BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2009,** *48,* 5623–5630 5623
DOI: 10.1021/bi900387t

A Helix Swapping Study of Two Protein Cages†

Rongli Fan,‡ Aimee L. Boyle,§ Vee Vee Cheong,‡ See Liang Ng,‖ and Brendan P. Orner*,‡

‡*Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371,* §*School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K., and* ‖*School of Biological Sciences, Nanyang Technological University, Singapore 637371*

Received March 7, 2009; Revised Manuscript Received April 19, 2009

ABSTRACT: Protein cages have been the focus of studies across multiple scientific disciplines. They have been used to deliver drugs, as templates for nanostructured materials, as substrates in the development of bio-orthogonal chemistry, and to restrict diffusion to study spatially confined reactions. Although their monomers fold into four-helix bundle structures, two cage proteins, DPS and BFR, self-assemble to form a 12-mer with tetrahedral symmetry and an octahedrally symmetric 24-mer, respectively. These monomers share strong similarities of both sequence and tertiary structure. However, they differ in the presence of a short additional helix. In BFR, the fifth helix is at the C-terminus and is positioned along the 4-fold symmetry axis, whereas with DPS, an extra helix helps to define the 2-fold axis in the cage and is located between the second and third helices in the monomer bundle. In an attempt to investigate if these short helices govern protein assembly, mutants were designed and produced that delete and swap these minidomains. All mutants form highly helical structures that unfold cooperatively as evidenced by thermal melting followed by circular dichroism. Dynamic light scattering, size exclusion chromatography, and sedimentation equilibrium experiments demonstrated that although many of the BFR mutants do not self-assemble and form lower-order complexes, many DPS mutants could form cages despite their unnatural design. Taken together, our data indicate that the BC helix is less important than the E helix for overall cage self-assembly, suggesting that dimerization may not play a role in nanostructure formation that is as key as previously assumed. Additionally, we found that fusing the minidomain from BFR onto DPS results in a mutant that assembles into a homogeneous population of a novel protein oligomer. This assembled cage while still formed from 12 subunits is larger in overall shape than that of the native protein.

Directly or indirectly, investigations into the structure of biomacromolecules have resulted in ~25% of the chemistry Nobel Prizes in the past 50 years (*1*). Despite these accomplishments and this level of enquiry, a complete understanding of protein folding remains an unsolved problem and one of the "big unanswered scientific questions" (*2*). The field has advanced enough that it is now possible to design proteins *de novo* (*3*). However, the majority of the resulting designed proteins have been small and monomeric. Surprisingly, considering that complex biostructural information is often built up by the spontaneous assembly of subunits, the rational design of multimeric protein structures has only recently begun to be studied (*4*−*7*). In addition, the fundamentals of self-assembly have been explored on the mesoscale (*8*) and as a pillar of supramolecular chemistry (*9*).

Protein cages, such as those in the ferritin family, can self-assemble into multisubunit, hollow, nanoscale structures (*10*). These proteins have been the focus of much recent attention as part of bioorthogonal methodology development (*11*), as part of drug delivery studies (*12*, *13*), and as platforms for nanostructured materials (*14*, *15*). Structural studies have revealed that many members of the ferritin family can self-assemble into nanocapsules of two types: maxiferritins form hollow spheres composed of 24 monomers, and miniferritins, ferritin-like proteins, are composed of 12 monomers. Because of the current interest in their application and the fact that they are composed of monomers folded into a four-helix bundle motif, a fold well studied by the *de novo* design community, ferritins could act as an important model system for developing the fundamentals of how to expand the dimensionality of rational protein engineering.

Ferritins are ubiquitously found in both prokaryotes and eukaryotes (*16*, *17*), and this wide distribution in nearly all organisms reflects the essential role sequestering iron plays in the cell (*18*, *19*). Due to the universal importance of the ferritins, they have been extensively studied, and many high-resolution structures of ferritins have been determined. This report focuses

†R.F. is supported by SPMS graduate scholarships. V.V.C. thanks the CBC undergraduate summer research fund, and S.L.N. thanks the URECA program for undergraduate research support. This paper fulfills SPMS's "18-month rule" for R.F. This research was supported by an SPMS start-up grant and a "Singapore Ministry of Education Academic Research Fund Tier 1 Grant" (RG 53/06).

*To whom correspondence should be addressed. Telephone: +65 6316 8757. Fax: +65 6791 1961. E-mail: orner@ntu.edu.sg.

 Published on Web 04/30/2009

on two such proteins. Bacterioferritin from *Escherichia coli* (BFR),[1] a maxiferritin (*20*), is composed of 24 identical subunits which self-assemble into a slightly distorted spherical cage with 432 octahedral symmetry. This cage has an exterior diameter of 12 nm and an interior cavity of 7.5 nm. Each monomer of BFR folds into a four-helix bundle motif with an additional short fifth helix ("E helix") at the C-terminus, which lies at an angle of 78° with respect to the bundle axis and points toward the core (Figure 1). The ferritin-like protein, DNA-binding protein from starved cells (DPS), has been shown to play an important role in protecting DNA from oxidative damage (*21*). Twelve DPS monomers self-assemble into a 32 symmetric tetrahedral, miniferritin protein cage with an exterior diameter of 9 nm and an interior cavity diameter of 4.5 nm. The subunits of DPS, like those of BFR, fold into four-helix bundles. DPS also has an additional fifth helix ("BC helix"), which unlike that of BFR, is arranged in a loop between the second and third helices of the bundle, running nearly orthogonal to the bundle axis, and is exposed on the outside of the cage (Figure 1) (*22*−*27*).

