Electronic Supplementary Material (ESI) for Chemical Communications

Supporting Information

Complete Shift of Ferritin Oligomerization toward Nanocage Assembly through Rationally Engineered Protein-Protein Interactions

Maziar S. Ardejani[1,2], Xiao Ling Chok [2], Ce Jin Foo[2] and Brendan P. Orner*[,1]

[1] Department of Chemistry, School of Biomedical Sciences, King's College London, London, United Kingdom

[2] Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

*To whom correspondence should be addressed.
E-mail: brendan.orner@kcl.ac.uk.

Computational methods:

Analysis of the pocket at the EcBfr three-fold protein-protein interface. The PDB entry 2vxi of *Ec*Bfr was used for computational analysis. This structure had the highest resolution of *Ec*Bfr (1.9 Å) at the time the analyses were performed. An *Ec*Bfr trimer was created by deleting all except three protein chains centered around a three-fold symmetry axis (chains A, P and I of the 2vxi structure). Using the FoldX[1] computer algorithm, this trimer was first minimized (at 298K, pH 7, 0.05M ionic strength) in order to generate a reference structure for subsequent energy calculations and structure comparisons. The optimized trimer then was submitted to CASTp[2] for detection and analysis of the structural pockets (Fig. S1).

Free energy simulations of virtual mutants at the threefold pocket of EcBfr. Residues involved in the threefold pocket were chosen to undergo a virtual mutation analysis. The BuildModel command of FoldX 3.0 beta, which implements a probability-based rotamer library of its own, was used to explore the changes in unfolding free energies (relative to the WT) of the 120 (6 residues involved in threefold pocket x 20 amino acids) possible point mutants (at 298K, pH 7, 0.05M ionic strength). Calculations for each set of triple mutations were performed five times; energy minimization for each replicate was initiated from a different rotamer set.

For the flexible backbone modeling, the WT structure optimized by FoldX was used as input for the backrub module of Rosetta 3.1 in which Dunbrack's backbone-dependent rotamer library (2002)[3] is used.[4] The backrub sampling radius was set to 6 Å around the site of mutation and an ensemble of ten structures was generated for each mutant before their Rosetta Energy was calculated using Rosetta full-atom scoring function and averaged.

The energy differences between WT and mutants calculated either by FoldX or Rosetta were averaged. The results of the calculation for some of the most stable mutants are compiled in table S1.

