 3.7 The role mutation plays in the thermal stability and folding reversibility of four charged BFR. (**Left**) Thermal transitions of the four charged BFR mutants as monitored by CD at 222 nm. (**Right**) CD spectra of the BFR derivatives before denaturation (solid lines) and after slow cooling post-thermal denaturation (dashed lines). The coloring follows that of the left. Data is the average of at least three replicates.

<i>Proteins</i>	<i>Assembly state</i>	<i>T_m</i> (°C)	<i>ΔT_m</i> (°C)	<i>ΔΔG</i> (kcal/mol)	<i>Reversibility</i> (%)
BFR wild type	24-mer & Dimer	68.2	-	-	71
R30D	Dimer	59.2	9.0	12.5	81
D56R	Dimer	64.4	3.8	6.8	66
E60H	24-mer & Dimer	72.3	-4.1	1.3	57
D56RE60H	Dimer	60.8	7.4	8.5	53

Table 3.3 Solution assembly state, melting temperature and reversibility of BFR and derived mutants taken from Figures 3.6 and 3.7. $\Delta\Delta G$ free binding energy calculated by FoldX is taken from Table 3.1.

Comparing this study to the alanine shaving mutagenesis performed in Chapter 2, indicates that the interactions at three-fold axis play more important roles in protein self-assembly and stability than interactions at two-fold axis which is surprising

considering the presumed importance of the dimer in ferritin assembly. All the mutations at the three-fold interface (E128A, Y114A, R61A) formed solely dimer and resulted in either the greatest affect on stability (either higher or lower) compared to the wild-type BFR. In contrast, among the selected key amino acid residues along the two-fold axis (R30, K33, Y45, D56), only R30 plays a significant role. Although hydrogen bonds and salt bridging interactions that are predicted to control the formation of dimers may be destroyed by mutation to their opposite charges, the dimer still remains.

It has usually been assumed that a dimer is a major building block in the ferritin mechanism of assembly. This dimer has been assumed to be similar to the one that exists at the two-fold axis in the cage. However in this thesis we have shown how manipulating the two-fold symmetry related protein-protein interface can result in destabilization of the cage with no disruption of the dimer. In addition, our lab has shown that in the course of attempting to rationally stabilize the ferritin, we generated a protein that was more thermally stable by manipulating the two-fold interface but this protein formed more dimer and less cage²³. There are a number of possible explanations a) the mutation could be inducing a conformational change within the monomer which changes the other interfaces; b) the dimer geometry could be changed resulting in a tighter dimer interaction but a dimer that can't fit into an assembled cage. We refer to this as the "Arch and Keystone Hypothesis"²³; c) the dimer is not the two-fold symmetric dimer. This dimer could be a dimer that is found in the cage or it could not be. This could be explored by calculating the energies of all the possible interfaces (see Figure 2.12) or by generating a geometry-specific cross linking detection strategy like bipartite FLaSH²⁴ to monitor assembly at specific interfaces.

3.4.2 pH stability analysis of wild-type BFR

The structural stability of wild-type BFR as a function of pH was studied using size exclusion chromatography, circular dichroism and native PAGE. The isolated 24-mer was incubated at pH range 1.0 ~ 12.0 at 4°C for 24 h, a time sufficient to reach an equilibrium condition and then analyzed by the various characterization techniques.

3.4.2.1 pH-dependent behavior of wild-type BFR

Native PAGE gels were used because they allowed small amounts of different oligomeric proteins to be investigated. However, as the running buffer and the sample buffer do not have the same pH, such gels cannot provide quantitative data because disassembly or reassembly may happen as the protein passes through the running buffer. Albeit this quantitative limitation, native PAGE gels do provide useful qualitative data based on the assumption that the time of experiment is sufficiently short compared to the equilibration time (see above). SEC allows the characterization of both small and large oligomers, and differences between population of small particles (such as monomer and dimer) can be observed. Hence, SEC can aid our understanding of the oligomerization state and help to further elucidate the role pH plays in the self-assembly of protein.

There is no significant alteration in the quaternary structure when the protein is incubated at pH values 7.0 to 3.5 as shown by the distinctive single band on native gel (Figure 3.8A). The electrophoretic mobility of the band at pH 7.0 to 3.5 corresponds to the 24-mer of the BFR mixture (last lane in Figure 3.8A). A fainter band which corresponds to the dimer at the pH range was also observed. In contrast, incubation of the purified BFR 24-mer at pH 3.0 to 1.0 resulted in marked changes to the quaternary structure. Broad bands with greater migration distance were observed. These broad bands represent a smaller protein which migrates through the porous gel

at a faster rate and corresponds to the dimer band in the BFR mixture. This result shows that there is a slight dissociation from 24-mer to dimer at pH 7.0 to 3.5, but BFR undergoes complete dissociation at pH 3.0 to 1.0. These findings are consistent with the SEC data (Figure 3.9 left).

Calibration of SEC with protein standards at different pHs (Table 3.4) indicates that the 24-mer BFR fraction obtained from BFR mixture (Figure 3.9) forms a peak with elution volumes of 11.3 mL and 16.3 mL after incubating at pH 7.0 to 5.0. This corresponds to 24-mer and dimer of BFR respectively. Results obtained from SEC at these pH correlate with those of the native PAGE as the samples contain mainly 24-mer. On the other hand, the 24-mer BFR fraction, after incubating at pH 3.0, eluted at 18.7 mL. This elution volume represents the monomer of BFR, and leads us to believe that the smaller protein observed in the native PAGE is the monomer instead of dimer and the monomer and dimer are co-eluting in native PAGE. Taken together, SEC data supports the pH dissociation trend observed in native PAGE. Slight differences in the elution volumes at various pHs were observed. This is due to the influence of pH on the performance of Superdex gel filtration column which was confirmed with the calibration. Furthermore, SEC data was not obtained for BFR at pH 4.0 as proteins tend to aggregate close to their isoelectric point (protein pI = 4.8).

Figure 3.8 Native PAGE (15%) of isolated 24-meric BFR at (A) acidic and (B) basic pH. (C) BFR protein boiled at 95°C for 15 min as a denatured control. All the lanes contain 24-meric BFR (0.35 mg/ml) dialyzed to the respective pH, with the exception of the last lane in (A) which is the BFR mixture of 24-mer and dimer before isolation of the 24-mer.

Figure 3.9 Size exclusion chromatography of isolated 24-meric BFR at (left) acidic and (right) basic pH. The elution volume for BFR 24-mer and dimer at pH 7.0 was 11.3 mL and 16.5 mL respectively. Column calibration at the various pHs are summarized in Table 3.4.

Under basic conditions, a slight dissociation of 24-mer to dimer which elutes at 16.2 mL was observed at pH 9.0. The protein was completely dissociated to a lower order dimer complex at pH 10.0. However, the peak from the pH 11.0 sample is asymmetric with a shoulder corresponding to a smaller elution volume and the pH 12.0 sample results a population that is solely a 14.5 mL elution volume peak. These results are surprising, however, native PAGE sheds some light.

The findings from SEC at basic pH mostly correspond to the results from the native PAGE. At pH 7.0 ~ 9.0, the quaternary structure of BFR is not affected as it still retains its 24-mer cage and dimer (Figure 3.8). Between pH 9.0 - 10.0 the dimer state grows in indicating that with increasing pH, the polar interactions are disrupted, preventing formation to a higher oligomerization state. This is consistent with SEC. However, at pH 11.0 a new band appears that is below the dimer band at pH 11.0 and at pH 12.0 this is the sole population. This band was assumed to be either the monomer or denatured protein. Therefore, to provide a control for denatured protein,

BFR was boiled at 95 °C for 15 minutes. This sample resulted in a similar band indicating that the protein begins to and is fully denatured at pH 11.0 and 12.0 respectively. This could be used to explain the SEC data. The shoulder at pH 11.0 and the new peak at 12.0 could be the result of denatured protein, that because it is no longer globular, falls off the calibration curve and elutes at a surprising volume. We sought to confirm the denaturation hypothesis by measuring the secondary structure of BFR at different pH using CD (see next section).

pH 3.0 (log MW = 6.6417 – 3.3548 K_{av}); V_0 = 8.56				
	MW	log MW	K_{av}	V_e
24-mer	444000	5.64738	0.29639	12.33
Dimer	37000	4.5682	0.61807	16.75
Monomer	18500	4.26717	0.7078	17.99
pH 4.0 (log MW = 6.3322 – 3.3822 K_{av}); V_0 = 8.41				
	MW	log MW	K_{av}	V_e
24-mer	444000	5.64738	0.20	11.57
Dimer	37000	4.5682	0.52	16.54
Monomer	18500	4.26717	0.61	17.93
pH 7.0 (log MW = 6.2952 – 3.4534 K_{av}); V_0 = 8.16				
	MW	log MW	K_{av}	V_e
24-mer	444000	5.64738	0.187589043	11.13
Dimer	37000	4.5682	0.500086871	16.08
Monomer	18500	4.26717	0.587256038	17.46
pH 10.0 (log MW = 6.2221 – 3.3906 K_{av}); V_0 = 8.10				
	MW	log MW	K_{av}	V_e
24-mer	444000	5.64738	0.169503923	10.80
Dimer	37000	4.5682	0.487789772	15.86
Monomer	18500	4.26717	0.576573468	17.26
pH 12.0 (log MW = 5.9959 – 3.1587 K_{av}); V_0 = 8.14				
	MW	log MW	K_{av}	V_e
24-mer	444000	5.64738	0.110336531	9.89
Dimer	37000	4.5682	0.451989743	15.31
Monomer	18500	4.26717	0.547291607	16.82

Table 3.4 Size exclusion chromatography calibration at pH 3.0, 4.0, 7.0, 10.0 and 12.0. The elution volumes (V_e) for BFR 24-mer, dimer and monomer at respective pH were determined via gel-phase distribution constant (K_{av}) by the following relation: $V_e = K_{av} (V_c - V_0) + V_0$, where V_0 is the void volume (assumed to be the elution volume of the blue dextran), and V_c is the column volume (24 mL). The calibration curves are presented in Appendix 7.

3.4.2.2 Secondary structure and thermostability of wild-type BFR at different pH

With the knowledge that the oligomerization state is affected by acidic conditions, we sought to determine the effect pH has on secondary structure using CD spectroscopy. The far-UV CD spectra obtained are characteristic for alpha helices,

which exhibit pronounced minimum at approximately 208 nm and 222 nm. The spectra of BFR at different pH conditions indicate that the secondary structure is predominantly α -helical (Figure 3.10). The consistency of the spectra demonstrates that acidic conditions do not significantly perturb the secondary structure content of BFR supporting the supposition that the disassembly of BFR at low pH results in folded monomers and dimers. Retention of the secondary structure made us question whether BFR will undergo reassociation when it is incubated back to physiological condition. Initial experiment suggests that completely dissociated BFR (monomer at pH 3.0) could reform back to a mixture of dimer and 24-mer but the percentage of 24-mer is low.

At pH 7.0 ~ 10.0, the presence of α -helical structure indicates that there is not much disruption of BFR secondary structure at this pH range again indicating that it is possible to disrupt the oligomerization into folded proteins. However, at pH 11.0 and 12.0, the substantial loss of pronounced double minima suggests that the protein is completely denatured as evidenced by drastic changes in α -helicity (Figure 3.10) suggesting that the interactions governing the folding of helical structure are destroyed at high pH. Therefore, this data reinforces the hypothesis that the band shown below the dimer band in the native PAGE at these pHs (Figure 3.8 B) is a denatured protein rather than a monomer as is the 14.5 mL SEC peak.

Under alkaline conditions, BFR exhibits a decreased thermal stability as the pH increases (Figure 3.11) and loss of cooperative melting at pH 11 and 12 further supporting the denaturation hypothesis. However, under acidic conditions, no clear trend of the melting temperature is evident. Surprisingly, the protein shows very high thermal stability at pH 3.5 and 4.5 (~ 90 °C) that is approximately 20 °C higher than the melting temperature at pH 7.0, indicating that stronger interactions are formed at

this pH range. At other pHs the T_m is depressed and the assumption of a two-state transition may not be a good one (Figure 3.11).

In summary, the data for wild-type BFR indicates that it is predominately a 24-mer over the pH range 6.0 ~ 8.0. Beyond this range, some of BFR is in a disassembled state, forming folded dimer between pH 5.0 ~ 6.0 and folded monomer at least to pH 4.0. It forms folded dimer up to pH 10 but is completely unfolded and non-helical between pH 11.0 ~ 12.0. This data correlates to a previous study using the bacterioferritin from *R. capsulatus*¹², and it demonstrates that similar stabilities to those of horse spleen ferritin, which is oligomeric over a pH range 2.8 ~ 10.6 with some of its subunits being monomeric below pH 3.0^{10,11,25}. Knowledge of the stability of these proteins can help direct further applications.

The pKa of Arg and Glu are 12.0 and 4.3 respectively. Under physiological pH, salt bridges are formed between the positively charged guanidinium group on one monomer (R30 or R61) and the negatively charged deprotonated carboxylate on another monomer (E128 or E60). Site directed mutagenesis which involves manipulating and disrupting these intermolecular interactions has demonstrated that these amino acid residues are energetically significant. These residues are hot spots and the stability of BFR is influenced by interactions between hot spots. Therefore, protonation or deprotonation at acidic or alkaline conditions destroys the electrostatic interactions, perhaps causing dissociation from 24-mer into folded dimer or monomer.

Figure 3.10 Far-UV CD spectra of wild-type BFR at different pH, as indicated. The protein concentration was 150 ~ 200 $\mu\text{g/ml}$ in 50 mM phosphate buffer.

Figure 3.11 Thermostability of wild-type BFR at different pH as indicated, monitored by the CD at 222nm. The protein concentration was 150 ~ 200 $\mu\text{g/ml}$ in 50 mM phosphate buffer. The scan rate is 1 $^{\circ}\text{C}/\text{min}$.

Wild-type BFR at different pH	T_m ($^{\circ}\text{C}$)	Assembly State
pH 1.0	-	Monomer*
pH 2.0	-	Monomer*
pH 3.0	73.2	Monomer
pH 3.5	93.9	24-mer + Dimer
pH 4.5	88.9	24-mer + Dimer
pH 5.0	-	24-mer + Dimer
pH 6.0	-	24-mer
pH 7.0	67.4	24-mer
pH 9.0	55.8	24-mer + Dimer
pH 10.0	51.4	Dimer
pH 11.0	-	Dimer + Denatured
pH 12.0	-	Denatured

Table 3.5 Thermostability and assembly state of wild-type BFR at different pH. The isolated 24-mer BFR is dialyzed to buffers (50 mM phosphate) at different pH adjusted using HCl or

NaOH¹². Monomer* indicates it maybe different from the monomer at pH 3.0 due to shape changes in the thermal unfolding curve.

3.4.3 Thermodynamic studies of BFR and its mutants R30A, Y114A and E128A by DSC

As it is known that dimers are the building blocks of the ferritin molecules, it is important to study the protein folding/unfolding process of 24-mer and dimer separately using differential scanning calorimetry, which is unable to detect by circular dichroism approach. The information obtained from CD experiments together with the DSC approach aid our understanding on the assembly/disassembly of the oligomeric protein. This calorimetric study is in progress.

3.4.3.1 The effect of different concentrations of dimer on the DSC thermogram

The thermal stability of BFR dimer was studied under different concentrations to determine the optimal concentration for the DSC experiments. After the dimer was isolated by SEC, the protein was diluted to three different concentrations in 50 mM phosphate buffer at pH 7.2. Each concentration of dimer was subjected to further SEC to confirm the oligomerization state (the dimer elutes at 16.3 mL) (Figure 3.12 B). The calorimetric profile of BFR dimer shows one peak at ~ 55 °C and a broad peak at higher temperature at 8.7 μM and 12.7 μM. However, the first peak is not observed for the lowest concentration (5.6 μM), more background noise is observed, and a possible new transition is evident at 73 °C (Figure 3.12 A). The temperature range of the thermal unfolding process investigated by DSC is similar to that of a temperature dependant CD experiment in that the protein starts to unfold at ~ 45 °C and the transition is complete at ~ 95 °C. Interestingly, an intermediate state for the higher concentrations of dimer is perhaps observable in the CD experiment which may be due to the first transition (55 °C) observable in the DSC. However, this intermediate state will mostly likely not affect the overall melting temperature of the

protein measured by CD because the 50% unfolding point is the same for these different concentrations of dimer (Figure 3.12 C). The curve fitting of the first peak in DSC varies, which is probably due to the difficult to define pre-transition baseline. Because the CD and the DSC agree on the 85 °C transition, this is probably due to unfolding of the protein. The lower temperature transition is most likely due to some other effect and it is concentration dependent, only observable at higher concentrations suggesting that is due to some sort of oligomerization or aggregation.