Why DPS and BFR form a miniferritin with tetrahedral symmetry and a maxiferritin with octahedral symmetry, respectively, when their monomers share similar four-helix bundle tertiary structure is still an open question. Although analyses have been proposed, what controls the nanoarchitecture of these proteins as well as what role the extra helices play in protein assembly is unknown. Grant et al. (*21*) have compared the crystal structures of BFR and DPS by structural alignment and demonstrated that the helices of the bundle are spaced in a remarkably similar manner (rmsd = 1.1 Å). The E helix (residues 146−151) from each of four BFR monomers join together to define the 4-fold axis of the protein cage, while in the DPS nanocapsule, the BC helix (residues 95−101) of one monomer interacts with the symmetry-related copy of another monomer's BC helix at the 2-fold axis. Some limited mutagenesis work on these ferritins has been performed with the goal of understanding their assembly. Andrews et al. engineered the C-terminal sequence of BFR with a 14-residue extension based on the λ peptide. Gel filtration studies demonstrated that although the wild type is a mixture of 24-mer and dimer, the mutant solely exists as a 24-mer, suggesting that the C-terminus, and perhaps the E helix (which is located at the C-terminus and defines the 4-fold octahedral axis, thereby distinguishing the 12-mer from the 24-mer), plays a role in controlling the oligomerization state (*28*). A study conducted by Santambrogio et al. involving the H-chain of the maxiferritin from humans focused on the mutagenesis of key residues positioned at symmetry axes. While only a small number of mutants were investigated, many of these mutations drastically altered the oligomerization state and stability of the proteins. Analysis of renaturation studies monitored by native PAGE resulted in the conclusion that mutants focused on the 2-fold axis had a more pronounced effect on 24-mer oligomerization than do the 4- and 3-fold axes. This conclusion leads to the proposed mechanism of assembly, for this system at least, which is initiated by dimerization followed by the aggregation of additional dimers to form the 24-mer (*29*).

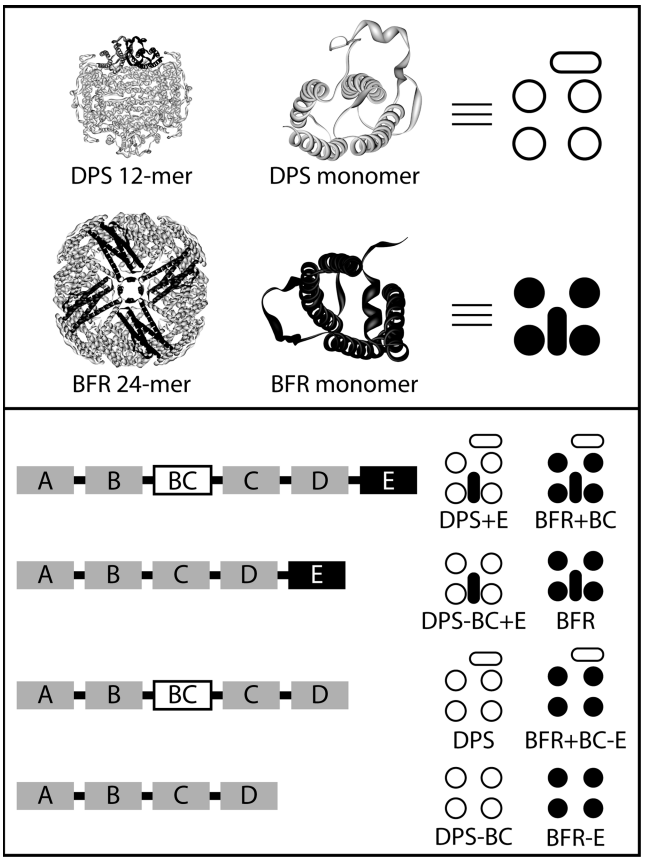

FIGURE 1: Crystal structures (top) of the DPS and BFR protein nanocages (Protein Data Bank entries 1DPS and 1BFR) with highlighted monomers, associated four-helix bundle monomers of which they consist (not to scale), and schematics indicating the A−D helices associated with the bundle (circles) and the additional BC and E helices (rounded rectangles) (white for DPS and black for BFR). Schematicized design (bottom) for proteins with added or deleted BC and E domains following the convention from above (see Figure S1 of the Supporting Information for full protein sequences).

We sought to clarify the role the E and BC helices play in protein assembly by swapping these domains between BFR and DPS. Although these are short helices (2−2.5 helical turns), we postulated that this swapping strategy would result in significant changes in stability and assembled structure since these helices are the major tertiary structural differences between these mini- and maxiferritins. Additionally, because the E helix is associated with the 4-fold axis in BFR and the BC helix of DPS with the 2-fold axis, switching these elements would also allow us to probe the importance of the symmetry-related interactions. Moreover, as we are not introducing any mutations within the four-helix bundles themselves [with the exception of some minor modifications within the D helix to better align the shorter DPS with BFR (details provided as Supporting Information)], it is thought that this experimental design would least likely disrupt the tertiary structure while providing an avenue for exploring assembly as independently as possible in a cooperatively folding multiple-subunit protein. We therefore designed six mutant proteins that either deleted (DPS−BC and BFR−E), added (DPS + E and BFR + BC), or replaced (DPS−BC + E, BFR + BC−E) these extra-bundle helices and set out to study their biophysical and self-assembly properties (Figure 1).

[1]Abbreviations: BFR, bacterioferritin from *E. coli*; DPS, DNA-binding protein from starved cells; PCR, polymerase chain reaction; Ek, enterokinase; LIC, ligation-independent cloning; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; EDTA, ethylenediaminetetraacetic acid; BCA, bicinchoninic acid; DTT, dithiothreitol; IPTG, isopropyl β-D-thiogalactopyranoside; SEC, size exclusion chromatography; CD, circular dichroism; TEM, transmission electron microscopy; DLS, dynamic light scattering; SE, sedimentation equilibrium; FPLC, fast protein liquid chromatography; rmsd, root-mean-square deviation.

MATERIALS AND METHODS

Protein and oligonucleotide sequences can be found in the Supporting Information along with schematics of gene cloning and protein purification strategies along with agarose and PAGE gels for all mutants for all steps (Figures S1−S15). Oligonucleotides were synthesized by solid-phase phosphoramidite technology by 1st Base Pte. Ltd. (Singapore). All concentrations quoted in molecular biological reactions are initial concentrations before dilution unless otherwise stated.

Production of Wild-Type and Mutant Ferritin Genes and Cloning. All the genes were synthesized using the PCR assembly approach described by Stemmer et al. (*30*). The individual steps in this method are described below.

Assembly PCR. The gene synthesis reactions were performed by adding the oligonucleotides (0.25 μL of each, 10 μM) to a solution consisting of a dNTP solution (Fermentas, 2.5 μL, dATP, dCTP, dTTP, and dGTP at 2 mM each), Pfu buffer (Fermentas, 2.5 μL, 10×), Pfu DNA polymerase (Fermentas, 0.5 μL, 2.5 units/μL), and deionized water to bring the total volume to 25 μL. The gene was constructed in a programmable thermal cycler (TaKaRa). The program consisted of 30 cycles with a denaturation stage (94 °C for 0.5 min), an annealing stage (55 °C for 0.5 min), and an extension stage (72 °C for 1 min). These 30 cycles were followed by a final extension step (72 °C for 5 min) (results presented in Figure S11).

Amplification PCR. The assembled gene was amplified by PCR. The unpurified assembled template (1 μL) was added to a solution of forward and reverse primers (these being the end primers from gene synthesis; 2 μL of each, 10 μM), dNTP solution (2.5 μL, dATP, dCTP, dTTP, and dGTP at 2 mM each), Pfu DNA polymerase (0.5 μL, 2.5 units/μL), Pfu buffer (2.5 μL, 10×), and deionized water (14.5 μL) to bring the total volume to 25 μL. The amplification program consisted of an initial denaturation step (94 °C for 5 min) followed by 30 cycles consisting of a denaturation stage (94 °C for 0.5 min), an annealing stage (55 °C for 0.5 min), and an extension stage (72 °C for 1 min). These 30 cycles were followed by a final extension step (72 °C for 5 min) (results presented in Figure S11).

Extension PCR. The DNA template (1 μL) was added to a solution of sense and antisense Ek/LIC primers (see section C of Figures S3−S10 for sequences; 2 μL of each, 10 μM), dNTP mix (2.5 μL, dATP, dCTP, dTTP, and dGTP at 2 mM each), Pfu buffer (2.5 μL, 10×), Pfu DNA polymerase (0.5 μL, 2.5 units/μL), and deionized water (14.5 μL) to bring the total volume to 25 μL. The PCR program consisted of an initial denaturation step (94 °C for 5 min) followed by 30 cycles consisting of a denaturation step (94 °C for 0.5 min), an annealing step (55 °C for 0.5 min), and an extension step (72 °C for 1 min). These 30 cycles were followed by a final extension step (55 °C for 5 min). The product was isolated by gel purification (Eppendorf Perfectprep Gel Cleanup kit) (results presented in Figure S12).

Preparation of Extended Single-Stranded Complementary Ends and Annealing to Ek/LIC Vector. The purified DNA product from the extension PCR (8 μL) was added to a solution consisting of dATP solution (Novagen, 2 μL, 25 mM), DTT (Novagen, 1 μL, 100 mM), T4 DNA polymerase buffer (Novagen, 2 μL, 10×), T4 DNA polymerase (Novagen, 0.4 μL, 2.5 units/μL), and deionized water (6.6 μL) to bring the total volume to 20 μL. The sample was incubated (22 °C for 30 min), after which the temperature was increased (75 °C for 30 min) to inactivate the enzyme. The gene insert (2 μL) was then added to a solution of deionized water (3.5 μL) containing the Ek/LIC vector (Novagen, 0.5 μL). The solution was incubated (22 °C for 5 min), and EDTA (Novagen Chemicals, 2 μL, 25 mM) was then added and the solution incubated for an additional 1 h.

The resulting vector construct was transformed by electroporation into electrocompetent Novablue cells (Novagen). The resulting carbenicillin-resistant colonies were screened for the recombinant gene by PCR (results presented in Figure S13) and dideoxy sequencing (1st Base) after miniprep (Qiagen).

Protein Expression. An overnight culture (3 mL) of BL21 (DE3) (Novagen) cells containing the expression construct (Novagen) was added to Luria broth (250 mL) supplemented with carbenicillin (Novagen, 250 μL, 50 mg/mL). The culture was incubated (150 rpm, 37 °C) until the OD_{600} reached ~0.4−0.6, at which time protein expression was induced (IPTG, Fermentas, 1 mL, 100 mM) and the culture incubated further (150 rpm, 30 °C, 3 h).

The cells were isolated by centrifugation, and the pellet, resuspended in equilibrium buffer [300 mM NaCl and 50 mM NaH_2PO_4 (USB Chemicals) at pH 7.0, 7.5 mL], was lysed by sonication (SONICS Vibra Cell sonicator, cycle of 3 s pulse at 40% amplitude followed by a 3 s rest repeated for a 4 min period). The lysis debris was cleared by centrifugation.

Protein Purification and Affinity Tag Cleavage. The His-tagged proteins were purified by incubation of the crude protein with metal affinity resin (TALON, Clontech) in batch mode as per the manufacturer's instructions. The proteins were eluted from the resin using affinity tag cleavage by the addition of enterokinase (New England Biolabs, 8 μL, 2 μg/mL), and the solution was incubated with gentle inversion (4 °C for 48 h). The protein was then concentrated, and the buffer was exchanged using ultrafiltration (Centricon). A 12% SDS−PAGE gel was run to verify the purity of the proteins (results presented in Figure S15).

On Resin Refolding. After induction and cell isolation, the bacterial pellet was resuspended in lysis buffer (100 mM NaH_2PO_4, 10 mM Tris, and 8 M urea at pH 8.0, 4 mL). The solution was centrifuged (10000 rpm, 4 °C, 20 min) and the pellet discarded. Resin (1 mL) was added to the supernatant and incubated with gentle inversion (1 h). The supernatant was then discarded and the resin washed with buffer [100 mM NaH_2PO_4, 10 mM Tris-HCl, and 8 M urea at pH 6.3 (for BFR + BC−E) or pH 6.7 (for DPS−BC + E)], to elute any nonspecifically bound protein. The resin was then washed with buffers containing decreasing amounts of urea (6, 4, 2, 1, and 0.5 M) to refold the protein. The buffer was then replaced with enterokinase buffer (20 mM Tris, 50 mM NaCl, and 2 mM $CaCl_2$ at pH 8.0, 2 mL), enterokinase (2 μL) added, and the solution incubated with gentle inversion (4 °C for 48 h) (results presented in Figure S15).

Temperature-Dependent CD Analysis. A purified protein solution was diluted with phosphate buffer (50 mM NaH_2PO_4 at pH 7.2) and then filtered (0.2 μm). The concentration of the protein was determined (BCA, Novagen) and diluted to 100 μg/mL.