Figure 3.12 The effect of different concentrations of dimer wild type on the DSC thermogram. The dimer samples are from the batch of SEC isolation. (A) The thermograms of different concentrations of dimer at a scan rate 90 °C/h. (B) SEC to confirm the oligomerization state of the dimer samples before DSC experiment. (C) Thermal unfolding monitored by CD at 222 nm and the scan rate is 90 °C/h. The UV absorbance is normalized by concentration. The fitting results of DSC thermograms are listed in the table.

3.4.3.2 Unfolding of the impure BFR dimer is scan rate dependent

An initial experiment at similar concentrations demonstrated that the thermal unfolding of isolated BFR dimers under different scan rates showed similar thermograms, indicating that the isolated dimer unfolding process is reversible as it is scan rate independent²⁶⁻²⁸ (Figure 3.13 A). However, the dimer from another batch of dimer isolation showed that the first transition was scan rate dependent (Figure 3.13 B), indicating that the first peak will certainly affect the fitting result of the second

thermal transition. Also this indicated that the transition at $\sim 65^\circ\text{C}$ is batch dependent further supporting that it maybe either due to an intermediate oligomerization state or aggregation.

Figure 3.13 Thermograms of isolated wild type dimer at different scan rates. (A) From one batch of isolated dimer at 5.0 and 6.9 μM (B) From a second batch at 8.0 μM . The fitting results are listed in the table.

3.4.3.3 The effect of different concentrations and scan rate on DSC thermal unfolding of isolated BFR 24-mer

The thermal unfolding of the isolated BFR 24-mer was examined at different concentrations (Figure 3.14). The integration under the peaks - the calorimetric enthalpy (ΔH_{cal}) - is the same though the melting temperature for each unfolding transition is slightly altered which is probably due to differences in the definition of

the pre-transition baseline. The concentration at $\sim 8.0 \mu\text{M}$ was used to determine the effects of different scan rate on protein unfolding (Figure 3.14 right). It was observed that the ΔH_{cal} is the same ($\sim 140 \text{ kcal/mol}$) and the transition for the 24-mer unfolding process is scan rate independent, indicating that the unfolding process is reversible

27,29.

Figure 3.14 The effect of different concentrations and scan rate on the thermal unfolding of isolated BFR 24-mer monitored by DSC. *(Left)* DSC profiles of isolated BFR 24-mer at different concentrations at a scan rate of $60 \text{ }^{\circ}\text{C/h}$. *(Right)* DSC profiles of isolated BFR 24-mer at scan rates of $60 \text{ }^{\circ}\text{C/h}$ and $90 \text{ }^{\circ}\text{C/h}$ with similar concentrations, $8.4 \mu\text{M}$ and $7.3 \mu\text{M}$ respectively.

3.4.3.4 Comparison of BFR 24-mer and dimer

BFR 24-mer and dimer isoalted by SEC before performing the DSC analysis. These isolations share the same transitions with the exception of the one observed in the 24-mer solution at $52 \text{ }^{\circ}\text{C}$ (Figure 3.15) suggesting that the disassembly of 24-mer to dimer occurs at this temperature. The transition occurring at the range of $60 \text{ }^{\circ}\text{C} \sim 100 \text{ }^{\circ}\text{C}$ is proposed to be the unfolding process of dimer to partially unfolded dimer or monomer, and eventually to the completely denatured protein. The fact that this is similar to the T_m that is observed with CD which monitors secondary structure supports this hypothesis. The calorimetric enthalpy (ΔH_{cal}) for this transition in the isolated 24-mer and dimer samples is 142.3 kcal/mol and 102.5 kcal/mol respectively.

It is not unreasonable that the excess heat uptake (~ 40 kcal/mol) is the energy for the dissociation of 24-mer to dimer. The calorimetric enthalpy measured is close to the previous reported ~ 440.9 kJ/mol (150.3 kcal/mol) for the L-ferritin subunit at pH 2.8 that the protein existed as dimer state ¹⁵.

It should be mentioned that the isolated dimer samples used in previous scan rate and concentration determination experiments (section 3.3.3.1 and Figure 3.13 B in section 3.3.3.2) are different from the isolated dimer samples used in this experiment from the evident absence of the peak at ~ 60 °C (Figure 3.12 and 3.13 B). The peak observed at 60 °C may be a “misfolded dimer” which is clearly scan rate dependant (Figure 3.13 B). The thermal unfolding profile of the misfolded dimer (with the extra transition at 60 °C) is different from that of the folded dimer (Figure 3.13 A and Figure 3.15 dimer) and it is not immediately observable in the SEC (Figure 3.12 B). This could also explain why there is an intermediate state in the temperature dependent CD experiment (Figure 3.12 C) which had not been previously observed and its transition temperature $50 \sim 70$ °C is correlated to the transition obtained from DSC experiment (Figure 3.12 A). Looking at this effect in another way, this data taken together demonstrates that DSC can therefore be used to detect whether there is misfolded dimer in the protein sample, a state that is undetectable using SEC. The unfolding of this misfolded dimer is evidently different from the dissociation of 24-mer to dimer (52 °C) and it is also scan rate dependent (Figure 3.13 B). Being aware of this transition can allow the comparison of the other states.

The thermal denaturation of dimer can also be deconvoluted into two transition peaks using peak analysis, 77°C and 85°C (Figure 3.15). However, it seems to be an oversimplification to equate these two transitions to the consecutive disassembly of dimer to monomer and the unfolding of the monomer itself. Considering the 50%

depression in secondary structure is at $\sim 75^{\circ}\text{C}$ in the thermal denaturation CD experiment (Figure 3.12 C), the first transition at 77°C in DSC thermal profile could not simply be the dimer disassociation. At the second transition (85°C), the partially unfolded dimer/monomer denatures completely.

Figure 3.15 DSC thermal profiles of isolated 24-mer and dimer wild type at the same concentration. The scan rate is $90^{\circ}\text{C}/\text{h}$. A non-two-state model is used to fit the thermograms. ΔH and ΔH_v represent for calorimetric enthalpy and van't Hoff enthalpy respectively. The unit is cal/mol. The curve fitting results are listed in the table.

3.4.3.5 Rescan of isolated BFR dimer and 24-mer results in similar thermal profiles

Repeated scanning of isolated BFR 24-mer and dimer was carried out to determine whether protein folding is reversible. Thermodynamic measurements revealed that the thermal unfolding for both dimer and 24-mer have little reversibility. After a full-

scale calorimetric recording from 30 to 110 °C, the heat absorption peak at 59 °C for isolated 24-mer was not observed in the rescan; it shifted to a broader peak at a higher temperature. There was a small fraction of misfolded dimer in the sample because a peak at 65 °C was observed and the misfolded protein aggregated strongly in the rescan. The rescan experiment of the isolated dimer resulted in a thermogram that strongly resembled the rescanned thermogram from the isolated 24-mer further supporting that the 65 °C transition is the result of a misfolded protein.

This analysis is consistent with a Lumry and Eyring model which places the native state in equilibrium with the unfolded state which can further denature irreversibly ³⁰:

where N is the native state, U is the reversibly unfolded intermediate, and D is the irreversibly denatured form. Unfolding probably leads to the exposure of a hydrophobic surface which results in an aggregated denatured state. Therefore, the unfolded form U either refolds to the native form N or it can be irreversibly denatured to the final aggregation form D. In many cases, protein aggregation occurs near the transition temperature resulting in no or little reversibility ^{31,32}.

	T_{m1} (°C)	T_{m2} (°C)	T_{m3} (°C)
24-mer	59	79	87

24-mer rescan	-	70	87
Dimer	65	82	87
Dimer rescan	-	67	85

Figure 3.16 Reversibility of DSC thermal unfolding profiles for isolated BFR 24-mer and dimer. The scan rate is 90 °C/h. The concentrations for 24-mer and dimer are 8.6 μM and 8.1 μM respectively. The non-two-state model is used for fitting. The curve fitting results are listed in the table. Each rescan is carried out continuously after cooling from the first scan. The table is the result of the peak analysis in the bottom four thermograms.

3.4.3.6 Comparison of DSC thermograms for mutants R30A, E128A, Y114A with the isolated dimer of wild-type BFR

The thermal unfolding of the isolated dimer of wild-type BFR and its mutants that solely exist in the dimer oligomerization state in solution (see Chapter 2) were analyzed and compared by DSC. The thermal transition temperatures are similar to those observed using CD in that the mutant R30A exhibits a higher thermal transition

than the wild type and E128A shows a depression in the melting temperature (Figure 3.17 and Figure 2.21 in Chapter 2). However, the mutant Y114A exhibits a similar transition temperature to that of wild type. It should be noted that it is not completely expected that the absolute melting temperatures obtained from the CD experiments will be the same as those measured from DSC. The T_m obtained from CD is defined as the temperature the α -helical structure of the protein is 50% unfolded. These DSC thermal diagrams also demonstrate that dimer is the only oligomerization state for these mutants because there is no observation of a 24-mer transition peak (52°C) further supporting SEC, native gel and DSC experiments (Figure 3.17 and Figure 2.19 in Chapter 2).

The ratio of $\Delta H_{vH} : \Delta H_{cal}$ which is usually used to indicate the number of subunits involved in the cooperative unfolding unit didn't give valuable information that some gave ~ 0.5 but some gave ~ 1.0 either in the first or second transition for all different types of dimers (Figure 3.17).

Protein	T_{m1} (°C)	ΔH_{cal1} kcal/mol	ΔH_{vH1} kcal/mol	$\Delta H_{cal1} / \Delta H_{vH1}$	T_{m2} (°C)	ΔH_{cal2} kcal/mol	ΔH_{vH2} kcal/mol	$\Delta H_{cal2} / \Delta H_{vH2}$
Wild-type dimer	77	43.1	65.7	0.66	86	62.0	100.0	0.62
R30A	86	98.9	75.5	0.76	92	54.4	130.0	0.42
Y114A	80	65.6	62.3	1.05	86	58.0	82.6	0.70
E128A	70	18.6	47.9	0.39	80	75.1	75.4	0.99

Figure 3.17 DSC thermograms for isolated BFR dimer and mutations that form only dimer in solution. The concentration for each protein is at the range of 4.0 ~ 6.0 μ M. The scan rate is 90 $^{\circ}$ C/h. The curve fitting results are listed in the table.

3.5 Conclusion

The interfacial amino acids at the two-fold axis that are involved in salt bridging interactions were selected based on the prediction of the energetic importance of their interactions. All of these four mutants formed cooperatively folded α -helical secondary structure and only BFR E60H possessed the ability to form nano-cage structure in solution as evidenced by SEC and native PAGE. The other three mutants, R30D, D56R, D56RE60H, only formed lower order oligomers in solution which is surprising as these interactions were replaced with repulsing residues. Furthermore, corresponding to computational calculations, R30D resulted in the largest reduction in melting temperature. This study further confirmed that these selected amino acid residues play an important role in the stability and self-assembly of the nano-cage protein.

Wild-type BFR is highly stable in acidic conditions, and changes to the oligomerization state were only observed at pH 3.0. Protonation of functional groups at low pH destroys the electrostatic interactions, causing complete dissociation from 24-mer into monomer. Wild-type BFR gradually denatures with an increase in pH, leading to substantial loss of the secondary structure. It disassembles at pH 9.0 and is completely unfolded at pH 12.0.

Studies using differential scanning calorimetry showed that the unfolding of 24-mer and dimer are scan-rate independent in the absence of denatured impurities. The unfolding process for 24-mer and dimer starts from 40 $^{\circ}$ C and 65 $^{\circ}$ C respectively and both completely unfold at 100 $^{\circ}$ C. Wild-type BFR possibly disassembles according to the following mechanism that the 24-mer disassociates to dimer; the dimer unfolds to

partially unfolded dimer/monomer state and eventually unfolds to the completely denatured protein. The reactions of dimer disassembly and unfolding of the subunits (dimer/monomer) themselves occur simultaneously. Further work is required to fully explore deeply the multiple nuances of this technique in particular and this class of proteins in general.

3.6 References

1. Cooper A, Nutley MA, Wadood A. Differential scanning microcalorimetry: in S. E. Harding and B. Z. Chowdhry (Eds.).
2. Ladbury JE, Doyle ML. Biocalorimetry 2: Applications of Calorimetry in the Biological Sciences: John Wiley & Sons Ltd; 2004.
3. Kemp RB. Handbook of Thermal Analysis and Calorimetry : From Macromolecules to Man (Handbook of Thermal Analysis and Calorimetry): Elsevier Science; 1999.
4. Jackson WM, Brandts JF. Thermodynamics of Protein Denaturation - a Calorimetric Study of Reversible Denaturation of Chymotrypsinogen and Conclusions Regarding Accuracy of 2-State Approximation. *Biochemistry* 1970;9(11):2294-&.
5. Sturteva.Jm. Some Applications of Calorimetry in Biochemistry and Biology. *Annu Rev Biophys Bio* 1974;3:35-51.
6. Sturtevant JM. Biochemical Applications of Differential Scanning Calorimetry. *Annu Rev Phys Chem* 1987;38:463-488.
7. Privalov PL, Khechina.Nn. Thermodynamic Approach to Problem of Stabilization of Globular Protein Structure - Calorimetric Study. *Journal of Molecular Biology* 1974;86(3):665-684.
8. Hu CQ, Sturtevant JM, Thomson JA, Erickson RE, Pace CN. Thermodynamics of Ribonuclease-T1 Denaturation. *Biochemistry* 1992;31(20):4876-4882.
9. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *Journal of Molecular Biology* 2002;320(2):369-387.
10. Stefanini S, Vecchini P, Chiancone E. On the Mechanism of Horse Spleen Apoferritin Assembly - a Sedimentation-Velocity and Circular-Dichroism Study. *Biochemistry* 1987;26(7):1831-1837.
11. Crichton RR, Bryce CFA. Subunit Interactions in Horse Spleen Apoferritin - Dissociation by Extremes of Ph. *Biochem J* 1973;133(2):289-299.
12. Kilic MA, Spiro S, Moore GR. Stability of a 24-meric homopolymer: Comparative studies of assembly-defective mutants of *Rhodobacter capsulatus* bacterioferritin and the native protein. *Protein Sci* 2003;12(8):1663-1674.
13. Jin KS, Kim M, Rho Y, Ahn B, Jung S, Kim H, Ree M. pH-Dependent Structures of Ferritin and Apoferritin in Solution: Disassembly and Reassembly. *Biomacromolecules* 2011;12(5):1629-1640.
14. Yoshizawa K, Mishima Y, Park SY, Hedde JG, Tame JRH, Iwahori K, Kobayashi M, Yamashita I. Effect of N-terminal residues on the structural

- stability of recombinant horse L-chain apoferritin in an acidic environment. *J Biochem* 2007;142(6):707-713.
15. Martsev SP, Vlasov AP, Arosio P. Distinct stability of recombinant L and H subunits of human ferritin: calorimetric and ANS binding studies. *Protein Eng* 1998;11(5):377-381.
 16. Stefanini S, Cavallo S, Wang CQ, Tataseo P, Vecchini P, Giartosio A, Chiancone E. Thermal stability of horse spleen apoferritin and human recombinant H apoferritin. *Arch Biochem Biophys* 1996;325(1):58-64.
 17. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 2002;320(2):369-387.
 18. Littlechild JA, Willies SC, Isupov MN, Garman EF. The binding of haem and zinc in the 1.9 angstrom X-ray structure of Escherichia coli bacterioferritin. *J Biol Inorg Chem* 2009;14(2):201-207.
 19. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to Measure and Predict the Molar Absorption-Coefficient of a Protein. *Protein Sci* 1995;4(11):2411-2423.
 20. Grant RA, Filman DJ, Finkel SE, Kolter R, Hogle JM. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* 1998;5(4):294-303.
 21. Zhang Y, Raudah S, Teo H, Teo GWS, Fan RL, Sun XM, Orner BP. Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* 2010;285(16):12078-12086.
 22. Fan R, Boyle AL, Cheong VV, Ng SL, Orner BP. A Helix Swapping Study of Two Protein Cages. *Biochemistry* 2009;48(24):5623-5630.
 23. Ardejani MS, Li NX, Orner BP. Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* 2011;50(19):4029-4037.
 24. Luedtke NW, Dexter RJ, Fried DB, Schepartz A. Surveying polypeptide and protein domain conformation and association with FIAsh and ReAsH. *Nat Chem Biol* 2007;3(12):779-784.
 25. Santambrogio P, Levi S, Arosio P, Palagi L, Vecchio G, Lawson DM, Yewdall SJ, Artymiuk PJ, Harrison PM, Jappelli R, Cesareni G. Evidence That a Salt Bridge in the Light Chain Contributes to the Physical Stability Difference between Heavy and Light Human Ferritins. *Journal of Biological Chemistry* 1992;267(20):14077-14083.
 26. Sanchezruiz JM, Lopezlacomba JL, Cortijo M, Mateo PL. Differential Scanning Calorimetry of the Irreversible Thermal-Denaturation of Thermolysin. *Biochemistry* 1988;27(5):1648-1652.
 27. Lepock JR, Ritchie KP, Kolios MC, Rodahl AM, Heinz KA, Kruuv J. Influence of Transition Rates and Scan Rate on Kinetic Simulations of Differential Scanning Calorimetry Profiles of Reversible and Irreversible Protein Denaturation. *Biochemistry* 1992;31(50):12706-12712.
 28. Carey PR, Davoodi J, Wakarchuk WW, Surewicz WK. Scan-rate dependence in protein calorimetry: The reversible transitions of *Bacillus circulans* xylanase and a disulfide-bridge mutant. *Protein Sci* 1998;7(7):1538-1544.
 29. Davoodi J, Wakarchuk WW, Surewicz WK, Carey PR. Scan-rate dependence in protein calorimetry: The reversible transitions of *Bacillus circulans* xylanase and a disulfide-bridge mutant. *Protein Sci* 1998;7(7):1538-1544.