CD analysis was performed in 1 mm quartz cells, using a Jasco J-710 spectropolarimeter fitted with a Peltier temperature controller. Spectra were recorded from 250 to 200 nm at 5 °C intervals over a temperature range of 20−90 °C with an equilibration time at each temperature of 10 min. At least three replicates were performed (results presented in Figure S16). For each experiment, the signal at 222 nm was plotted versus temperature and fitted to the equation $CD = CD_{fold} + (CD_{unfold} - CD_{fold})/[1 + (T/T_m)^B]$, where CD_{fold} is the signal of

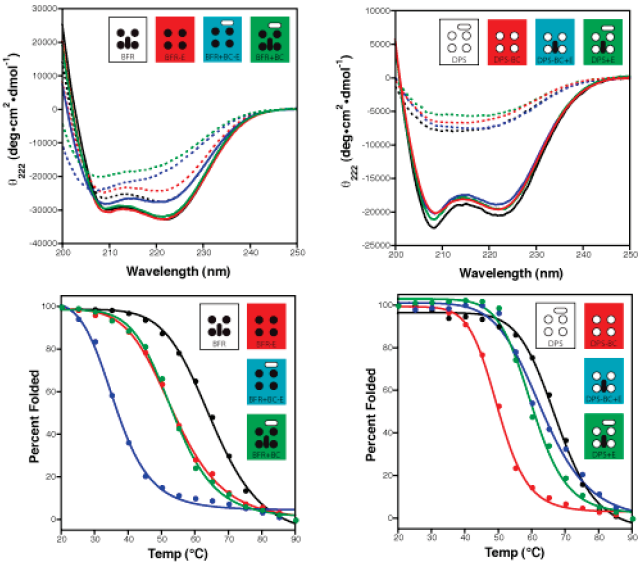

FIGURE 2: Circular dichroism (CD) spectra (top) of the cage proteins and mutants at 20 °C before thermal denaturation (solid lines) and at 20 °C after thermal denaturation (dotted lines) to indicate the level of unfolding reversibility. CD signals (bottom) of the cage proteins and mutants at 222 nm during thermal denaturation. Full scans of the denaturation experiments at each temperature point can be found in Figure S16, actual melting point values in Figure S17, and individual protein reversibility comparisons in Figure S18, both of which are summarized in Table 1. All data are the average of at least three experiments. The schematic convention follows that of Figure 1.

the fully folded protein, CD_{unfold} is the signal of the fully unfolded protein, T is the temperature, T_m is the melting point, and B is a fitting parameter which gives a sense of the cooperativity of the system (Figure S17 and Figure 2). After the proteins had melted, the solutions were cooled slowly to 20 °C over 30 min, and the spectra were compared to those obtained by the unfolding experiment (Figure S18 and Figure 2).

Dynamic Light Scattering (DLS). The concentration of the protein was determined (BCA, Novagen), and the solution was diluted to 100 μg/mL with phosphate buffer. DLS analysis was performed using a Brookhaven 90 Plus Particle Sizer and a 1 cm path length quartz cell. The data set was collected for 3 min, and three data sets were collected for each protein. An average hydrodynamic diameter was calculated from at least three replicates (Figure 4).

Size Exclusion Chromatography. Size exclusion was performed on a GE Akta FPLC system with a Superdex 200 10/300 GL gel filtration column at a flow rate of 0.5 mL/min (running buffer, 50 mM phosphate and 0.15 M NaCl at pH 7.0). Calibration was performed by injection of blue dextran, thyroglobin, horse spleen ferritin, aldolase, conalbumin, and ovalbumin (GE Biosystems). The elution volumes (V_e, average of two runs) were used to determine the gel-phase distribution constant (K_{av}) by the relationship $K_{av} = (V_e - V_o)/(V_c - V_o)$, where V_o is the void volume (assumed to be the elution volume of the blue dextran) and V_c is the column volume (24 mL). These K_{av} values were correlated to the Stokes radii of the proteins (Figure S19). Conalbumin was excluded from the Stokes radii correlation due to the fact that its radius is unknown (GE Biosystems Calibration Kit Manual).

Transmission Electron Microscopy. The solution of the purified protein (10 μL, 20−100 μg/mL) was dropped onto a sheet of parafilm; the grid (Formvar/carbon, copper 300 mesh) was placed on the drop (1 min), and filter paper was used to wick

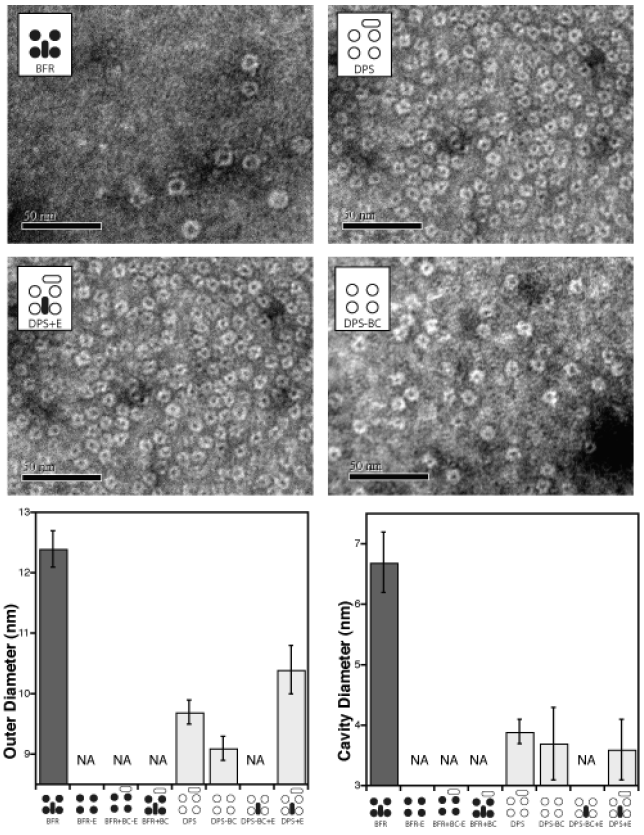

FIGURE 3: Representative negatively stained TEM micrographs of protein cages formed by BFR, DPS, DPS + E, and DPS−BC (top four panels) with outer (bottom left) and cavity (bottom right) diameters of the particles from analysis of micrographs. Additional micrographs can be found in Figure S19. The schematic convention follows that of Figure 1.

away excess solution. The grid was then placed on a solution of uranyl acetate stain [7 μL, 1% (w/v), 1 min]. The grid was dried with filter paper and then air-dried in a closed box.