30. Lumry R, Eyring H. Conformation Changes of Proteins. *J Phys Chem-US* 1954;58(2):110-120.
31. Manly SP, Matthews KS, Sturtevant JM. Thermal denaturation of the core protein of lac repressor. *Biochemistry* 1985;24(15):3842-3846.
32. Brandts JF, Hu CQ, Lin LN, Mas MT. A Simple-Model for Proteins with Interacting Domains - Applications to Scanning Calorimetry Data. *Biochemistry* 1989;28(21):8588-8596.

CHAPTER 4 Alanine shaving of mini-ferritin, *E. coli* DPS test*

4.1 Introduction

The first Dps protein family member was isolated from *Escherichia coli* grown in starvation conditions by Roberto Kolter in 1992. They found that the protein forms an extremely stable complex with DNA in vitro without apparent sequence specificity. Therefore it was named DNA-binding protein from starved cells (Dps)¹. These proteins remove cellular Fe (II) and H₂O₂ which are potentially toxic in aerobic conditions as “Fe (II) forms spontaneously insoluble Fe (III) and reacts with molecular oxygen or its reduced forms to yield the highly [DNA] damaging hydroxyl radicals” through the so-called Fenton reaction: $\text{Fe (II)} + \text{H}_2\text{O}_2 \rightarrow \text{Fe (III)} + \text{OH}^- + \text{OH}^\cdot$ ². All Dps proteins possess this intrinsic reactivity by using hydrogen peroxide to oxidize Fe (II) to neutralize the toxicity. Apart from this unique DNA protection capability, there are several members of the family that protect DNA physically by shielding it in Dps-DNA complexes. The crystal structure of *E. coli* Dps was solved by R. A. Grant *et al.* This protein forms a dodecamer with 23 (tetrahedral) point group symmetry. The protein has a hollow core where Fe ions are stored³ demonstrating that Dps is a structural analogue of ferritin³.

4.2 Structure features of Dps proteins

Ferritins, whose structure and function are highly evolutionally conserved, are mainly composed of twenty-four subunits assembled into essentially spherical proteins with a central cavity⁴. Dps proteins (DNA-binding proteins from starved cells), which form oligomers of twelve subunits, protect bacteria primarily from oxidative damage. Generally there is only one Dps protein that is encoded in bacterial

* Parts of this chapter have been published previously with the author of this thesis as the first author. *Protein Sci.* 2011

genomes however there are a few organisms that express two or three of these proteins⁵.

4.2.1 Tetrahedral symmetry of Dps and the structural comparison between Bfr and Dps

The Dps monomer has a four-helix bundle core that is essentially the same core as the bacterioferritin (Bfr) monomer (Figure 4.1). In both structures, the first two helices (the A and B helices) are connected by a short loop, as are the third and the fourth helices (the C and D helices). A long loop stretches across the surface of the four-helix bundle and connects the two pairs of helices. In Bfr, there is no secondary structure in this BC loop, but there is a small E helix at the C terminus. The E helices in the Bfr 24-mer affords contacts between the monomers at the four-fold symmetry axes. The Dps monomer has no E helix, but there is a short helix in the middle of BC loop. This small helix provides additional contacts between the monomers at the two-fold axis. In addition, the terminal ends of the A and D helices of Bfr are a bit longer than the relative helices of Dps. Along with the presence of the E helix, this makes the Bfr monomer a little longer than Dps at the C-terminal end of the bundle (Figure 4.2 and 4.3).

Similar to the packing of Bfr monomers, Dps monomers assemble around the two-fold and the three-fold symmetry axes (Figure 4.1 (b-d)). The major difference is the result of the lack of the E helix, which makes four-fold interactions in Bfr. Dps therefore forms a smaller, lower symmetry oligomer. In contrast to the 432 (O_h) point group symmetry of Bfr, the Dps dodecamer has 32 (tetrahedral, D_{3h}) point group symmetry. The Dps dodecamer measures ~ 9 nm in diameter and has a central cavity of ~ 4.5 nm in which an iron core of ~ 500 Fe^{3+} iron per molecule can be formed³.

There are two major differences in the architectures of Dps and Bfr. First, there are no four-fold interactions in the Dps dodecamer (see above). Second, Bfr's 432 symmetry determines that only one kind of trimeric interaction is possible as the chemical environment around the three-fold axes is the same on the "front" and "back" sides in octahedral symmetry³. In contrast, due to the 32 symmetry of the Dps dodecamer, there are two nonequivalent three-fold axes, one corresponding to the interactions between monomers on the "front" of the dodecamer, and the other is on the "back" side around the same three-fold axis³. One is formed by the N-terminal ends of the subunits which is called "ferritin -like" since its packing is similar to the packing in ferritin (Figure 4.1 (c)). In both protein families iron utilizes the electrostatic potential generated by the negatively charged residues lining the pores at the N-terminal three-fold axes to enter and exit the protein inner cavity^{6,7} (Figure 4.1 (e)). The other type of three-fold symmetry axis in Dps is formed by the C-terminal ends of the monomers. This axis is unique to this protein family and is thus called "Dps-type" (Figure 4.1 (d)). The pore formed at this axis and is smaller, due to hydrophobic constriction, and less acidic³ (Figure 4.1 (f)).

Figure 4. 1 (a) Ribbon diagram of the *E. coli* DPS monomer. Helices A to D are shown in rainbow colors. DPS protein cages in views along the (b) two-fold axis (c) ferritin-like three-fold axis and (d) DPS-like three-fold axis. (e) The expansion shows the aspartic acid residues lining the ferritin-like pore. (f) The expansion shows the hydrophobic amino acids lining the DPS-like pore. Note: in e and f, due to symmetry, the residues are labeled on only one monomer (PDB ID: 1dps, figures are generated using Chimera)

```

DSSP .....LLLLHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHh11LHHHHHHHHHH
BFR .....MKGDTKVINYLNKLLGNELVAINQYFLHARMFkwnwgLKRLNDVEYH 46
ident                ||
DPS skatnlllytrndvsDSEKKATVELLNRQVIQFIDLSLITKQAHWNMrganFIAVHEMLDG 60
DSSP l1111111111111111HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHLL1111HHHHHHHHHH

DSSP HHHHHHHHHHHHHHHHHHHHHHLLLLLLLLL.....LLLLLLLLL...LLHHHHHHHHHHHHHHHH
BFR ESIDEMKHADRYIERILFLEGLPNLQ.....DLGKLNIG...EDVEEMLRSDLALELDG 97
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DPS FRTALIDHLDTMAERAVQLGVALGTtqvinsKTPLKSYPl diHNVQDHLKELADRYAIV 120
DSSP HHHHHHHHHHHHHHHHHHHHHHLLLLLLLLLhhhhhhhLLLLLLLLL1111LHHHHHHHHHHHHHHHH

DSSP HHHHHHHHHHHHHh11LHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHh11111
BFR AKNLREAIGYADsvhDYVSRDMMIEILRDEEGHIDWLLETELDliqkmgqlqnylqaqiree 157
ident | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DPS ANDVRKAIGEAK...DDDTADILTAASRDLDKFLWFIECNIE..... 159
DSSP HHHHHHHHHHLL...LHHHHHHHHHHHHHHHHHHHHHHHHHHHHHLL.....

```

Figure 4.2 Structure alignment of the amino acid sequences of BFR and DPS. The grey shadings indicate the alpha helix regions and the red color residues highlight shared identity. This alignment was generated using the DaliLite online program: <http://www.ebi.ac.uk/Tools/dalilite/index.html>. Lower case means there is no alignment at the specific position relative to the reference structure. “DSSP” stands for dictionary of second structure pattern, “H” stands for helix, “L” stands for loop.

Figure 4. 3 Superimposition of BFR (grey) and DPS (dark) monomers generated using the DaliLite online program⁸: <http://www.ebi.ac.uk/Tools/dalilite/index.html>. The RMSD of the structural superimposition between BFR and DPS is 2.1 Å.

4.2.2 The ferroxidase center

The ferroxidase center is highly conserved in the Dps family and is located at the two-fold symmetry axis and is made up of residues from two monomers rather than being embedded in the four-helix bundle of a single subunit as it is in the ferritins.

In ferritins, the ferroxidase centers are typically bimetallic. Two iron atoms are connected by an oxo-bridge with a distance of close to 3 Å⁹. The iron coordinates to histidine and carboxylates (Asp or Glu) at site A and binds to only acidic residues and with lower affinity in binding site B⁹. In *E. coli* Dps, one subunit provides two histidine residues (H51 and H63, *E. coli* numbering) and the other subunit donates two acidic ligands (D78 and E82, *E. coli* numbering). The two iron atoms also bind two water molecules (Figure 4.4).

Generally, ferritins use O₂ as iron oxidant with the production of hydrogen peroxide¹⁰, whereas bacterioferritins can employ both O₂ and H₂O₂¹¹. Dps proteins typically prefer H₂O₂ in carrying out the iron oxidation with efficiency 100-fold higher than with O₂¹²⁻¹⁴. During the Dps ferroxidation process, iron (II) ions first coordinate with the ferroxidase centers and are rapidly oxidized by hydrogen peroxide. Then they are mineralized in the negatively charged inner cavity. The overall process of iron deposition and detoxification by Dps is described by the following equations (e.q. 4.1 - 4.3)¹⁵:

For ferroxidation:

For mineralization:

Figure 4.4 The two-fold symmetry interface of *E. coli* Dps (**left**) and an expansion of the iron binding residues at the ferroxidase center with two water molecules (orange red) occupying the iron binding sites (**right**). (PDB ID: 1dps, figures were generated using Chimera)

4.2.3 DNA binding

It is known that Dps can interact with DNA and protect DNA from oxidative damage by sequestering iron. The protein acts as a physical shield and therefore prevents Fenton radical chemistry from causing DNA degradation^{13,16}. It has been found that DNA binding is not sequence specific and no DNA binding motif has been identified. However, the mechanism of DNA interaction is not completely established.

The *E. coli* Dps binds DNA via the highly flexible and positively charged N-terminal region of each subunit. The N-terminal regions in the Dps dodecamer extend away from the compact sphere and therefore are able to interact with the negatively charged DNA backbone^{3,17}. It is proposed that DNA binds through the three lysines residues at positions 5, 8, 10 of the N-terminal regions which are unstructured in the PDB structure¹⁷. This model is consistent with the fact that residues 11 to 22 of the N-terminal loop are visible and clearly point away from the cage and towards the solvent (Figure 4.5). It is also reported that deletion of these lysines results in the loss

of DNA binding¹⁷. There are many Dps proteins that do not bind DNA because the N-terminal region either lacks positively charged amino acids or either is very short, as in *L. innocua* Dps¹⁸, or lacks flexibility as in *Agrobacterium tumefaciens* Dps.¹⁹ In these cases other strategies are adopted to protect against oxidative cleavage.

Figure 4.5 View of the *E. coli* Dps emphasizing the N-terminal extensions. The N-terminal extensions are colored in blue and the four-helix bundles are colored in grey.

4.3 Aims and objectives

DNA-binding protein from starved cells (DPS) and Bacterioferritin (BFR) both belong to the ferritin superfamily of proteins. However, the BFR monomer is a four helix bundle which oligomerizes into a “maxi-ferritin” such that twenty-four ferritin monomers associate to form a large, roughly spherical particle with 432 (octahedral) symmetry. In contrast, the DPS monomers, which have essentially the same protein fold (four helix bundle), pack into a “mini-ferritin”, a dodecamer which has 32 (tetrahedral) symmetry. It is of great value to understand the protein-protein interaction and self-assembly of this high-ordered symmetrical protein with the intention of applying what we learn here to the self-assembly of nano-quaternary structure in general. It is very useful to scrutinize the importance of various amino acids at each of the protein-protein interfaces defining different symmetry axes to

understand how these two structurally analogous proteins assemble into nano-structures with different sizes and shapes.

This study, following a similar methodology to that performed with BFR, focuses on mutating single amino acids to alanine around the symmetry axes of DPS and studying how the stability and structure of the tetrahedral DPS is affected by the mutation. In total, eight amino acids were mutated and ten proteins (including wild type and a double mutant to explore follow up on trends observed with the most interesting single mutants) were expressed during the course of this project (Figure 4.12). The amino acids selected for mutations are charged (Aspartic acid (D) and Arginine (R)) components of salt bridges and aromatic amino acids (Tryptophan (W), Tyrosine (Y)) which have been suggested by Bogan *et al.*²⁰ to play major roles in hot spots.

4.4 Materials and methods

4.4.1 Virtual alanine scanning of DPS

A rapid and computationally non-intensive approach, developed by the Baker lab²¹, to predict the energetically important amino acid residues at protein-protein interfaces is available at the following address: <http://robetta.bakerlab.org/alaninescan>.

The algorithm of this computational virtual alanine scanning identifies all the interfacial residues in a protein-protein interface automatically. In this method, an interface residue is based on defined by the following criteria: (i) in the protein complex, at least one atom from the residue is within 4 Å of an atom in the residue from the other protein, or (ii) the number of beta carbon atoms surrounding the beta carbon atom of the residue in question increases and therefore this residue is one that is buried upon formation of the protein-protein interaction although it may not be making a direct contact. The program replaces each of the interface residues

individually with alanine and the free binding energy is then calculated derived from the effect this mutation has on the protein complex.

The input for this analysis in Dps was modified from the crystal structure coordinates (PDB ID: 1dps). Two monomers on the two-fold axis, and three monomers around the three-fold axis were used. The output of the computational analysis predicted a “hot spot” if the $\Delta\Delta G$ for mutagenesis to alanine was greater than 1 kcal/mol.

4.4.2 Gene synthesis of DPS and its mutants

The protein sequence of DNA binding protein from starved cells *E. coli* (PDB ID: 1dps) was obtained from the Protein Data Bank (PDB). The protein sequence of wild type and mutants were converted to gene sequences by utilizing optimal codons for maximizing the expression level in *E. coli*. Codons and the primer design are similar as it is described for BFR and the gene preparation and cloning procedure are similar too (see Chapter 2). The mutant constructs for DPS R18A, DPS R83A and DPS R83A/R133A were generated using the QuikChange site-directed mutagenesis kit (Stratagene) with the DPS construct as the template. The sequences of all the oligonucleotides are listed in Appendix 3.

4.4.3 Protein expression and purification

Plasmids harboring the desired genes were transformed into *E. coli* BL21(DE3) (Novagen) cells. A starter culture was incubated overnight with shaking (150 rpm, 37 °C) in LB (5 ml) containing carbenicillin (Novagen, 50 ug/ml). This culture was used to inoculate pre-warmed LB (500 ml) and further incubated with shaking (150 rpm, 37 °C, ~ 3 hrs) until OD₆₀₀ was 0.4 - 0.6. IPTG (Fermentas, 2 ml, 0.1 M) was then added to induce protein expression and the culture was further incubated (200 rpm, 3h, 30 °C). The cells were harvested by centrifugation (10000 rpm, 20 min, 4 °C) and

the pellet was resuspended in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0). “Complete, EDTA-free” protease inhibitor solution (Roche, 0.6 ml, 25x stock solution) was added to inhibit protease activity. The suspension was lysed by sonication (SONICS Vibra Cell sonicator, cycles of 3 s pulse at 40 % amplitude for two 5 mins periods) at 4 °C. The lysis debris was centrifuged (12,000 rpm, 15 min, 4 °C) and the supernatant was filtered.

The clarified lysate was then added to Ni-NTA (nickel-nitrilotriacetic acid) resin (Qiagen) that was pre-equilibrated with lysis buffer. The mixture was incubated by gentle inversion for 1 hour at 4 °C. The suspension was centrifuged (700 g, 5 min) to separate the resin from solution. The resin was washed twice with wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH 8.0), followed by affinity cleavage using on-resin or off-resin digestion methods (Figure 4.6).

Figure 4.6 Schematic diagram of on-resin and off-resin digestion methods using affinity Ni-NTA resin.

For on-resin digestion, the resin was pre-equilibrated with enterokinase buffer (20 mM Tris-HCl, 50 mM NaCl, 2 mM CaCl₂, pH 8.0) before enterokinase (New England Biolabs, 6 μ l, 2 μ g/ml) was added. The cleavage reaction was carried out either at 4 °C for 36 hrs or at room temperature for 16 hrs. For the off-resin digestion, the fusion protein was eluted from the resin by washing with elution buffer (50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0). The elutant was dialyzed against enterokinase buffer before enterokinase was added. The digestion of the fusion protein was performed at 4 °C for 36 hrs. Additional pre-equilibrated Ni-NTA resin (1.5 ml, 50% slurry) was added to remove the cleaved tags. The target protein was concentrated via ultrafiltration (Sartorius Vivaspin 6) for the further gel filtration purification step. A 12% SDS-PAGE gel was carried out to monitor the samples throughout the whole protein expression and purification process and to check the purity of the target proteins. The identity of the purified proteins was confirmed by MALDI-TOF mass spectroscopy.

4.4.4 Transmission electron microscopy (TEM)

A drop of a solution of purified protein (5 μ l, $\sim$ 20 μ g ml⁻¹) was placed on a piece of parafilm and a grid (Formvar/carbon, copper 300 mesh) was placed on top of the drop (1 min). Excess sample was removed by wicking with a piece of filter paper and the grid was placed on a drop of uranyl acetate (5 μ l, 1% w/v, 1 min). Excess solution was removed by wicking with filter paper. TEM data was obtained using JOEL JEM-1400 microscope operation at 120 keV. Captured TEM micrographs were analyzed using ImageJ (NIH) to measure the longest diameter of the protein cages. For each protein, 100 particles were measured except for DPS Y16A where only 50 particles were measured. Unpaired t-test was run to compare the means of the particle sizes

between the wild type and DPS Y16A using the online service:
<http://www.graphpad.com/quickcalcs/ttest1.cfm>

4.4.5 Size Exclusion chromatography (SEC)

Size exclusion chromatography experiments were performed on an ÄKTAFPLC™ (GE Healthcare) system equipped with a UV detector ($\lambda = 280$ nm) using a Superdex 200 10/300 GL gel filtration column pre-equilibrated with running buffer (50 mM NaH_2PO_4 , 150 mM NaCl, pH 7.0). The SEC experiments were all conducted at 4 °C with a flow rate 0.5 ml/min. The column was previously calibrated using six well-characterized proteins as standards (GE Biosystems Calibration Kit). The elution volumes (V_e) 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 molecular weight of the proteins. Oligomerization state was determined from the calculated molecular weight and the predicted monomer molecular weight.

4.4.6 Temperature-dependent circular dichroism (CD)

CD spectra in the far-UV range (200 nm - 250 nm) were collected on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller using a 1 mm quartz cell. Protein concentration was determined (BCA Protein Assay Kit, Novagen) and the stock solutions were diluted ($100 \mu\text{g ml}^{-1}$) with phosphate buffer (50 mM NaH_2PO_4 , pH 7.2). Ellipticity was measured at 222 nm and the temperature was increased from 20 to 90 °C in 5 °C steps. Samples were allowed to equilibrate for 8 min at each temperature. After melting, the protein solutions were cooled slowly to 20 °C in ten minutes, and the spectra were compared to those obtained before unfolding. At least three replicates were performed.

4.4.7 Native PAGE

A 7% gel was run using 10 µg of wild type DPS and double mutant DPS R83A/R133A. Gel electrophoresis was carried out on a Bio-Rad Laboratories electrode (S1, 80 V, 20 min; S2, 100 V, 40 min). The gel was stained (0.025% commassie brilliant blue, 1 h) and subsequently destained (1 h) before the gel was visualized.

4.5 Results and discussion

In analogy to our work with *E. coli* bacterioferritin (BFR), DPS was subjected to a strategy where amino acids at protein-protein interfaces were mutated to identify residues that control assembly and stability²². As opposed to traditional “alanine shaving” studies where the energetic role of each residue is determined relative to the wild-type through the measurement of the equilibrium between the oligomerized state and disassociated state, our strategy, due to the complexities associated with highly oligomeric structures and mathematical challenges in measuring their equilibrium, focused on measuring the energetic effect of mutating a residue to alanine on the equilibrium between the folded, oligomeric state and the unfolded, disassociated state through denaturation experiments²³. Changes in protein-protein interaction stability should be reflected in changes in overall stability. We reasoned that the energy of the overall stability, as estimated by the melting transition (ΔG_{melt}), could be broken into disassociation (the most significant contribution of which are the protein-protein energetics), monomer unfolding energies, and a solvation term ($\Delta G_{\text{disassoc}}$, ΔG_{unfold} , and ΔG_{solv} respectively):

$$\Delta G_{\text{melt}} \sim \Delta G_{\text{disassoc}} + \Delta G_{\text{unfold}} + \Delta G_{\text{solv}}$$

The challenges with this analysis lie in uncoupling the sometimes cooperative folding and association energies and in addressing the solvation. The unfolding energy can

be further broken into energies associated with inherent conformational preferences which stabilize secondary structure, and residue-residue interactions which mainly stabilize tertiary structure via forming the hydrophobic core of the protein. Comparing mutant and wild-type proteins results in the change in melting energy due to mutation ($\Delta\Delta G_{\text{melt}}$) which is made up of the residual quaternary ($\Delta\Delta G_{\text{prot-prot}}$), tertiary ($\Delta\Delta G_{\text{tertiary}}$), and secondary structure ($\Delta\Delta G_{\text{secondary}}$) components:

$$\Delta\Delta G_{\text{melt}} \sim \Delta\Delta G_{\text{prot-prot}} + \Delta\Delta G_{\text{tertiary}} + \Delta\Delta G_{\text{secondary}} + \Delta\Delta G_{\text{solv}}$$

In a well-behaved, and properly designed and selected system, this relationship can be employed to establish a ranking of the importance of various residues at a protein-protein interface. The residues we selected are located at the protein-protein interfaces and are hence positioned on the outside of the folded monomers. Therefore they are playing little if any role in the stabilization of the protein tertiary structure allowing the assumption that $\Delta\Delta G_{\text{tertiary}}$ is small upon mutation. Moreover, because they are buried at the protein-protein interface, solvation changes between the mutant and wild-type are also assumed to be negligible. Furthermore, ferritins are dominated by α -helix secondary structure. Residual helix propensities for each amino acid are well known and would be the major contributor to $\Delta G_{\text{secondary}}$ ²⁴. Thus, a mutation to alanine, due to the fact that this residue has the highest helix propensity, should stabilize the folded state and thus increase the melting temperature. Thus, a decrease in melting temperature will reflect a substantial effect which can be attributed to $\Delta\Delta G_{\text{prot-prot}}$ and the potential identification of a “hot spot”. Therefore, we reason that decreases in the melting temperatures of our DPS mutants, where interfacial residues are mutated to alanine, should indicate that these residues are contributing to the protein-protein interactions that govern self-assembly.

4.5.1 Selection of key interfacial residues for mutagenesis

Although saturation mutagenesis is a powerful strategy for the identification of residues that play key energetic or structural roles at protein-protein interfaces^{20, 25}, the logistics required for this technology are beyond those available to our laboratory. Therefore we opted for a more directed approach. We targeted residues involved in salt bridges, because their identification is relatively straight forward, and aromatic residues due to the fact that they have a high tendency to be found energetic hot spots between proteins^{26, 27}. The protein has three types of protein-protein interfaces that are related to the two-fold and the Dps-like and the ferritin-like three-fold symmetry axes. The Dps-like three-fold interface buries only a small amount of monomer surface area and salt bridging interactions or aromatic amino acids that are prevalent in hot spots (Trp or Tyr) are not evident in this interface. Computational analysis (see below) confirms that it may not be the most fertile interface for our hot spot search (Table 4.1). Therefore the selection of potential hot spot residues was mainly focused on the two-fold and the ferritin-like three-fold interfaces. Through simple inspection of the *E. coli* DPS crystal structure³ (PDB ID: 1DPS) eight residues were immediately identified that appear to be clearly involved in protein-protein interactions and were selected for mutagenesis. These are Y16, R18, W52, D78, R83, R133, D146, and W160 (Figure 4.7).

Figure 4.7 Predicted structurally and energetically important residues at the protein-protein interfaces of the nano-cage protein, *E. coli* DPS. (**Left**) The DPS crystal structure (PDB ID: 1DPS) with mutated residues highlighted with respect to the axes of symmetry and schematized diagrams representing the protein-protein interactions defining these symmetries. (**Right**) The residues of DPS mutated in this study and their schematized position at the symmetry related protein-protein interfaces.

Two of these, W52 and D78, make two-fold symmetric interactions near the ferroxidase center^{3, 7, 28}. It should be noted that the aromatic residue W52 was identified at both the two-fold and the Dps-like three-fold interface by PISA (Protein Interfaces, Surfaces and Assemblies) analysis, however the energetic effect was predicted to be more significant at the two-fold interface (data not shown). The acidic residue D78 appears to be interacting with both K48 and R70 (Figure 4.8A).

Five of the selected residues, R18, R83, R133, D146, and W160, are positioned around the three-fold symmetry axes. Intriguingly, the acidic residue, D146, is located in an ion channel which has been shown to have a negative electrostatic gradient which has been shown to have a negative electrostatic gradient which guides iron toward the ferroxidase centers⁷. Residue R18 appears to form salt bridges with E163 and D123 (Figure 4.8 B). Residues R83 and R133 seem to be part of an extended network involving D143, D156, and D20 (Figure 4.8B). Because these residues seem to be working in concert, it was thought that whatever effect the single mutations might have would be amplified together. Therefore the double mutant R83A/R133A was also chosen for study²⁹. The aromatic residue, W160, is located at both types of three-fold interfaces, but, again, the energetic effect at the ferritin-like three-fold interface is predicted to be more significant than at Dps-like symmetry interface (Table 4.1).

Figure 4.8 Hydrogen bonding networks at the DPS two-fold and three-fold symmetry interfaces. **(A)** Two-fold symmetry interface (left) and expansion of the interaction between D78, K48, and R70 (right). **(B)** Three-fold symmetry interface (left) and blow up of the interaction between five charged residues (R83, R133, D156, D143, D20) and between R18, E163 and D123 (right) The heavy atom distances between the interacting groups are indicated as dotted lines and the distances are given in Å.

To add support to our residue selection, sequence alignment analysis of DPS proteins from thirteen diverse species was performed. This analysis revealed that six of the selected residues are conserved (Figure 4.9). The two that are not found in all the other proteins, Y16 and R18, are located in a N-terminal region that is not present in the majority of the proteins analyzed. However, the only other protein that has this N-terminal region, DPS from *Yersinia pestis*, shares Tyr and Arg residues in identical positions.

Figure 4.9 Sequence alignment of representative Dps proteins. EcDPS, DNA-binding protein Dps (*E. coli*); LiDPS, non-heme iron-containing ferritin (*Listeria. innocua*); HpDPS, neutrophil-activating protein A (*Helicobacter pylori*); BaDPS, gene stress protein 20U (*Bacillus anthracis* str. A0193); AtDPS, DNA protection during starvation protein (*Agrobacterium tumefaciens* str. C58); MsDPS, DNA protection during starvation protein (*Mycobacterium smegmatis*); VcDPS, DPS family protein (*Vibrio cholerae* MZO-3); YpDPS, DNA protection during starvation protein (*Yersinia pestis* bv. Antiqua str. Angola); PhDPS, stress response DNA-binding protein Dps (*Pseudoalteromonas haloplanktis* str. TAC 125); PaDPS, putative DNA-binding protein, DPS (*psychrobacter arcticus* str. DSM 17307); SoDPS, DPS family protein (*Shewanella oneidensis*); TfDPS, hypothetical protein Tfu_0799 (*Thermobifida fusca* YX); RmDPS, Metalloreulation DNA-binding stress protein (*Ralstonia metallidurans* str. CH34). Alignment was created with ClustalW2. Helical secondary structure elements are indicated at the top. The residues at two-fold and three-fold interfaces in EcDPS are highlighted in light gray and dark gray respectively. The residues mutated in this study are underlined.

To further add weight to our selections, virtual alanine scanning³⁰ was also performed to identify energetically important residues (Table 4.1). A calculated $\Delta\Delta G$

of greater than 1 kcal/mol in this analysis suggests that a residue may be an energetic hot spot, and all but one of our selected residues were predicted to be hot spots. The outlier was Y16 which is a solvent exposed tyrosine that the calculation did not identify as an interfacial residue (for formal definition used by this software see Section 4.4.1). Nevertheless, we decided to generate this mutant. Additionally, the D146 was predicted to be only a weak hot spot. Because of the placement of this residue at an important ion channel (see above) this mutant was also produced. Of the selected residues R83 and R133 were predicted to be the energetically most significant.