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

Particles sizes were measured from the micrographs using Image J (National Institutes of Health, Bethesda, MD). A circle was drawn around the particle, and the diameter of this circle was calculated. At least 50 particles were measured from each image, except in the cases where there were fewer than 50 particles in a particular image, in which case the diameter of every particle in the image was measured (Figure S20 and Figure 3)

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

RESULTS

Protein Design. Inspection of the BFR and DPS four-helix bundle monomer structures reveals that the two primarily differ in the placement of an additional fifth helix: DPS has a "BC helix" between the second and third helices of the bundle, and BFR has an "E helix" at its C-terminus. The BC helix is found at the interface between two dimers near the 2-fold axis in DPS, whereas the E helix is located at the tetramerization interface and is positioned near the 4-fold symmetry axis in BFR. The goal of this study is to determine what role the E and BC helices play in BFR and DPS, respectively, by determining changes in monomer structure, thermodynamic stability, folding reversibility, size of the oligomer, and oligomerization number when these minihelices are swapped between the two proteins. We therefore designed six mutant proteins that explore all possible permutations. The proteins that were studied were DPS−BC and BFR−E which have neither of the extra-bundle helices, DPS and BFR + BC−E which possess only the BC helix, DPS−BC + E and BFR which incorporate only the E helix, and DPS + E and BFR + BC, each of which has both BC and E helices (Figure 1 and Figure S1).

Cloning and Protein Expression and Purification. The proteins were produced through bacterial overexpression followed by affinity purification. The genes for expression were generated through the assembly of synthetic oligonucleotides by means of a double-PCR procedure (*30*) (Figures S2−S13). This method proved to be general and high-yielding and should be amenable to eventual higher-throughput production. These genes were conjugated to a pET-32 Ek/LIC expression vector (Novagen). There are several advantages of the use of this expression vector. (a) It uses ligation-independent cloning which is high-yielding and therefore could prove useful in the construction of libraries. (b) It provides proteins fused to a series of affinity tags to simplify purification. (c) These tags can be removed through cleavage at an enterokinase cleavage site. (d) This site 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 (Figures S14 and S15). The benefit of this strategy is that because the monomer is bound to the resin and only self-assembles upon release (the large affinity tags presumably prevent self-assembly), it would conceivably be relatively simple to load for drug delivery studies. This procedure appeared to be general in that it worked well for all the proteins with the exception of BFR + BC−E, in which after overexpression all the protein was taken up in insoluble inclusion bodies. After unsuccessfully attempting to optimize the expression of BFR + BC−E, we employed an on-resin refolding strategy in which the inclusion bodies were solubilized in urea and applied to the resin and the denatured protein was refolded by gradual buffer exchange into enterokinase buffer and cleaved as normal. This procedure may be general, with eventual optimization, as it also was successful for the purification of DPS−BC + E which was difficult to purify to homogeneity using the conventional process. We are currently optimizing this on-resin refolding method and exploring its further generality.

Circular Dichroism-Based Studies. Circular dichroism (CD) indicates that all the proteins fold into highly helical secondary structures and that mutation, as expected, does not strongly affect the overall helicities of the proteins. Interestingly, the CD spectra of the mutants at 20 °C resemble those of the parent proteins (Figure 2), suggesting that the structures of the mutants derived from BFR are more similar to BFR than they are to DPS and vice versa. As would be expected by making such gross, minidomain-based mutations without optimization for packing, the thermal stability of the mutants was depressed in all cases as evidenced by melting experiments following the CD band at 222 nm, although they all exhibit highly cooperative unfolding, suggesting well-packed structures. The maxiferritin, BFR, was more sensitive to mutation than DPS. For the BFR series, either the deletion of the E helix (BFR−E) or the addition of the BC helix (BFR + BC) resulted in an ∼11 °C depression in T_m, indicating that the E helix is essential for stability and that introduction of the BC helix disrupts the stability. This trend is reinforced by the observation that removal of the E helix and inclusion of the BC helix (i.e., BFR + BC−E) has the strongest effect on BFR stability, reducing the T_m by ∼27 °C. Taken together, these data demonstrate the importance of interactions across the 4-fold axis and imply that the introduction of additional interactions across the 2-fold symmetry axis is detrimental rather than advantageous as would be expected if dimerization drives the assembly. For the mutants derived from DPS, as expected, removal of the BC helix (DPS−BC) has the largest melting point depression (∼16 °C) presumably due to losing essential contacts across the 2-fold axis. However, addition of the E helix (DPS−BC + E) nearly fully recovers the stability, depressing the T_m by only ∼5 °C. The protein with both helices (DPS + E) preserves the trend with a T_m depression of ∼7 °C (Figure S17 and Table 1).

Table 1: Melting Temperatures and Reversibility of Proteins Used in This Study

protein	T_m (°C)	ΔT_m (°C)	reversibility (%)
BFR	65.3 ± 0.7		83
BFR−E	53.9 ± 0.2	11.4 ± 0.9	74
BFR + BC−E	38.1 ± 2.7	27.2 ± 3.4	67
BFR + BC	53.6 ± 0.2	11.7 ± 0.9	51
DPS	67.6 ± 0.3		35
DPS−BC	50.1 ± 0.8	17.5 ± 1.1	30
DPS−BC + E	63.4 ± 1.4	4.2 ± 1.7	38
DPS + E	60.3 ± 0.4	7.3 ± 0.7	27

Controlled refolding experiments result in an additional surprise. Although the DPS series is more thermally stable, slow cooling after denaturation indicated that DPS and its mutants were much less likely to return to their native fold than were the BFR mutants (Figure 2, Figure S18, and Table 1). These results imply that the DPS series possesses a larger kinetic barrier to refolding. Understanding the nature of this possible barrier warrants further investigation. Taken together, with the observation that BFR is more thermally sensitive to mutation, these data suggest that the mechanisms governing the folding of the maxiferritins are possibly dissimilar to those that control the miniferritins.

Transmission Electron Microscopy (TEM). With the knowledge that all the proteins form cooperatively folded, thermodynamically stable, helical structures, we explored their ability to assemble into hollow cages by transmission electron microscopy (TEM) using negative staining (Figure 3 and Figure S19). The electron micrographs of the native proteins BFR and DPS revealed structures similar in shape and size to those reported in the literature. Of the mutants, only the DPS

series formed any uniform structure that could be detected by TEM. This is consistent with CD data which demonstrated the greater sensitivity of BFR to mutagenesis relative to that of DPS. Image analysis of the DPS−BC micrographs indicates structures slightly smaller than DPS. Interestingly, DPS + E produced nanostructured particles of a size intermediate between those of BFR and DPS. As evidenced by the uptake of negative stain, the mutants formed hollow spheres. Although their outer diameters were of different sizes, the size of their inner cavities remains unchanged from that of the native DPS.

Dynamic Light Scattering (DLS) and Size Exclusion Chromatography (SEC). To gain further insight into the nature of the nanostructures and their relative populations in solution, and to determine the self-assembled state of the mutants that do not exhibit cage structures in the TEM, two solution-phase techniques were employed. Dynamic light scattering (DLS) supported the formation of particles ~13 and ~10 nm in diameter for BFR and DPS, respectively (Figure 4). These values are slightly larger than those published. However, DLS measures hydrodynamic radii, so the measured diameters would be expected to be larger than those obtained by TEM or AFM analysis. The BFR mutants (~1−3 nm) are much smaller than their parent, implying that those mutants lose protein cage structure, which is consistent with the TEM data. The DPS series formed larger particles than the BFR mutants and ranged in size from ~6 to ~13 nm, suggesting that some sort of oligomer was formed. Analysis of the mutants showed that DPS + E formed particles larger than the parent, whereas DPS−BC formed nanostructures that were smaller which is also consistent with the electron microscopy data.

Size exclusion chromatography (SEC) allows the exploration of both small and large oligomers, and differences between populations of small particles (such as monomer and dimer) can be clearly observed (Figure 4). SEC therefore can sharpen our perception of oligomerization state and therefore help to further refine the role the BC and E helices play in assembly of the protein cages. Size exclusion experiments, possibly because of the associated solvation sphere surrounding these proteins, suggested that BFR and DPS (BFR, ~13.0 nm; DPS, ~10.5 nm) were slightly larger than those compounds described in the literature (Figure 4). Consistent with the literature (*28*), the BFR 24-mer in our hands exists as a mixture with a much smaller oligomer which correlates to a dimer. Isolation of one of these peaks, followed by concentration, results in the reappearance of both peaks, suggesting that they are in equilibrium. None of the BFR mutants form the larger oligomer, again emphasizing the sensitivity of BFR to mutation. Removing the E helix from BFR (BFR−E) results in a protein that exists entirely in the dimeric state, further emphasizing the great importance of the E helix in controlling the assembly of the protein cage. Addition of the BC helix to both BFR (BFR + BC) and the E helix-deleted mutant (BFR + BC−E) results in mixtures of the dimer and a protein which correlates to the monomer. Light is shed on the sensitivity of BFR to mutagenesis through the realization that native BFR is an equilibrium mixture; this delicate balance may be easily tipped. The E helix may have near-complete control over this equilibrium as evidenced by the fact that BFR-E forms a homogeneous population of the lower-order complex. Since the BC helix exists at the 2-fold symmetry axis between dimers and no packing optimization was performed on the mutants, it stands to reason that addition of the BC helix affects the dimer−monomer transition.

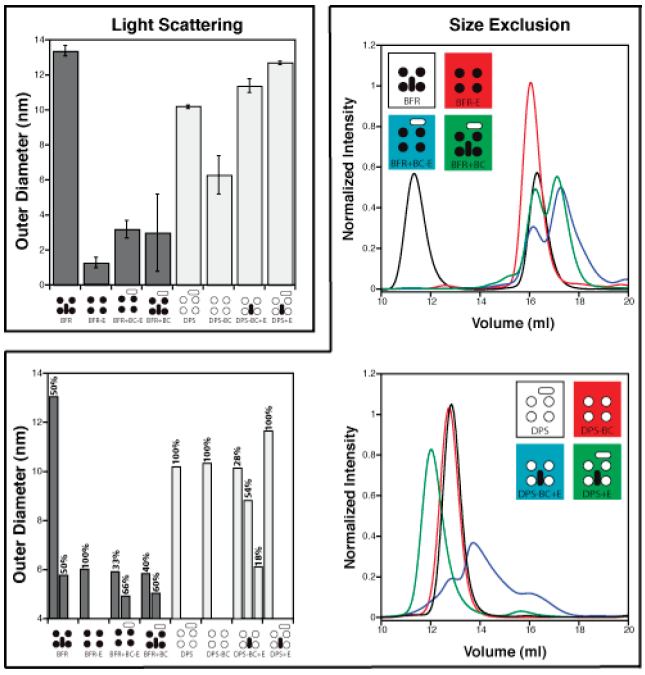

FIGURE 4: Dynamic light scattering (top left) and size exclusion chromatograms (right) of cage proteins and mutants. The chromatogram peaks were correlated to elution volumes of standard proteins (Figure S19) to determine Stokes diameters (bottom left). The chromatograms were normalized to 1, and the populations of particles of each diameter are indicated. The schematic convention follows that of Figure 1.

The DPS series is consistent with the conclusion based on the CD and DLS data that this protein's structure is less sensitive to mutation. Analysis of the DPS series also supports the hypothesis that the E helix has greater oligomerization power than the BC helix. Removal of the BC helix (DPS−BC) results in no observable change in the oligomerization state, whereas addition of the E helix to this mutant (DPS−BC + E) leads to a protein that exists as a complex mixture of states composed of the dimer, the natural 12-mer oligomerization state, and, intriguingly, a well-populated population (54%) of a state that is intermediate between the two (approximately hexamer). We are currently trying to determine the nature of this structure and whether it is possible to stabilize it through further mutation. Most striking of all is DPS + E which, consistent with the DLS data, forms a unique, homogeneous, and stable population (as evidenced by the melting data) of an oligomer intermediate in size (~11.5 nm) between BFR and DPS, supporting the notion that the E helix domain of BFR controls higher-order oligomerization.