Residue	Chain	Int_id	$\Delta\Delta G$ (bind)	ΔG (partner)	Interface
15	F	0	0.01	0.32	C3
17	F	1	0.65	1.47	C3
18	F	1	3.93	-0.21	C3
19	F	1	1.15	0.79	C3
20	F	1	0.26	-0.38	C3
21	F	1	0.05	0.54	C3
79	F	1	0.05	0.19	C3
82	F	1	0.02	-0.39	C3
83	F	1	9.84	2.68	C3
85	F	1	0.19	1.07	C3
86	F	1	3.2	0.24	C3
123	F	1	1.03	-1.16	C3
130	F	1	4.21	0.61	C3
133	F	1	6.11	3.17	C3
134	F	1	0.4	-0.47	C3
136	F	1	0.32	1.9	C3
141	F	0	2.06	1.47	C3
142	F	1	-0.17	-0.99	C3
143	F	1	1.47	-0.89	C3
146	F	1	1.48	-1.25	C3
147	F	1	0.19	1.51	C3
149	F	1	0.32	-0.06	C3
152	F	1	0.14	-0.88	C3
153	F	1	2.46	-0.07	C3
156	F	1	6.56	0.85	C3
157	F	1	0.04	0.21	C3
159	F	1	1.02	1.98	C3
160	F	1	2.42	-0.13	C3
163	F	1	3.95	-0.09	C3
164	F	0	0.01	5.32	C3
15	F	1	0.73	0.32	C2
41	F	1	0.62	1.3	C2
42	F	1	0.5	-1.54	C2
45	F	1	1.73	0.8	C2
46	F	0	0.15	2.04	C2
48	F	1	0.6	1.27	C2
49	F	1	2.51	0.05	C2
51	F	0	-0.02	1.14	C2
52	F	1	3.35	0.9	C2
53	F	1	2.25	0.42	C2

55	F	1	0.05	-0.05	C2
70	F	1	0.58	0.41	C2
74	F	1	0.41	0.58	C2
77	F	1	0.28	2.57	C2
78	F	1	1.09	-1.25	C2
85	F	1	0.66	1.07	C2
90	F	1	0.77	0.94	C2
92	F	1	0.66	1.27	C2
94	F	1	0.78	0.12	C2
95	F	1	1.22	-0.05	C2
96	F	1	1.71	-0.84	C2
97	F	0	0.23	1.01	C2
98	F	1	0.44	2.24	C2
99	F	1	0.5	-0.77	C2
102	F	0	0.08	1.45	C2
106	F	1	0.4	-0.22	C2
107	F	0	0.07	3.83	C2
109	F	1	1.15	0.51	C2
112	F	1	0.62	0.83	C2

Tabel 4.1 Results from “virtual alanine scanning” of the protein-protein interface at the DPS three-fold axis. Red indicates residues involved in this study. The website used was: <http://robetta.bakerlab.org/alaninescan> ³¹

The abbreviations used are as follows: “pdb#” is the residue number that was mutated to alanine using the numbering system in the PDB coordinates; “chain” is the pdb chain identifier; “int_id” describes the nature of the interfacial residue. A value of 1 indicates that, in the complex, at least one atom from the residue is within 4 Å of an atom in a residue from the other protein. A value of 0 indicates that the number of beta carbon atoms surrounding the beta carbon atom of the residue in question increases upon complex formation and therefore this residue is one that is buried upon formation of the protein-protein interaction although it may not be making a direct contact; “ $\Delta\Delta G(\text{bind})$ ” is the change in the binding energy between the partners due to the mutation of the residue in question to alanine. A positive value indicates the mutation has a destabilizing affect whereas a residue causing a negative value upon mutation is stabilizing; “ $\Delta G(\text{partner})$ ” is an indication of how a mutation affects the stability of the interacting partner. ²²

4.5.2 Verification of cloning

All the genes of the mutants were successfully synthesized and cloned (Figure 4.10). After cloning, the plasmids that harbored the gene inserts were sequenced. The resulting sequences were aligned with the designed gene sequences. These alignments are listed in Appendix 4. These alignments indicated that all the genes were successfully introduced into the plasmid.

Figure 4.10 Assessment of gene synthesis and cloning by electrophoretic analysis for DPS D146A. (A) Lane 1: 100bp DNA ladder; Lane 2: amplification PCR product; Lane 3: extension PCR using LIC primers. (B) Colony PCR result using LIC primers.

4.5.3 Protein over-expression and purification

SDS-PAGE was used to monitor the samples collected throughout the general protein expression and purification procedure (Figure 4.11). The addition of IPTG leads to the induction of the protein and the thick band with the MW between 25 - 37 kDa in the third lane of the gels for all the constructs suggesting the presence of the desired proteins. After sonication, soluble and insoluble fractions were examined to determine the degree to which the target protein is soluble. From comparing lanes four and five, it was observed that with the exception of DPS W160A and DPS D78A, the majority of the proteins were in the soluble fraction. Even for these two exceptions a substantial amount of protein was found in the soluble fraction. To

purify the soluble recombinant protein, it was bound to Ni-NTA affinity resin (lanes six and seven; faint bands indicate that the protein is being removed from solution) and released through enterokinase cleavage as a protein around 18 kDa (lane eight). The molecular weight of the purified proteins were confirmed by MALDI-TOF mass spectroscopy (Table 4.2).

Figure 4.11 Electrophoretic monitoring of the purification of the DPS-derived proteins. Lane 1: ladder. Lane 2: before induction. Lane 3: post induction. Lane 4: soluble fraction. Lane 5: insoluble fraction. Lane 6: flow through. Lane 7: wash. Lane 8: cleavage.

Protein	MW (calculated)	MW (measured)
DPS	18711	18732.8
DPS Y16A	18619.2	18623.4
DPS R18A	18626.2	18647.4
DPS D78A	18667.3	18653.8
DPS W52A	18596.2	18619.7
DPS R83A	18626.2	18633.6
DPS R133A	18626.2	18638.1
DPS D146A	18667.3	18668.3
DPS W160A	18596.2	18616.2
DPS R83A/R133A	18541.1	18562.0

Table 4.2 MALDI-TOF analysis of expressed proteins.

4.5.4 Transmission electron microscopy

The pure proteins were imaged using TEM to determine whether they are able to assemble into nano-cages. TEM is a microscopic technique which involves the transmission of a high energy electron beam through a very thin sample. The samples were supported by a carbon-coated copper grid and were stained with a negative stain, uranyl acetate, which provides the contrast between ferritins and background resulting in an outline of the ferritin particle. The protein cages appear as light rings surrounding dark cores which is consistent with the literature³². The light rings suggested the ferritin shell and the dark cores are evidence of the uptake of the negative stain in the hollow protein cavities (Figure 4.12).

TEM images indicated that all the proteins, except DPS R83A/133A, formed nano-cages under the experimental conditions (Figure 4.12), suggesting that mutagenesis does alter the energetics to an extent that would disrupt self-assembly.

Figure 4.12 All single mutants can form nano-cages in TEM conditions. Negatively stained TEM micrographs indicate that all of the DPS-derived single mutants can form nano-cages. Scale bars indicate 50 nm.

The TEM images were analyzed using ImageJ software to calculate the average Feret's diameter of the protein capsules. The average Feret's diameter of wild type DPS is around 9.2 nm and this value is similar to that reported by Grant *et al*³. The

size of the mutants was either equal to or slightly bigger than the wild-type protein (Figure 4.13) with DPS Y16A being significantly larger than DPS. It should be mentioned that in similar systems we have observed, conditions for obtaining TEM micrographs can encourage oligomerization that is not observed in solution²² so this type of analysis, while suggestive and not unuseful, should be taken with a grain of salt.

Figure 4.13 Quantification of nano-cage sizes determined by TEM. Error bars represent the standard deviation of 100 replicates. T-Test: *** $P < 0.0001$ (between DPS and Y16A).

4.5.5 Size exclusion chromatography (SEC)

Size exclusion chromatography is a widely used technique for the purification and analysis of biological polymers, such as proteins. With size exclusion chromatography, compounds can be well separated rapidly based on a wide range of molecular sizes with good sensitivity. The advantages of this technique are that there is no sample loss because solutes do not interact with the stationary phase and the biological activity of the particles are preserved because it can be performed in buffer.

To determine the role the mutated residues play in assembly in solution, the mutants were analyzed by SEC (Figure 4.15). The retention volumes of each species

were correlated to the molecular weights of protein standards (Figure 4.14) which were then used to calculate approximate oligomerization states. Wild-type DPS protein formed solely 12-mer, consistent with the literature^{5, 33}. All of the DPS single mutants assembled into 12-mers but four mutants also formed smaller oligomers. Mutation of W52 resulted in some unassembled monomer (6%) and the mutant Y16A formed a small amount (16%) of dimer along with 12-mer. Individual mutation of R83 or R133 resulted in 40% and 50% dimer formation respectively. Following this trend, double mutation of both these residues destroyed the ability of the protein to assemble into 12-mer, cleanly forming 100% dimer. This is consistent with the TEM micrographs, which showed that R83A/R133A produced no observable nano-cages. The hydrogen bonding network in three-fold interfaces were disrupted by the mutations as evidenced by the fact of protein disassembly to partially or solely dimers, which suggested that these amino acids are hot spots.

Figure 4.14 Size exclusion chromatography calibration. (*Left*) Chromatogram of calibration proteins. (*Right*) Molecular weight correlation to K_{av} as defined in the experimental part. The trace with blue dextran indicates the void volume.

Figure 4.15 Two residues control the assembly from dimer to 12-mer. (*Left*) SEC chromatograms of the mutants. (*Right*) Quantitation of the populations determined from the SEC chromatograms with respect to oligomerization state.

4.5.6 Temperature-dependent circular dichroism analysis

Circular dichroism (CD) is a widely used method for studying secondary structure and protein folding. It measures differences in the absorption of left-handed versus right-handed polarized light which arise from structural asymmetry. The far-UV absorbance spectra of proteins and peptide molecules are dominated by the π - π^* and n - π^* transitions of amide groups. The CD spectra are influenced by the geometries of the polypeptide backbones which are reflective of different types of secondary structures³⁴. For example, α -helical proteins have characteristic negative bands at 208 nm and 222 nm and a positive band at 193 nm³⁵. Proteins with well-defined antiparallel β -pleated sheets have negative bands at 218 nm and positive bands at 195 nm³⁶.

Because of the high dependency on the conformation of the protein spectra, CD can be used to evaluate the structure of unknown proteins or conformation changes due to the mutation or environmental changes³⁷. It is also widely used to determine protein stability and thermal reversibility by following the unfolding and folding as a function of temperature.

Ellipticity (θ) is the unit of CD signal and is measured in millidegree (mdeg) precision. The most commonly used unit in current literature is mean residue ellipticity θ_{mrw} ($\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$) which is normalized for concentration and number of residues of proteins or peptides.

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

Where θ_{obs} is the observed ellipticity in mdeg, M is the molecular weight of protein, c is the concentration of protein in mg/mL, l is the path length of the cell in cm and n is the number of residues.

Wild type Dps is a highly helical protein as observed from its crystal structure. To examine the effects that mutations have on protein secondary structure and thermal stability, CD experiments were carried out to monitor the disruption of helical structure with increasing temperature. None of the interfacial residues affect the ability of DPS to fold into α -helical protein as evidenced by prominent minima at 208 and 222 nm in the CD spectra of the proteins (Figure 4.17 (right)), although all of the mutants exhibited reduced secondary structure. The lone exception is DPS D78A, which gives a nearly identical spectrum to DPS itself. The residue that causes the greatest reduction in structure upon mutation is W160. The double mutant, R83A/R133A, is less structured than its respective single mutants.

Increasing the temperature results in the melting of the secondary structure (Figures 4.16 and 4.17). By monitoring the 222 nm band, it is evident that, for all the proteins, this unfolding is cooperative, and that these residues do not control the ability of DPS to form a tight, stable fold. Most of the mutants melt at a lower temperature than wild-type DPS (Figure 4.18), demonstrating that even single interfacial mutations can affect the protein thermal stability and these residues are indeed hot spots. The exception to this trend is D146A which has nearly identical

stability as wild-type DPS. The residues that affect the stability the greatest upon single mutation to alanine are R83 and R133. The melting temperature of these mutants which shave the side chains of these residues individually were the lowest among the point mutants, and complete disruption of the hydrogen bond network through double mutation resulted in the least stable protein with a melting point depressed by nearly 25 °C. These two residues were also consistently predicted to be the most energetically significant by virtual alanine scanning (Table 4.1). As the T_m depression for DPS R83A/133A was less than the sum of that for R83A and R133A, it also suggests that the individual mutations were not additive and thus that the two residues work cooperatively in the wild-type protein as would be consistent with the analysis of the crystal structure.

Figure 4.16 CD spectra for thermal melts used in Figure 4.18 for DPS-derived proteins. These are averages of at least three melts.

Slowly reducing the temperature following thermal melting allows for the exploration of folding reversibility (Figure 4.17 (right)) by comparing the structure before and after the melt. Although some visible precipitation for a few of the mutants was observed under these conditions, poor reversibility was also observed for the wild-type protein, suggesting that this is a property of DPS itself and not due to mutagenesis. This is consistent with other work we have performed with this family of proteins. Interestingly we have repeatedly seen that most BFR derivatives do not share this property. Although no clear trends are evident, thermal stability does not seem to follow that of reversibility which is consistent with other work we have done in these systems^{33a, 22}. Interestingly, the double mutant R83A/R133A which was the least stable is among the most reversible which is possibly due to the fact that it only forms dimer in solution.

Figure 4.17 The role mutation plays in nano-cage protein thermal stability and ability to refold after denaturation. (*Left*) Thermal transitions of the DPS derivatives as monitored by CD. (*Right*) CD spectra of the DPS derivatives before denaturation (solid lines) and after slow cooling post-thermal denaturation (dashed lines).

Figure 4.18 The identification of thermally significant interfacial residues. (*Left*) Quantified transition melting temperatures from the data presented in Figures 4.16, 4.17. Error bars represent the standard deviation of three replicates. (*Right*) Residular thermal transition temperatures mapped onto a DPS monomer.

4.6 Conclusion

Residue	Interface	$\Delta\Delta G^*$ (kcal/mole)	Tm (°C)	ΔT_m (°C)	Oligomerization state from SEC
Wild type	-	-	66.7±1.0	-	12-mer (100%)
Y16A	-	-	54.8±0.3	11.9±1.3	12-mer (84%) + Dimer (16%)
R18A	C3	3.93	55.2±0.4	11.5±1.4	12-mer (100%)
R83A	C3	9.84	50.4±0.7	16.3±1.7	12-mer (61%) + Dimer (39%)
R133A	C3	6.11	48.1±0.4	18.6±1.4	12-mer (51%) + Dimer (49%)
R83A/R133A	C3	-	41.4±0.4	25.3±1.4	Dimer (100%)
D146A	C3	1.48	66.6±0.9	0.1±1.9	12-mer (100%)
W160A	C3	2.42	57.3±0.7	9.4±1.7	12-mer (100%)
W52A	C2	3.35	56.3±0.4	10.4±1.4	12-mer (94%) + monomer (6%)
D78A	C2	1.09	61.9±0.7	4.8±1.7	12-mer (100%)

Table 4.3 The properties of DPS proteins determined by virtual alanine shaving computational calculation, CD and SEC analysis. $\Delta\Delta G^*$ is obtained from the virtual alanine scanning studies.

Protein-protein interactions are crucial for cell signaling and building up complex oligomeric architectures essential for cell structure and maintenance. Disorders associated with disruption of protein-protein interactions or their stabilization can manifest as cancer or senility for example. The ferritins and related proteins make up a class of proteins that self-assemble into hollow nano-cages, help maintain iron-

homeostasis and are directly or indirectly involved in preventing DNA damage through reactive oxygen species. In addition, a disease caused by neurotoxicity has recently been attributed to the misfolding and aggregation of ferritin proteins in human brain³⁸. Through research carried out on the self-assembly of ferritin protein cages, it is hoped that not only can insight be established into this important and ubiquitous protein class, but also into the fundamentals of protein-protein interactions in general. Moreover, these proteins have recently been applied to the formation of inorganic nano-materials^{28,39-41} and understanding the fundamentals of their folding and assembly could help to advance this research.

The ferritin-like DNA-binding protein from starved cells (DPS) from *E. coli* was used in this study with the goals of building upon the fundamentals of protein-protein interactions and self-assembly. By inspection of the crystal structure in concert with virtual alanine scanning (Figures 4.8, Table 4.1), specific residues at symmetry related protein-protein interfaces in the DPS protein cage were selected for mutagenesis to determine their role in stability and oligomerization. These interfacial amino acids were either aromatic or involved in salt bridging interactions. Six of the selected residues are at the three-fold axis, and two residues are at the two-fold axis (Figure 4.7). All nine of the mutants formed cooperatively folded α -helical structures (Figure 4.17) and all the mutants, except the double mutant R83A/R133A, possessed the ability to form nano-cages as evidenced by SEC (Figure 4.15) and TEM (Figure 4.12). All DPS mutants melted at a lower temperature than wild-type DPS (Figure 4.18), suggesting that even single point mutations can affect protein thermal stability if they are at appropriate positions. Taken together, these data confirm that these residues are indeed hot spots and that hot spots can be rationally identified in some systems.