Sedimentation Equilibrium. Because of the intriguing and unexpected TEM, DLS, and SEC results suggesting that DPS + E forms capsules that are intermediate in size between those of native DPS and BFR, we sought to determine whether this size difference is due to a different oligomerization state or to some alteration in the assembled conformation. While SEC can be used to determine difference in size of protein oligomers, sedimentation equilibrium (SE) is routinely used to determine differences in molecular mass (*31*). Analysis of the SE data for DPS and DPS + E suggests that DPS forms an oligomer with a molecular mass of 220 kDa, whereas the molecular mass of the DPS + E protein cage is 240 kDa (Figure 5). These data are consistent with the formation of a DPS + E multimer with a stoichiometry similar to that of the DPS 12-mer (note that the DPS + E monomer molecular mass is 19 kDa; the DPS

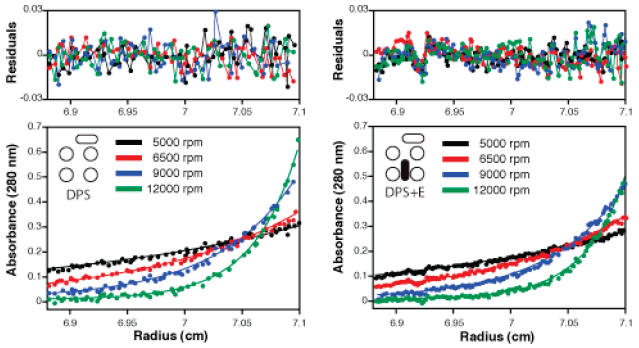

FIGURE 5: Sedimentation equilibrium for DPS (bottom left) and DPS + E (bottom right) at four angular velocities. The data (points) are best fit to 11.7-mer DPS and 11.5-mer DPS + E (lines), and the residuals of these fits are presented above the data traces.

monomer molecular mass is 21 kDa). Therefore, we conclude that DPS + E, instead of assembling into a unique and novel oligomerization state, forms a uniquely sized protein cage of the same order as the native protein. We are actively seeking a broader understanding regarding the origin of this alteration in size and whether it can be exploited in broadening the fundamental understanding of protein self-assembly, and whether it can be applied to the formation of novel nanomaterials.

DISCUSSION

The E helix from BFR and the BC helix from DPS were swapped between the two parent proteins, resulting in a series of mutants with all possible permutations. Although all mutants exhibited decreased thermostability compared to their parent, they all formed helical structures and unfolded cooperatively, suggesting tight, near-native folds. All BFR mutants exhibited a larger decrease in thermostability, suggesting a higher sensitivity of this protein to mutation compared to that of DPS. Although BFR is more thermodynamically sensitive to mutation, all members of the BFR series exhibited a higher degree of thermal folding reversibility than did the DPS series, implying that DPS is blocked kinetically from returning to the folded state. Of the mutants, only DPS + E and DPS − BC could form protein capsules as evidenced by TEM. Techniques for measuring the size of the oligomers that did form (DLS and SEC) demonstrated that none of the BFR mutants formed structures larger than a dimer, whereas all members of the DPS series exhibited some populations of higher-order oligomers. The weak oligomerization power of the BC helix was demonstrated by DPS − BC, which formed a nanostructure identical in size to the native DPS, and by BFR + BC and BFR + BC − E, both of which formed populations of the dimer smaller than that of BFR and BFR − E. However, the greater power exhibited by the E helix in controlling oligomerization is made evident by the removal of the E helix from BFR (BFR − E) which results in obliteration of the 24-mer to form a dimer as a homogeneous population. In addition, a comparison between the DPS − BC + E and DPS − BC mutants further demonstrates the high level of importance of the E helix in oligomerization, as addition of the E helix in the former resulted in a mixed population of 12-mer, dimer, and a majority of ~6-mer, while the latter with only the removal of the BC helix failed to demonstrate this. Finally, addition of the E helix to native DPS (DPS + E) dramatically emphasizes the importance of the E helix. The techniques of TEM, DLS, and SEC all show that this mutant forms a unique and homogeneous population of a protein that is intermediate in size between the BFR and DPS native proteins. The determination of its molecular mass using SE supports the running hypothesis that this protein is not self-assembling into an oligomer with a greater number of subunits but that is in fact just larger, most likely due to a change in assembly structure, monomer conformation, or both. Taken together, these data indicate that the E helix in BFR plays a stronger role in the formation of quaternary structure than does the BC helix in DPS and that stabilization of the dimer may not always be the most important self-assembly step in cage formation as was previously assumed.

We are actively seeking a broader understanding of these systems, attempting to further characterize some of the novel structures that were formed, and to apply these mutants to the templation of novel nanoscale materials. As a larger goal of this project, we are attempting to build more comprehensive systems to investigate other factors that determine protein cage self-assembly, such as grafting bundle surface residues, varying amino acid content, and changing helix length. Additionally, we hope that this research can provide preliminary data for the establishment of these cage proteins as model systems for exploring self-assembly and quaternary structure formation within a protein design context.

ACKNOWLEDGMENT

We thank Profs. Julien Lescar and Ahn Tuan Phan for use of their fast protein liquid chromatography systems, Ms. Ai Hua Seow for use of the CBC teaching lab CD, and the NUS School of Medicine for use of their TEM. We also thank Dr. Eric Underbakke and Prof. Hongyu Chen for cloning and DLS advice, respectively, and Alison Lester for critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Sequences of designed proteins, schematics of cloning and protein purification strategies, sequences of oligonucleotides used for gene synthesis, gels for each step in cloning and protein purification, CD melting and refolding data, SEC calibration, and TEM micrographs. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Seringhaus, M., and Gerstein, M. (2007) Chemistry Nobel rich in structure. *Science 315*, 40–41.
2. Editorial (2005) So much more to know. *Science 309*, 78–102.
3. Hill, R. B., Raleigh, D. P., Lombardi, A., and Degrado, N. F. (2000) De novo design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res. 33*, 745–754.
4. Padilla, J. E., Colovos, C., and Yeates, T. O. (2001) Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U.S.A. 98*, 2217–2221.
5. Papapostolou, D., Smith, A. M., Atkins, E. D. T., Oliver, S. J., Ryadnov, M. G., Serpell, L. C., and Woolfson, D. N. (2007) Engineering nanoscale order into a designed protein fiber. *Proc. Natl. Acad. Sci. U.S.A. 104*, 10853–10858.
6. Grueninger, D., Treiber, N., Ziegler, M. O. P., Koetter, J. W. A., Schulze, M. S., and Schulz, G. E. (2008) Designed protein-protein association. *Science 319*, 206–209.
7. Klug, A. (1999) The tobacco mosaic virus particle: Structure and assembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci. 354*, 531–535.
8. Whitesides, G. M., and Grzybowski, B. (2002) Self-assembly at all scales. *Science 295*, 2418–2421.