The most striking results involve R83 and R133. These residues, which are involved in an extended hydrogen bond network (Figure 4.8) at the three-fold symmetry axis and were, of the residues chosen for mutation, predicted to be the most energetically significant by virtual alanine scanning (Table 4.1). Moreover arginine and lysine are found in similar positions in DPS from a wide array of species (Figure 4.9). The melting temperature (T_m) of mutants which shave the side chains of these residues individually were the lowest among the point mutants, and complete disruption of the hydrogen bond network through double mutation resulted in the least stable protein emphasizing the importance of these residues. It is still unclear whether it is additive since the T_m doesn't represent for energy value even though the T_m depression of 25.2 °C for double mutant R83A/R133A was less than the sum of that for R83A and R133A. The importance of these residues is further reinforced by the fact that individual mutation of R83 and R133 to alanine results in proteins that form a mixture of 12-mer and a substantial amount dimer (39% and 49% respectively). Consistent with the thermal profile, combining these mutations results in a protein that only forms dimer. TEM also confirms that mutating these residues results in protein cage disassembly.

There has been some previous work with Dps homologs to alter the oligomerization state^{42,43}. For example *D. radiodurans* Dps-1 was reported to switch from a 12-mer to a dimer upon buffer exchange from high to low salt⁴⁴ and the *L. innocua* DPS 12-mer can be made to dissociates to a dimer by lowering the pH below 2.0.⁴⁵ Moreover, the Dps from *M. smegmatis* dissociates into lower order oligomers at low temperature (4 °C)^{46,47}. These examples all involve changing solution conditions and while they may be providing relevant information that can be extrapolated, ways to alter the oligomerization state in standard conditions with minimal perturbation are necessary

for delivery and materials applications and as controls to understand the role of oligomerization in DNA binding for example. The few mutants that have been discovered to alter the Dps oligomerization state involved deletions of relatively large subdomains that have been implicated in other roles such as making interactions with DNA or in iron transport.⁴⁸ Therefore the discovery of a simple double mutant that cleanly alters the oligomerization state not only provides insight into fundamentals of this system, it could also provide direction to the development of unique applications and better tools.

The straightforward rationality of this system is further emphasized through recalling that these two residues are located at the ferritin-like three-fold symmetry axis and hence the trimer interface. As the dimer has been shown to be an intermediate in the assembly process⁴⁹ it is reasonable that destabilization of the three-fold symmetry would result in a protein that can only assemble as a dimer. We are conducting a more systematic study to probe the generality of a design rule that links disruption of a symmetry related interface with the nature of the oligomerization state produced.

It is useful to compare the results of this study to those using the maxi-ferritin BFR. In that previous research, we discovered four residues that completely obliterate cage formation when mutated to alanine²². Surprisingly one of these is located at the two-fold axis of symmetry suggesting that there either is a global conformational change or that an oligomerization non-productive dimer is possible. In addition we also discovered two residues that increase the overall stability of the protein when mutated. This result is remarkable not only in that overall thermal and chemical stability was enhanced, but that both of these mutations only form dimer, further supporting the running hypothesis of the existence of a non-productive dimer, the possibility of

which we are currently exploring⁴⁹. These complexities contrasted with the relatively very rational and straightforward DPS further emphasizes the structural energetic distinctness of these mini- and maxi-ferritins. Fundamental studies of how protein-protein interactions control the stability and oligomerization of cage proteins could aid in the development of future applications, provide controls for further studies, and direct investigations into using these proteins as building blocks in the design and discovery of novel oligomerized architectures.

4.7 References

1. Almiron, M.; Link, A. J.; Furlong, D.; Kolter, R., A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes & Development* **1992**, *6* (12b), 2646-2654.
2. Crichton, R. R.; Wilmet, S.; Leggsyler, R.; Ward, R. J., Molecular and cellular mechanisms of iron homeostasis and toxicity in mammalian cells. *J Inorg Biochem* **2002**, *91* (1), 9-18.
3. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* **1998**, *5* (4), 294-303.
4. Lewin, A.; Moore, G. R.; Le Brun, N. E., Formation of protein-coated iron minerals. *Dalton T* **2005**, (22), 3597-3610.
5. Chiancone, E.; Ceci, P., The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2010**, *1800* (8), 798-805.
6. Romão, C. V.; Mitchell, E. P.; McSweeney, S., The crystal structure of *Deinococcus radiodurans* Dps protein (DR2263) reveals the presence of a novel metal centre in the N terminus. *JBIC Journal of Biological Inorganic Chemistry* **2006**, *11* (7), 891-902.
7. Bellapadrona, G.; Stefanini, S.; Zamparelli, C.; Theil, E. C.; Chiancone, E., Iron Translocation into and out of *Listeria innocua* Dps and Size Distribution of the Protein-enclosed Nanomineral Are Modulated by the Electrostatic Gradient at the 3-fold "Ferritin-like" Pores. *Journal of Biological Chemistry* **2009**, *284* (28), 19101-19109.
8. Holm, L.; Park, J., DaliLite workbench for protein structure comparison. *Bioinformatics* **2000**, *16* (6), 566-567.
9. Lawson, D. M.; Treffry, A.; Artymiuk, P. J.; Harrison, P. M.; Yewdall, S. J.; Luzzago, A.; Cesareni, G.; Levi, S.; Arosio, P., Identification of the Ferroxidase Center in Ferritin. *Febs Lett* **1989**, *254* (1-2), 207-210.
10. Zhao, G. H.; Bou-Abdallah, F.; Arosio, P.; Levi, S.; Janus-Chandler, C.; Chasteen, N. D., Multiple pathways for mineral core formation in mammalian apoferritin. The role of hydrogen peroxide. *Biochemistry* **2003**, *42* (10), 3142-3150.

11. Bunker, J.; Lowry, T.; Davis, G.; Zhang, B.; Brosnahan, D.; Lindsay, S.; Costen, R.; Choi, S.; Arosio, P.; Watt, G., Kinetic studies of iron deposition catalyzed by recombinant human liver heavy, and light ferritins and bacterioferritin using O and HO as oxidants. *Biophysical Chemistry* **2005**, *114* (2-3), 235-244.
12. Franceschini, S.; Ceci, P.; Alaleona, F.; Chiancone, E.; Ilari, A., Antioxidant Dps protein from the thermophilic cyanobacterium *Thermosynechococcus elongatus*. *FEBS Journal* **2006**, *273* (21), 4913-4928.
13. Zhao, G., Iron and Hydrogen Peroxide Detoxification Properties of DNA-binding Protein from Starved Cells. A ferritin-like DNA-binding protein of *Escherichia coli* *Journal of Biological Chemistry* **2002**, *277* (31), 27689-27696.
14. Tonello, F.; Pellizzari, R.; Pasqualato, S.; Grandi, G.; Peggion, E.; Montecucco, C., Recombinant and truncated tetanus neurotoxin light chain: Cloning, expression, purification, and proteolytic activity. *Protein Expression Purification* **1999**, *15* (2), 221-227.
15. Su, M. H.; Cavallo, S.; Stefanini, S.; Chiancone, E.; Chasteen, N. D., The so-called *Listeria innocua* ferritin is a Dps protein. Iron incorporation, detoxification, and DNA protection properties. *Biochemistry* **2005**, *44* (15), 5572-5578.
16. Martinez, A.; Kolter, R., Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *J Bacteriol* **1997**, *179* (16), 5188-5194.
17. Ceci, P., DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus. *Nucleic Acids Research* **2004**, *32* (19), 5935-5944.
18. Ilari, A.; Stefanini, S.; Chiancone, E.; Tsernoglou, D., The dodecameric ferritin from *Listeria innocua* contains a novel intersubunit iron-binding site. *Nat Struct Biol* **2000**, *7* (1), 38-43.
19. Ceci, P., The Dps protein of *Agrobacterium tumefaciens* does not bind to DNA but protects it toward oxidative cleavage: X-ray crystal structures, iron binding, and hydroxyl-radical scavenging properties. *Journal of Biological Chemistry* **2003**, *278* (22), 20319-20326.
20. Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **1998**, *280* (1), 1-9.
21. Kortemme, T.; Kim, D. E.; Baker, D., Computational alanine scanning of protein-protein interfaces. *Sci STKE* **2004**, *2004* (219), pl2
22. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R. L.; Sun, X. M.; Orner, B. P., Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* **2010**, *285* (16), 12078-12086.
23. (a) Cunningham, B. C.; Wells, J. A., High-Resolution Epitope Mapping of Hgh-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science* **1989**, *244* (4908), 1081-1085; (b) Jaenicke, R.; Rudolph, R., Refolding and association of oligomeric proteins. *Methods Enzymol* **1986**, *131*, 218-50.
24. Chou, P. Y.; Adler, A. J.; Fasman, G. D., Lac Repressor - Conformational Prediction and Circular-Dichroism Studies. *Fed Proc* **1974**, *33* (5), 1373-1373.

25. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J., Hot spots-A review of the protein-protein interface determinant amino-acid residues. *Proteins* **2007**, *68* (4), 803-812.
26. Jin, L.; Cohen, F. E.; Wells, J. A., Structure from Function - Screening Structural Models with Functional Data. *P Natl Acad Sci USA* **1994**, *91* (1), 113-117.
27. Cunningham, B. C.; Jhurani, P.; Ng, P.; Wells, J. A., Receptor and Antibody Epitopes in Human Growth-Hormone Identified by Homolog-Scanning Mutagenesis. *Science* **1989**, *243* (4896), 1330-1336.
28. Fan, R. L.; Chew, S. W.; Cheong, V. V.; Orner, B. P., Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* **2010**, *6* (14), 1483-1487.
29. Jencks, W. P., On the Attribution and Additivity of Binding-Energies. *P Natl Acad Sci-Biol* **1981**, *78* (7), 4046-4050.
30. Kortemme, T.; Baker, D., Computational design of protein-protein interactions. *Curr Opin Chem Biol* **2004**, *8* (1), 91-97.
31. (a) Kortemme, T.; Joachimiak, L. A.; Bullock, A. N.; Schuler, A. D.; Stoddard, B. L.; Baker, D., Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **2004**, *11* (4), 371-379; (b) Kortemme, T.; Baker, D., A simple physical model for binding energy hot spots in protein-protein complexes. *P Natl Acad Sci USA* **2002**, *99* (22), 14116-14121.
32. Massover, W. H., Ultrastructure of Ferritin and Apoferritin - a Review. *Micron* **1993**, *24* (4), 389-437.
33. (a) Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009**, *48* (24), 5623-5630; (b) Swift, J.; Wehbi, W. A.; Kelly, B. D.; Stowell, X. F.; Saven, J. G.; Dmochowski, I. J., Design of functional ferritin-like proteins with hydrophobic cavities. *J Am Chem Soc* **2006**, *128* (20), 6611-6619.
34. Whitemore, L.; Wallace, B. A., Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **2007**, *89* (5), 392-400.
35. Holzwart, G.; Doty, P., Ultraviolet Circular Dichroism of Polypeptides. *J Am Chem Soc* **1965**, *87* (2), 218-228.
36. Greenfield, N.; Fasman, G. D., Computed Circular Dichroism Spectra for Evaluation of Protein Conformation. *Biochemistry* **1969**, *8* (10), 4108-4116.
37. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **2007**, *1* (6), 2876-2890.
38. Levi, S.; Cozzi, A.; Arosio, P., Neuroferritinopathy: a neurodegenerative disorder associated with L-ferritin mutation. *Best Pract Res Clin Ha* **2005**, *18* (2), 265-276.
39. Meldrum, F. C.; Wade, V. J.; Nimmo, D. L.; Heywood, B. R.; Mann, S., Synthesis of Inorganic Nanophase Materials in Supramolecular Protein Cages. *Nature* **1991**, *349* (6311), 684-687.
40. Uchida, M.; Klem, M. T.; Allen, M.; Suci, P.; Flenniken, M.; Gillitzer, E.; Varpness, Z.; Liepold, L. O.; Young, M.; Douglas, T., Biological containers: Protein cages as multifunctional nanoplatfoms. *Adv Mater* **2007**, *19* (8), 1025-1042.
41. Uchida, M.; Flenniken, M. L.; Allen, M.; Willits, D. A.; Crowley, B. E.; Brumfield, S.; Willis, A. F.; Jackiw, L.; Jutila, M.; Young, M. J.; Douglas, T.,

- Targeting of cancer cells with ferrimagnetic ferritin cage nanoparticles. *J Am Chem Soc* **2006**, *128* (51), 16626-16633.
42. Chowdhury, R. P., Vijayabaskar, M. S., Vishveshwara, S., Chatterji, D. Molecular mechanism of in vitro oligomerization of Dps from *Mycobacterium smegmatis*: Mutations of the residues identified by "interface cluster" analysis. *Biochemistry* **2008**, *47* (42), 11110-11117.
 43. Chowdhury, R. P., Saraswathi, R., Chatterji, D. Mycobacterial Stress Regulation: The Dps "Twin Sister" Defense Mechanism and Structure-Function Relationship. *Iubmb Life* **2010**, *62* (1), 67-77.
 44. Grove, A., Wilkinson, S. P. Differential DNA binding and protection by dimeric and dodecameric forms of the ferritin homolog Dps from *Deinococcus radiodurans*. *Journal of Molecular Biology* **2005**, *347* (3), 495-508.
 45. Chiaraluce, R., Consalvi, V., Cavallo, S., Ilari, A., Stefanini, S., Chiancone, E. The unusual dodecameric ferritin from *Listeria innocua* dissociates below pH 2.0. *Eur J Biochem* **2000**, *267* (18), 5733-5741.
 46. Ceci, P., Ilari, A., Falvo, E., Giangiacomo, L., Chiancone, E. Reassessment of protein stability, DNA binding, and protection of *Mycobacterium smegmatis* Dps. *Journal of Biological Chemistry* **2005**, *280* (41), 34776-34785.
 47. Gupta, S., Chatterji, D. Bimodal protection of DNA by *Mycobacterium smegmatis* DNA-binding protein from stationary phase cells. *Journal of Biological Chemistry* **2003**, *278* (7), 5235-5241.
 48. Roy, S., Saraswathi, R., Gupta, S., Sekar, K., Chatterji, D., Vijayan, M. 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* **2007**, *370* (4), 752-767.
 49. Ardejani, M. S., Li, N. X., Orner, B. P. Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* **2011**, *50* (19), 4029-4037.

Appendix 1. The gene design of BFR synthesis

BFR	
A MKGDTKVINYLNLKLLGNELVAINQYFLHARMFKNWGLKRLNDUEYHESIDEMKHADRYIERILFLEGLPNLQDLGKLNIGEDVEEMLRSDLALELDGAKNLREAIGYADSUHDYVSRDMMIEILRDEEGHIDWLETDLIQKMLQNYLQAQIREEG	
B BFR (474bp) 5'-3'	
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	5'-3' 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-ACG-TTC R8 CAG-TTT-ACC-CAG-ATC-CTG-CAG-GTT-CCG R9 CAT-TTC-TTC-CAC-ATC-TTC-ACC-GAT-GTT R10 CAG-TTC-CAG-TGC-CAG-ATC-GCT-ACG-CAG R11 TGC-TTC-ACG-CAG-GTT-TTT-TGC-ACC-ATC R12 ATC-ATG-CAC-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
C 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	

Figure A1. 1 (A) Protein sequence of BFR. (B) Forward and reverse oligonucleotides used to synthesize the BFR gene. (C) Primers used for extension PCR.

BFR E157A	R17 ACC- <u>TGC</u> -TTC-ACG-GAT-CTG-TGC-CTG-CAG antisense GAG-GAG-AAG-CCC-GGT-TCA-ACC- <u>TGC</u> -TTC-ACG-GAT
BFR E128A	F15 GAA- <u>GCA</u> -GGT-CAT-ATC-GAT-TGG-CTG-GAA R14 GAT-ATG-ACC- <u>TGC</u> -TTC-ATC-ACG-CAG-GAT
BFR Y114A	F13 GAT-AGC-GTG-CAT-GAT- <u>GCA</u> -GTG-AGC-CGT R13 TTC-GAT-CAT-CAT-ATC-ACG-GCT-CAC- <u>TGC</u>
BFR Y45A	F5 CTG-AAA-CGT-CTG-AAC-GAT-GTG-GAA- <u>GCA</u> R5 ATC-GAT-GCT-TTC-ATG- <u>TGC</u> -TTC-CAC-ATC
BFR K33A	F4 CAT-GCA-CGT-ATG-TTC- <u>GCA</u> -AAC-TGG-GGT R4 GTT-CAG-ACG-TTT-CAG-ACC-CCA-GTT- <u>TGC</u>
BFR D56A	F7 GCA- <u>GCA</u> -CGT-TAC-ATC-GAA-CGT-ATC-CTG R6 GAT-GTA-ACG- <u>TGC</u> -TGC-ATG-TTT-CAT-TTC
BFR D56A E60A	F7 GCA- <u>GCA</u> -CGT-TAC-ATC- <u>GCA</u> -CGT-ATC-CTG R6 GAT-GTA-ACG- <u>TGC</u> -TGC-ATG-TTT-CAT- <u>TGC</u>

(All the oligos are 5'→3')

BFR R61A

R61A-F (5'-3') CAGATCGTTACATCGAAGCAATCCTGTTCTGGAAGG
R61A-R (5'-3') CCTTCCAGGAACAGGATTGCTTCGATGTAACGATCTG
BFR R30A
R30A-F (5'-3') CCAGTACTTCCTGCATGCAGCAATGTTCAAAAAGTGGGGTC
R30A-R (5'-3') GACCCAGTTTTTGAACATTGCTGCATGCAGGAAGTACTGG

Figure A1. 2 (*Top*) Replaced oligos from the scheme in Figure S1 to generate the mutant BFR genes. The mutant codons are boxed. (*Bottom*) Primers used in QuikChange mutagenesis for BFR R61A and BFR R30A using the BFR as a template.

Appendix 2. Sequencing results of BFR and BFR derived genes

The sequencing results were aligned with the BFR gene using BLAST from the NCBI website. The mutant codon is highlighted in light gray.

BFR

Query is BFR native, Sbjct is the sequencing result.

Score = 912 bits (474), Expect = 0.0

Identities = 474/474 (100%), Gaps = 0/474 (0%)

Strand=Plus/Plus

Query	1	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG	60
Sbjct	24	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG	83
Query	61	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGTCTGAAACGTCTG	120
Sbjct	84	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGTCTGAAACGTCTG	143
Query	121	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA	180
Sbjct	144	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA	203
Query	181	CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	240
Sbjct	204	CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	263
Query	241	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGTGGTGCAAAAAAC	300
Sbjct	264	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGTGGTGCAAAAAAC	323
Query	301	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	360
Sbjct	324	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	383
Query	361	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTG	420
Sbjct	384	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTG	443
Query	421	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	474
Sbjct	444	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	497

BFR E157A

E-GAA

A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 906 bits (471), Expect = 0.0
 Identities = 473/474 (99%), Gaps = 0/474 (0%)
 Strand=Plus/Plus

```

Query 1  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 26  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 85

Query 61  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGCTGAAACGTCTG 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 86  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGCTGAAACGTCTG 145

Query 121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 146  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 205

Query 181  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 206  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT 265

Query 241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGTGGTCAAAAAAC 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 266  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGTGGTCAAAAAAC 325

Query 301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG 360
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 326  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG 385

Query 361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTG 420
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 386  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTG 445

Query 421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAGGT 474
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 446  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAGGT 499

```

BFR E128A

E-GAA
 A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 906 bits (471), Expect = 0.0
 Identities = 473/474 (99%), Gaps = 0/474 (0%)
 Strand=Plus/Plus

```

Query 1  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 30  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 89

Query 61  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGCTGAAACGTCTG 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 90  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGCTGAAACGTCTG 149

Query 121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 150  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 209

Query 181  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 210  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT 269

Query 241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGTGGTCAAAAAAC 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Sbjct  270  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAAC  329

Query  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  330  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  389

Query  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420
          ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Sbjct  390  ATCGAAATCCTGCGTGATGAAGCAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  449

Query  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  474
          |||||||||||||||||||
Sbjct  450  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  503

```

BFR Y114A

Y-TAC

A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 894 bits (465), Expect = 0.0

Identities = 471/474 (99%), Gaps = 0/474 (0%)

Strand=Plus/Plus

```

Query   1  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  25  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG  84

Query   61  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACGGGGTCTGAAACGTCTG  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  85  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACGGGGTCTGAAACGTCTG  144

Query   121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  145  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  204

Query   181  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  205  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  264

Query   241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAAC  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  265  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAAC  324

Query   301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360
          |||||||||||||||||||||||||||||||||||||||||||
Sbjct  325  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATGCAGTGAGCCGTGATATGATG  384

Query   361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420
          ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Sbjct  385  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  444

Query   421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  474
          |||||||||||||||||||
Sbjct  445  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  498

```

BFR R61A

R-CGT

A-GCA

Query is BFR native, Sbjct is the mutation sequence

Score = 841 bits (932), Expect = 0.0

Identities = 471/474 (99%), Gaps = 0/474 (0%)
Strand=Plus/Plus

```

Query 1  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 25  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 84

Query 61  GCAATCAACCAGTACTTCTGCATGCACGTATGTTCAAAAAGTGGGTCTGAAACGTCTG 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 85  GCAATCAACCAGTACTTCTGCATGCACGTATGTTCAAAAAGTGGGTCTGAAACGTCTG 144

Query 121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 145  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 204

Query 181  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAACTGAACATCGGT 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 205  GCAATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAACTGAACATCGGT 264

Query 241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAAC 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 265  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAAC 324

Query 301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG 360
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 325  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG 384

Query 361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTG 420
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 385  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTG 444

Query 421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT 474
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 445  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT 498

```

BFR Y45A

Y-TAC
A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 894 bits (465), Expect = 0.0
Identities = 471/474 (99%), Gaps = 0/474 (0%)
Strand=Plus/Plus

```

Query 1  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 24  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG 83

Query 61  GCAATCAACCAGTACTTCTGCATGCACGTATGTTCAAAAAGTGGGTCTGAAACGTCTG 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 84  GCAATCAACCAGTACTTCTGCATGCACGTATGTTCAAAAAGTGGGTCTGAAACGTCTG 143

Query 121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 180
      |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct 144  AACGATGTGGAAGCACATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA 203

Query 181  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAACTGAACATCGGT 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 204  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAACTGAACATCGGT 263

Query 241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAAC 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 264  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAAC 323

```

```

Query 301 CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 324 CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG 383

Query 361 ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 384 ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG 443

Query 421 ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 444 ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT 497

```

BFR K33A

K-AAA

A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 892 bits (464), Expect = 0.0
 Identities = 468/470 (99%), Gaps = 0/470 (0%)
 Strand=Plus/Plus

```

Query 5 AAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTGGCAA 64
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 33 AAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTGGCAA 92

Query 65 TCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTGAACG 124
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 93 TCAACCAGTACTTCCTGCATGCACGTATGTTCCGCAAACTGGGGTCTGAAACGTCTGAACG 152

Query 125 ATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAACGTA 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 153 ATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAACGTA 212

Query 185 TCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGAAG 244
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 213 TCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGAAG 272

Query 245 ATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAACCTGC 304
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 273 ATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAACCTGC 332

Query 305 GTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGATCG 364
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 333 GTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGATCG 392

Query 365 AAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTGATCC 424
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 393 AAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTGATCC 452

Query 425 AGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT 474
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 453 AGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT 502

```

BFR D56A

D-GAT

A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 900 bits (468), Expect = 0.0
 Identities = 472/474 (99%), Gaps = 0/474 (0%)
 Strand=Plus/Plus

Query	1	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG	60
Sbjct	29	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG	88
Query	61	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACCTGGGGTCTGAAACGTCTG	120
Sbjct	89	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACCTGGGGTCTGAAACGTCTG	148
Query	121	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA	180
Sbjct	149	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGCACGTTACATCGAA	208
Query	181	CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	240
Sbjct	209	CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	268
Query	241	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACCTGGATGGTGCAAAAAAC	300
Sbjct	269	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACCTGGATGGTGCAAAAAAC	328
Query	301	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	360
Sbjct	329	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	388
Query	361	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG	420
Sbjct	389	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG	448
Query	421	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	474
Sbjct	449	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	502

BFR D56A E60A

D-GAT
E-GAA
A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 890 bits (463), Expect = 0.0
Identities = 469/472 (99%), Gaps = 0/472 (0%)
Strand=Plus/Plus

Query	3	GAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTGGC	62
Sbjct	25	GAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTGGC	84
Query	63	AATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACCTGGGGTCTGAAACGTCTGAA	122
Sbjct	85	AATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACCTGGGGTCTGAAACGTCTGAA	144
Query	123	CGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAACG	182
Sbjct	145	CGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGCACGTTACATCGAACG	204
Query	183	TATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGA	242
Sbjct	205	TATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGA	264
Query	243	AGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACCTGGATGGTGCAAAAAACCT	302
Sbjct	265	AGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACCTGGATGGTGCAAAAAACCT	324
Query	303	GCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGAT	362

```

Sbjct  325  GCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGAT  384

Query  363  CGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGAAACCGAACTGGATCTGAT  422
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  385  CGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGAAACCGAACTGGATCTGAT  444

Query  423  CCAGAAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  474
      |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  445  CCAGAAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  496

```

BFR R30A

R-CGT
A-GCA

Query is BFR native, Sbjct is the mutation sequence.

Score = 841 bits (932), Expect = 0.0
Identities = 471/474 (99%), Gaps = 0/474 (0%)
Strand=Plus/Plus

```

Query   1  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG  60
      |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  25  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG  84

Query   61  GCAATCAACCAGTACTTCTGCATGCACGTATGTTCAAAAACGGGGTCTGAAACGTCTG  120
      ||||||||||||||||||| |||||||||||||||||||||||||||
Sbjct  85  GCAATCAACCAGTACTTCTGCATGCACATGTTCAAAAACGGGGTCTGAAACGTCTG  144

Query  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180
      |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  145  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  204

Query  181  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240
      |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  205  CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  264

Query  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAAC  300
      |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  265  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAAC  324

Query  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360
      |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  325  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  384

Query  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGAAACCGAACTGGATCTG  420
      |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  385  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGAAACCGAACTGGATCTG  444

Query  421  ATCCAGAAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  474
      |||||||||||||||||||||||||||||||||||||||||||
Sbjct  445  ATCCAGAAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  498

```


Appendix 3. The gene design of DPS synthesis

DPS

A

MSTAKLUKSKATNLLYTRNDUSDSEKKATVELLNRRQVIQFIDLSLITKQAHWNMRGANFIAUHEMLDGFRTALIDHLDMAERA
UQLGGVALGTTQVINSKTPKSYPLDIHNUQDHLKELADRYAIVANDURKAIGEAKDDDTADILTAASRDLDKFLWFIECNIE

B

5' -3'		5' -3'	
F1	ATG-AGC-ACC-GCA-AAA-CTG-GTG-AAA-AGC	R1	CAG-GTT-GGT-TGC-TTT-GCT-TTT-CAC-CAG
F2	AAA-GCA-ACC-AAC-CTG-CTG-TAC-ACC-CGT	R2	ATC-GCT-CAC-ATC-GTT-ACG-GGT-GTA-CAG
F3	AAC-GAT-GTG-AGC-GAT-AGC-GAA-AAA-AAA	R3	CAG-TTC-CAC-GGT-TGC-TTT-TTT-TTC-GCT
F4	GCA-ACC-GTG-GAA-CTG-CTG-AAC-CGT-CAG	R4	GAT-GAA-CTG-GAT-CAC-CTG-ACG-GTT-CAG
F5	GTG-ATC-CAG-TTC-ATC-GAT-CTG-AGC-CTG	R5	TGC-CTG-TTT-GGT-GAT-CAG-GCT-CAG-ATC
F6	ATC-ACC-AAA-CAG-GCA-CAT-TGG-AAC-ATG	R6	GAA-GTT-TGC-ACC-ACG-CAT-GTT-CCA-ATG
F7	CGT-GGT-GCA-AAC-TTC-ATC-GCA-GTG-CAT	R7	ACC-ATC-CAG-CAT-TTC-ATG-CAC-TGC-GAT
F8	GAA-ATG-CTG-GAT-GGT-TTC-CGT-ACC-GCA	R8	CAG-ATG-ATC-GAT-CAG-TGC-GGT-ACG-GAA
F9	CTG-ATC-GAT-CAT-CTG-GAT-ACC-ATG-GCA	R9	CTG-CAC-TGC-ACG-TTC-TGC-CAT-GGT-ATC
F10	GAA-CGT-GCA-GTG-CAG-CTG-GGT-GGT-GTG	R10	GGT-GGT-ACC-CAG-TGC-CAC-ACC-ACC-CAG
F11	GCA-CTG-GGT-ACC-ACC-CAG-GTG-ATC-AAC	R11	CAG-CGG-GGT-TTT-GCT-GTT-GAT-CAC-CTG
F12	AGC-AAA-ACC-CCG-CTG-AAA-AGC-TAC-CCG	R12	GTT-ATG-GAT-ATC-CAG-CGG-GTA-GCT-TTT
F13	CTG-GAT-ATC-CAT-AAC-GTG-CAG-GAT-CAT	R13	TGC-CAG-TTC-TTT-CAG-ATG-ATC-CTG-CAC
F14	CTG-AAA-GAA-CTG-GCA-GAT-CGT-TAC-GCA	R14	ATC-GTT-TGC-CAC-GAT-TGC-GTA-ACG-ATC
F15	ATC-GTG-GCA-AAC-GAT-GTG-CGT-AAA-GCA	R15	TTT-TGC-TTC-ACC-GAT-TGC-TTT-ACG-CAC
F16	ATC-GGT-GAA-GCA-AAA-GAT-GAT-GAT-ACC	R16	GGT-CAG-GAT-ATC-TGC-GGT-ATC-ATC-ATC
F17	GCA-GAT-ATC-CTG-ACC-GCA-GCA-AGC-CGT	R17	GAA-TTT-ATC-CAG-ATC-ACG-GCT-TGC-TGC
F18	GAT-CTG-GAT-AAA-TTC-CTG-TGG-TTC-ATC	R18	TTC-GAT-GTT-GCA-TTC-GAT-GAA-CCA-CAG

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 A3. 1 (A) Protein sequence of DPS. (B) Forward and reverse oligonucleotides used to synthesize the DPS gene. (C) Primers used for extension PCR

DPS Y16A	F2	AAA-GCA-ACC-AAC-CTG-CTG-GCA-ACC-CGT
	R2	ATC-GCT-CAC-ATC-GTT-ACG-GGT-TGC-CAG
DPS W52A	F6	ATC-ACC-AAA-CAG-GCA-CAT-GCA-AAC-ATG
	R6	GAA-GTT-TGC-ACC-ACG-CAT-GTT-TGC-ATG
DPS D78A	F9	CTG-ATC-GAT-CAT-CTG-GCA-ACC-ATG-GCA
	R9	CTG-CAC-TGC-ACG-TTC-TGC-CAT-GGT-TGC
DPS D146A	F17	GCA-GCA-ATC-CTG-ACC-GCA-GCA-AGC-CGT
	R16	GGT-CAG-GAT-TGC-TGC-GGT-ATC-ATC-ATC
DPS W160A	F18	GAT-CTG-GAT-AAA-TTC-CTG-GCA-TTC-ATC
	R18	TTC-GAT-GTT-GCA-TTC-GAT-GAA-TGC-CAG

(All the oligos are 5'→3')

DPS R18A

R18A-F (5'-3') ACCAACCTGCTGTACACCGCAAACGATGTGAGCGATAGC

R18A-R (5'-3') GCTATCGCTCACATCGTTTGCGGTGTACAGCAGGTTGGT

DPS R83A

R83A-F (5'-3') CTGGATACCATGGCAGAAGCAGCAGTGCAGCTGGGTGGT

R83A-R (5'-3') ACCACCCAGCTGCACTGCTGCTTCTGCCATGGTATCCAG

DPS R133A

R133A-F (5'-3') TCGTGGCAAACGATGTGGCAAAAGCAATCGGTGAAGC

R133A-R (5'-3') GCTTCACCGATTGCTTTTGCCACATCGTTTGCCACGA

Figure A3. 2 (*Top*) Oligos substituted into the strategy described in Figure S1 to generate the mutant DPS genes. The mutant codons are boxed. (*Bottom*) Primers used in QuikChange mutagenesis for DPS R18A, DPS R83A, DPS R133A, DPS R83A/R133A using the DPS construct as a template.