9. Lehn, J. M. (1988) Supramolecular chemistry: Scope and perspectives molecules, supermolecules, and molecular devices. *Angew. Chem., Int. Ed. 27*, 89–112.
10. Liu, X. F., and Theil, E. C. (2005) Ferritins: Dynamic management of biological iron and oxygen chemistry. *Acc. Chem. Res. 38*, 167–175.
11. Wang, Q., Lin, T. W., Tang, L., Johnson, J. E., and Finn, M. G. (2002) Icosahedral virus particles as addressable nanoscale building blocks. *Angew. Chem., Int. Ed. 41*, 459–462.
12. Destito, G., Yeh, R., Rae, C. S., Finn, M. G., and Manchester, M. (2007) Folic acid-mediated targeting of cowpea mosaic virus particles to tumor cells. *Chem. Biol. 14*, 1152–1162.
13. Ren, Y., Wong, S. M., and Lim, L. Y. (2007) Folic acid-conjugated protein cages of a plant virus: A novel delivery platform for doxorubicin. *Bioconjugate Chem. 18*, 836–843.
14. Meldrum, F. C., Wade, V. J., Nimmo, D. L., Heywood, B. R., and Mann, S. (1991) Synthesis of inorganic nanophase materials in supramolecular protein cages. *Nature 349*, 684–687.
15. Uchida, M., Klem, M. T., Allen, M., Suci, P., Flenniken, M., Gillitzer, E., Varpness, Z., Liepold, L. O., Young, M., and Douglas, T. (2007) Biological containers: Protein cages as multifunctional nanoplatforms. *Adv. Mater. 19*, 1025–1042.
16. Andrews, S. C., Smith, J. M. A., Yewdall, S. J., Guest, J. R., and Harrison, P. M. (1991) Bacterioferritins and ferritins are distantly related in evolution: Conservation of ferroxidase-center residues. *FEBS Lett. 293*, 164–168.
17. Harrison, P. M., and Arosio, P. (1996) Ferritins: Molecular properties, iron storage function and cellular regulation. *Biochim. Biophys. Acta 1275*, 161–203.
18. Arosio, P., and Levi, S. (2002) Ferritin, iron homeostasis, and oxidative damage. *Free Radical Biol. Med. 33*, 457–463.
19. Andrews, N. C. (2008) Forging a field: The golden age of iron biology. *Blood 112*, 219–230.
20. Dautant, A., Meyer, J. B., Yariv, J., Precigoux, G., Sweet, R. M., Kalb, A. J., and Frolow, F. (1998) Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from *E. coli*. *Acta Crystallogr. D54*, 16–24.
21. Grant, R. A., Filman, D. J., Finkel, S. E., Kolter, R., and Hogle, J. M. (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat. Struct. Biol. 5*, 294–303.
22. Roy, S., Gupta, S., Das, S., Sekar, D., Chatterji, D., and Vijayan, M. (2004) X-ray analysis of *Mycobacterium smegmatis* DPS and a comparative study involving other DPS and DPS-like molecules. *J. Mol. Biol. 339*, 1103–1113.
23. Ceci, P., Ilari, A., Falvo, E., Giangiacomo, L., and Chiancone, E. (2005) Reassessment of protein stability, DNA binding, and protection of *Mycobacterium smegmatis* DPS. *J. Biol. Chem. 280*, 34776–34785.
24. Bellapadrona, G., Chiaraluce, R., Consalvi, V., Ilari, A., Stefanini, S., and Chiancone, E. (2007) The mutations Lys 114-Gln and Asp 126-Asn disrupt an intersubunit salt bridge and convert *Listeria innocua* DPS into its natural mutant *Listeria monocytogenes* DPS. Effects on protein stability at low pH. *Proteins: Struct., Funct., Bioinf. 66*, 975–983.
25. Roy, S., Saraswathi, Gupta, S., Sekar, K., Chatterji, D., and Vijayan, M. (2007) Role of N and C-terminal tails in DNA binding and assembly in DPS: Structural studies of *Mycobacteriaum smegmatis* DPS deletion mutants. *J. Mol. Biol. 370*, 752–767.
26. Roy, S., Saraswathi, R., Chatterji, D., and Vijayan, M. (2008) Structural studies on the second *Mycobacterium smegmatis* DPS: Invariant and variable features of structure, assembly, and function. *J. Mol. Biol. 375*, 948–956.
27. Chodhury, R. P., Vihayabaskar, M. S., Vishveshwara, S., and Chatterji, D. (2008) Molecular mechanism of *in vitro* oligomerization of DPS from *Mycobacterium smegmatis*: Mutations of the residues identified by "interface cluster" analysis. *Biochemistry 47*, 11110–11117.
28. Andrews, S. C., Smith, J. M. A., Hawkins, C., Williams, J. M., Harrison, P. M., and Guest, J. R. (1993) Overproduction, purification and characterization of the bacterioferritin of *Escherichia coli* and a C-terminally extended variant. *Eur. J. Biochem. 213*, 329–338.
29. Santambrogio, P., Pinto, P., Levi, S., Cozzi, A., Rovida, E., Albertini, A., Artymiuk, P., Harrison, P. M., and Arosio, P. (1997) Effects of modifications near the 2-, 3- and 4-fold symmetry axes an human ferritin renaturation. *Biochem. J. 322*, 461–468.
30. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene 164*, 49–53.
31. Laue, T. M., and Rhodes, D. G. (1990) Determination of Size, Molecular-Weight, and Presence of Subunits. *Methods Enzymol. 182*, 566–587.