Appendix 4. Sequencing results of DPS and DPS derived genes

The sequencing results (using the sense extension primer (Figure A3) unless otherwise noted) were aligned with the DPS gene using BLAST (NCBI). The mutant codon is highlighted in light gray. If part of the sequence did not align it was re-sequenced.

DPS

Query is DPS wild type, Sbjct is the sequencing result.

```
Query 1   ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT 60
          |||
Sbjct 24   ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT 82
          |||
Query 61  GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC 120
          |||
Sbjct 83  GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC 142
          |||
Query 121 ATCGATCTGAGCCTGATACCAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC 180
          |||
Sbjct 143 ATCGATCTGAGCCTGATACCAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC 202
          |||
Query 181 GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG 240
          |||
Sbjct 203 GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG 262
          |||
Query 241 GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCAACAGGTGATCAACAGC 300
          |||
Sbjct 263 GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCAACAGGTGATCAACAGC 322
          |||
Query 301 AAAACCCCGCTGAAAAGCTACCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 360
          |||
Sbjct 323 AAAACCCCGCTGAAAAGCTACCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 382
          |||
Query 361 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 420
          |||
Sbjct 383 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 442
          |||
Query 421 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 480
          |||
Sbjct 443 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 502
          |||
Query 481 TTCATCGAATGCAACATCGAA 501
          |||
Sbjct 503 TTCATCGAATGCAACATCGAA 523
```

Because one “A” is missing (in dark gray above), the same sample was sent for re-sequencing using the antisense extension primer (Figure S1). The BLAST result is as follows (the A in question is highlighted in dark gray):

Query	1	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	60
Sbjct	484	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	425
Query	61	GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC	120
Sbjct	424	GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC	365
Query	121	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAATTCATC	180
Sbjct	364	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAATTCATC	305
Query	181	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	240
Sbjct	304	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	245
Query	241	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	300
Sbjct	244	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	185
Query	301	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	360
Sbjct	184	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	125
Query	361	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	420
Sbjct	124	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	65
Query	421	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTG	477
Sbjct	64	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATC-NGATAAATTCCTG	9

DPS Y16A

Y-TAC
A-GCA

Query is DPS wild type, Sbjct is the sequencing result.

Query	1	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	60
Sbjct	24	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	82
Query	61	GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC	120
Sbjct	83	GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC	142
Query	121	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAATTCATC	180
Sbjct	143	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAATTCATC	202
Query	181	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	240
Sbjct	203	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	262
Query	241	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	300
Sbjct	263	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	322
Query	301	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	360
Sbjct	323	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	382
Query	361	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	420
Sbjct	383	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	442

```

Query 421 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 480
          |||
Sbjct 443 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 502

Query 481 TTCATCGAATGCAACATCGAA 501
          |||
Sbjct 503 TTCATCGAATGCAACATCGAA 523

```

Because one “A” is missing (in dark gray above), the same sample was sent for re-sequencing using the antisense extension primer (Figure S1). The BLAST result is as follows (the A in question is highlighted in dark gray):

```

Query 1 ATGAGCACCAGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT 60
          |||
Sbjct 483 ATGAGCACCAGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTGCAACCGTAACGAT 424

Query 61 GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC 120
          |||
Sbjct 423 GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC 364

Query 121 ATCGATCTGAGCCTGATACCAAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC 180
          |||
Sbjct 363 ATCGATCTGAGCCTGATACCAAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC 304

Query 181 GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG 240
          |||
Sbjct 303 GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG 244

Query 241 GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC 300
          |||
Sbjct 243 GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC 184

Query 301 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 360
          |||
Sbjct 183 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 124

Query 361 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 420
          |||
Sbjct 123 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 64

Query 421 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCG 458
          |||
Sbjct 63 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCG 26

```

DPS R18A

R-CGT
A-GCA

Query is DPS wild type, Sbjct is the sequencing result.

```

Query 1 ATGAGCACCAGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT 60
          |||
Sbjct 27 ATGAGCACCAGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGCAACGAT 86

Query 61 GTGAGCGATAGCGGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC 120
          |||
Sbjct 87 GTGAGCGATAGCGGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC 146

Query 121 ATCGATCTGAGCCTGATACCAAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC 180
          |||
Sbjct 147 ATCGATCTGAGCCTGATACCAAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC 206

```

Query	181	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	240
Sbjct	207	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	266
Query	241	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	300
Sbjct	267	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	326
Query	301	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	360
Sbjct	327	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	386
Query	361	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	420
Sbjct	387	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	446
Query	421	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	480
Sbjct	447	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	506
Query	481	TTCATCGAATGCAACATCGAA	501
Sbjct	507	TTCATCGAATGCAACATCGAA	527

DPS W52A

W-TGG
A-GCA

Query is DPS wild type, Sbjct is the sequencing result.

Query	1	ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT	60
Sbjct	32	ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT	91
Query	61	GTGAGCGATAGCGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC	120
Sbjct	92	GTGAGCGATAGCGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC	151
Query	121	ATCGATCTGAGCCTGATCACCAAACAGGCACATTGCAACATGCGTGGTGCAAACCTTCATC	180
Sbjct	152	ATCGATCTGAGCCTGATCACCAAACAGGCACATTGCAACATGCGTGGTGCAAACCTTCATC	211
Query	181	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	240
Sbjct	212	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	271
Query	241	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	300
Sbjct	272	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	331
Query	301	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	360
Sbjct	332	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	391
Query	361	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	420
Sbjct	392	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	451
Query	421	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	480

Sbjct	452	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	511
Query	481	TTCATCGAATGCAACATCGAA	501
Sbjct	512	TTCATCGAATGCAACATCGAA	532

DPS D78A

D-GAT
A-GCA

Query is DPS wild type, Sbjct is the sequencing result.

Query	1	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	60
Sbjct	28	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	87
Query	61	GTGAGCGATAGCGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC	120
Sbjct	88	GTGAGCGATAGCGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC	147
Query	121	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC	180
Sbjct	148	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC	207
Query	181	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG	240
Sbjct	208	GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGCAACCATG	267
Query	241	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	300
Sbjct	268	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	327
Query	301	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	360
Sbjct	328	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	387
Query	361	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	420
Sbjct	388	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	447
Query	421	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	480
Sbjct	448	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	507
Query	481	TTCATCGAATGCAACATCGAA	501
Sbjct	508	GTCATCGAATGCAACATCGAA	528

DPS R83A

R-CGT
A-GCA

Query is DPS wild type, Sbjct is the sequencing result.

Query	1	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	60
Sbjct	24	ATGAGCACCGCAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	83
Query	61	GTGAGCGATAGCGGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC	120

Sbjct	84	GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC	143
Query	121	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC	180
Sbjct	144	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC	203
Query	181	GCAGTGCATGAAATGCTGGATGGTTCCGTACCGCACTGATCGATCATCTGGATACCATG	240
Sbjct	204	GCAGTGCATGAAATGCTGGATGGTTCCGTACCGCACTGATCGATCATCTGGATACCATG	263
Query	241	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	300
Sbjct	264	GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	323
Query	301	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	360
Sbjct	324	AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA	383
Query	361	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	420
Sbjct	384	CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA	443
Query	421	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	480
Sbjct	444	GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG	503
Query	481	TTCATCGAATGCAACATCGAA	501
Sbjct	504	TTCATCGAATGCAACATCGAA	524

DPS R133A

R-CGT

A-GCA

Query is DPS wild type, Sbjct is the sequencing result

Query	1	ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	60
Sbjct	25	ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCGTAACGAT	84
Query	61	GTGAGCGATAGCGaaaaaaGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC	120
Sbjct	85	GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGAACCGTCAGGTGATCCAGTTC	144
Query	121	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC	180
Sbjct	145	ATCGATCTGAGCCTGATACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC	204
Query	181	GCAGTGCATGAAATGCTGGATGGTTCCGTACCGCACTGATCGATCATCTGGATACCATG	240
Sbjct	205	GCAGTGCATGAAATGCTGGATGGTTCCGTACCGCACTGATCGATCATCTGGATACCATG	264
Query	241	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	300
Sbjct	265	GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC	324

```

Query 301 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 360
      |||
Sbjct 325 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 384

Query 361 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 420
      |||
Sbjct 385 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 444

Query 421 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 480
      |||
Sbjct 445 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 504

Query 481 TTCATCGAATGCAACATCGAA 501
      |||
Sbjct 505 TTCATCGAATGCAACATCGAA 525

```

DPS D146A

D-GAT
A-GCA

Query is DPS wild type, Sbjct is the sequencing result

```

Query 1 ATGAGCACC GCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT 60
      |||
Sbjct 31 ATGAGCACC GCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT 90

Query 61 GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGTAACCGTCAGGTGATCCAGTTC 120
      |||
Sbjct 91 GTGAGCGATAGCGAAAAAAGCAACCGTGGAAGTCTGTAACCGTCAGGTGATCCAGTTC 150

Query 121 ATCGATCTGAGCCTGATCACC AAACAGGCACATTGGAACATGCGTGGTGCAAATTCATC 180
      |||
Sbjct 151 ATCGATCTGAGCCTGATCACC AAACAGGCACATTGGAACATGCGTGGTGCAAATTCATC 210

Query 181 GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG 240
      |||
Sbjct 211 GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG 270

Query 241 GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC 300
      |||
Sbjct 271 GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC 330

Query 301 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 360
      |||
Sbjct 331 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 390

Query 361 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 420
      |||
Sbjct 391 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 450

Query 421 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 480
      |||
Sbjct 451 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 510

Query 481 TTCATCGAATGCAACATCGAA 501
      |||
Sbjct 511 TTCATCGAATGCAACATCGAA 531

```

DPS W160A

W-TGG
A-GCA

Query is DPS wild type, Sbjct is the sequencing result

```

Query   1   ATGAGCACCACGAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   30   ATGAGCACCACGAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT   89

Query   61   GTGAGCGATAGCGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   90   GTGAGCGATAGCGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC   149

Query   121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   150  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC   209

Query   181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   210  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG   269

Query   241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   270  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC   329

Query   301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   330  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA   389

Query   361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   390  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA   449

Query   421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   450  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGGCA   509

Query   481  TTCATCGAATGCAACATCGAA   501
          ||||||||||||||||||
Sbjct   510  TTCATCGAATGCAACATCGAA   530

```

DPS R83A/R133A

R-CGT

A-GCA

Query is DPS wild type, Sbjct is the sequencing result

```

Query   1   ATGAGCACCACGAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   26   ATGAGCACCACGAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT   85

Query   61   GTGAGCGATAGCGaaaaaaaGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   86   GTGAGCGATAGCGAAAAAAGCAACCGTGGAACCTGCTGAACCGTCAGGTGATCCAGTTC   145

Query   121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   146  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACCTTCATC   205

Query   181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   206  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG   265

Query   241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC   300
          ||||||||||||||||||

```

```

      |||||
Sbjct 266 GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC 325

Query 301 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 360
      |||||
Sbjct 326 AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA 385

Query 361 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA 420
      |||||
Sbjct 386 CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGGCAAAGCAATCGGTGAAGCAAAA 445

Query 421 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 480
      |||||
Sbjct 446 GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG 505

Query 481 TTCATCGAATGCAACATCGAA 501
      |||||
Sbjct 506 TTCATCGAATGCAACATCGAA 526

```

Appendix 5. The primers used for four charged mutants at BFR two-fold axis

Primers used in QuikChange mutagenesis for BFR R30D, BFR D56R, BFR E60H, BFR D56RE60H using the BFR construct as a template.

R30D

R30D-F (5'-3') CCAGTACTTCCTGCATGCAGACATGTTCAAAAAGTGGGGTC
R30D-R (5'-3') GACCCAGTTTTTTGAACATGTCTGCATGCAGGAAGTACTGG

D56R

D56R-F (5'-3') CGATGAAATGAAACATGCACGTCGTTACATCGAACGTATCC
D56R-R (5'-3') GGATACGTTTCGATGTAACGACGTGCATGTTTCATTTTCATCG

E60H

E60H-F (5'-3') CATGCAGATCGTTACATCCACCGTATCCTGTTTCCTGGAAG
E60H-R (5'-3') CTTCCAGGAACAGGATACGGTGGATGTAACGATCTGCATG

D56R E60H

D56R E60H-F (5'-3') CATGCACGTCGTTACATCCACCGTATCCTGTTTCCTGGAAG
D56R E60H-R (5'-3') CTTCCAGGAACAGGATACGGTGGATGTAACGACGTGCATG

Appendix 6. Sequencing results of four charged mutants at BFR two-fold axis

R30D

R-CGT; D-GAC

Query is BFR wild type, Sbjct is the sequencing result

```

Query 5 AAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTGGCAA 64
      |||||
Sbjct 24 AAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTGGCAA 83

```

Query	65	TCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACCTGGGGTCTGAAACGTCTGAACG	124
Sbjct	84	TCAACCAGTACTTCCTGCATGCAGACATGTTCAAAAACCTGGGGTCTGAAACGTCTGAACG	143
Query	125	ATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAACGTA	184
Sbjct	144	ATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAACGTA	203
Query	185	TCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGAAG	244
Sbjct	204	TCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGAAG	263
Query	245	ATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAACCTGC	304
Sbjct	264	ATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAACCTGC	323
Query	305	GTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGATCG	364
Sbjct	324	GTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGATCG	383
Query	365	AAATCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTGATCC	424
Sbjct	384	AAATCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAACCGAACTGGATCTGATCC	443
Query	425	AGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	474
Sbjct	444	AGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	493

D56R

D-GAT; R-CGT

Query is BFR wild type, Sbjct is the sequencing result

Query	1	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG	60
Sbjct	21	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG	80
Query	61	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACCTGGGGTCTGAAACGTCTG	120
Sbjct	81	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACCTGGGGTCTGAAACGTCTG	140
Query	121	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA	180
Sbjct	141	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCACGTCGTTACATCGAA	200
Query	181	CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	240
Sbjct	201	CGTATCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	260
Query	241	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAAC	300
Sbjct	261	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACGGATGGTGCAAAAAAC	320
Query	301	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	360

Sbjct	321	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	380
Query	361	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG	420
Sbjct	381	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG	440
Query	421	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	474
Sbjct	441	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	494

E60H

E-GAA; H-CAC

Query is BFR wild type, Sbjct is the sequencing result

Query	1	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG	60
Sbjct	21	ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTG	80
Query	61	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAACTGGGGTCTGAAACGTCTG	120
Sbjct	81	GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAACTGGGGTCTGAAACGTCTG	140
Query	121	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA	180
Sbjct	141	AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCCAC	200
Query	181	CGTATCCTGTTCTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	240
Sbjct	201	CGTATCCTGTTCTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT	260
Query	241	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAAC	300
Sbjct	261	GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAAC	320
Query	301	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	360
Sbjct	321	CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG	380
Query	361	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG	420
Sbjct	381	ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG	440
Query	421	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	474
Sbjct	441	ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	494

D56RE60H

D-GAT; R-CGT; E-GAA; H-CAC

Query is BFR wild type, Sbjct is the sequencing result

Query	5	AAGGTGATACCAAAGTGATCAACTACCTGAACAACTGCTGGGTAACGAACTGGTGGCAA	64

Sbjct	24	AAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTGGCAA	83
Query	65	TCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGGTCTGAAACGTCTGAACG	124
Sbjct	84	TCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAGTGGGGTCTGAAACGTCTGAACG	143
Query	125	ATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAACGTA	184
Sbjct	144	ATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCCACCGTA	203
Query	185	TCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGAAG	244
Sbjct	204	TCCTGTTCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGTGAAG	263
Query	245	ATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAACCTGC	304
Sbjct	264	ATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAAGTGGATGGTGCAAAAAACCTGC	323
Query	305	GTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGATCG	364
Sbjct	324	GTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATGATCG	383
Query	365	AAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTGATCC	424
Sbjct	384	AAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTGATCC	443
Query	425	AGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	474
Sbjct	444	AGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT	493

Appendix 7. The calibration for the size exclusion chromatography column Sephadex 200 10/300 at different pH

Size exclusion chromatography calibration. Molecular weight correlation to Kav is defined in the experimental part (see Chapter 3).