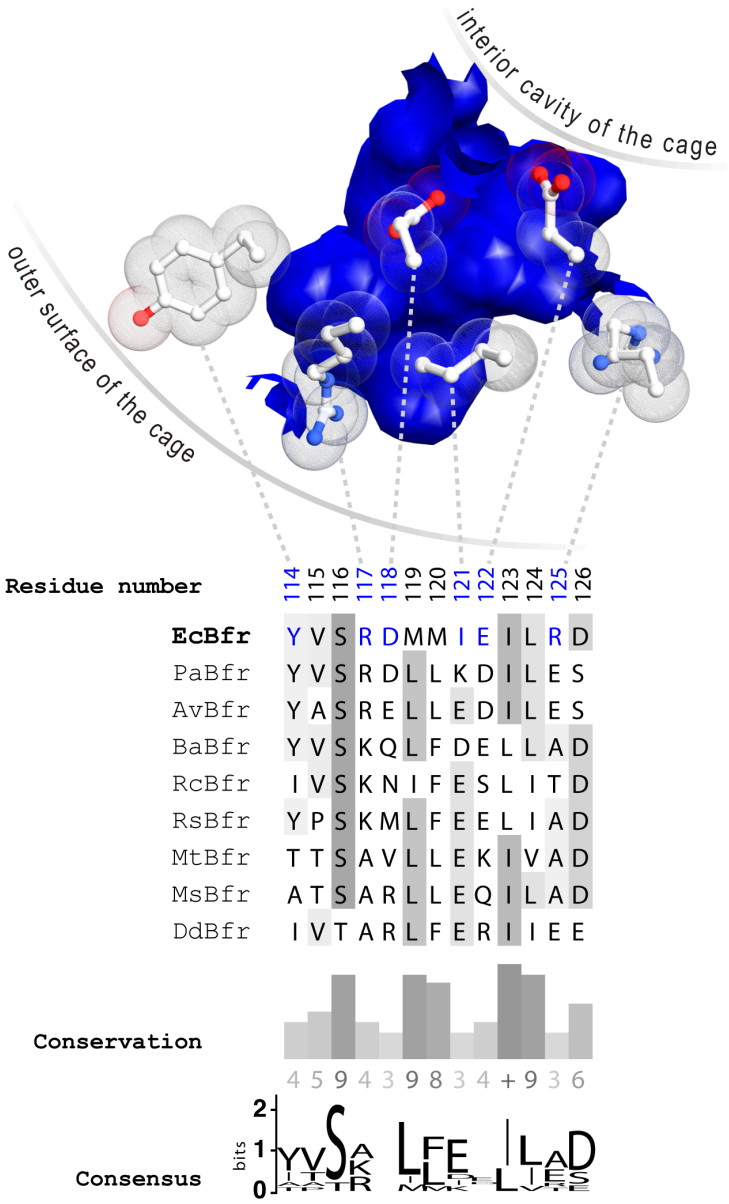

Fig. S1. Structure-based multiple sequence alignment of the subset of residues involved in the pocket of *Ec*Bfr at the C3 symmetrical trimerization interface. Six residues, Y114, R117, D118, I121, E122 and R125 (in blue font), of each *Ec*Bfr protein chain encase a pocket as determined by CASTp analysis. The blue surface represents the pocket volume. The *Ec*Bfr residues forming walls of the pocket are only weakly conserved among bacterioferritins with known crystal structures. Position 118 shows no clear preference for any type of amino acid. Bacterioferritins from the following bacteria with the accession and PDB codes respectively mentioned in parentheses used for the analyses: *Ec*Bfr from *Escherichia coli* (P0ABD3, 2vxi), *Pa*Bfr from *Pseudomonas aeruginosa* (Q9HY79, 3isf), *Av*Bfr from *Azotobacter vinelandii* (P22759, 2fl0), *Ba*Bfr from *Brucella abortus* (Q2YKI4, 3fvb), *Rc*Bfr from *Rhodobacter capsulatus* (Q59738, 1jgc), *Rs*Bfr from *Rhodobacter sphaeroides* (Q3J696, 3gvy), *Mt*Bfr from *Mycobacterium tuberculosis* (P63697, 2wtl), *Ms*Bfr from *Mycobacterium smegmatis* (A0QY79, 3bkn), *Dd*Bfr from *Desulfovibrio desulfuricans* (Q93PP9, 1nfv). These are the only bacterioferritins for which a crystal structure is available. In the sequence alignment, residues are shaded according to the percentage of the residues in each column that agree with the consensus sequence. Only the residues that agree with the consensus residue for each column are shaded. The darkest grey shade highlights residues that have percentage identity higher than 80% among the bacterioferritins and lightest grey shade indicates residues that only have a percentage identity between 40-50%. In the conservation histogram, regions of the alignment where physicochemical properties are conserved are shaded so that the columns with highest conservation index[5] in each group are colored with darkest grey, and the least conserved are the lightest grey. Graphs were prepared using Jalview.[6]

Table S1. Results of virtual mutation calculations of pocket forming residues at the C3 symmetrical trimerization interface of *Ec*Bfr ranked by their difference in free energy of unfolding versus WT (ΔΔG) as calculated using FoldX[1] and in Rosetta Energy Units (REU) as computed by Rosetta Backrub[4].

	Energy difference	
Mutation	**FoldX fixed backbone (kcal/mol)**	**Rosetta flexible backbone (REU)**
*Ec*Bfr **D118F**	-7.09	-142.49
*Ec*Bfr **D118L**	-4.00	-141.05
*Ec*Bfr **D118M**	-5.86	-139.37
*Ec*Bfr **D118W**	-6.45	-140.47
*Ec*Bfr **D118Y**	-3.20	-143.53

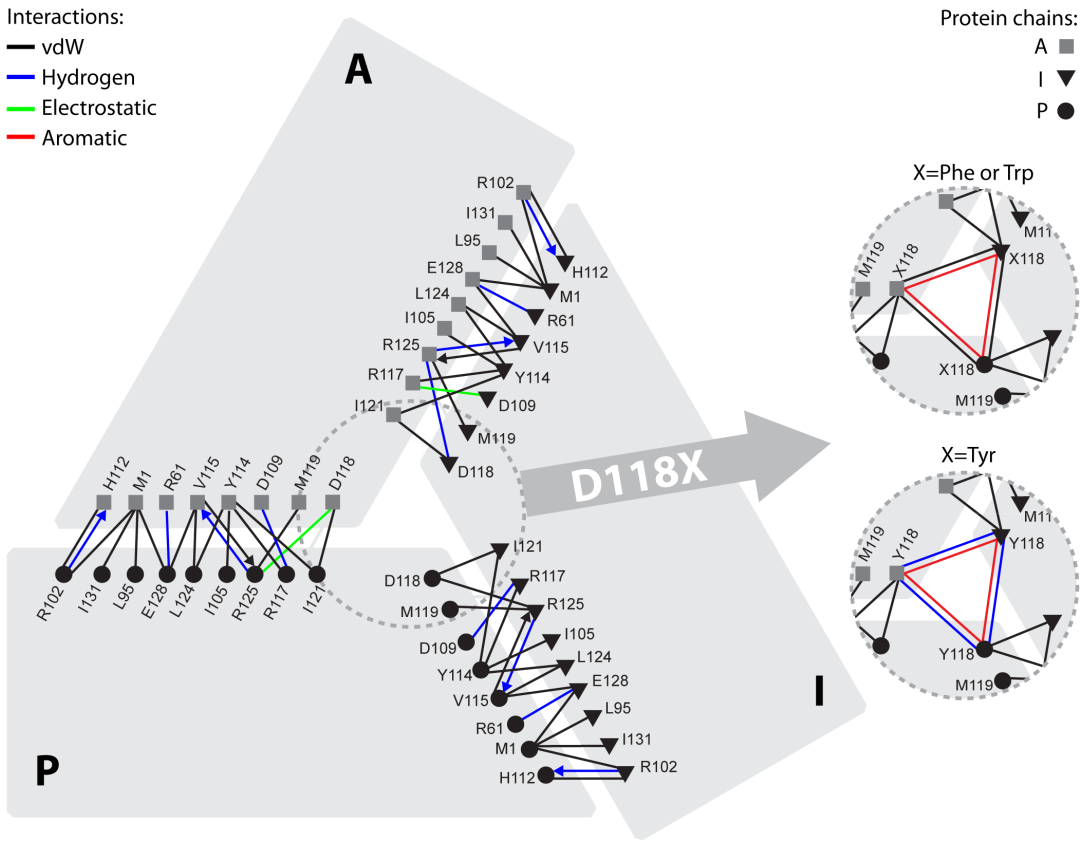

Fig. S2. The effect of D118 mutations on the interaction networks at the trimeric interface between chains **A**, **I**, and **P** of the *Ec*Bfr nanocage: wild type protein-protein interactions at threefold interface cluster into three separate groups with no detectable interaction between them. The aspartic acid at position 118 of the WT protein is involved in a vdW interaction with I121 from the neighboring protein chain. Intermolecular interaction of D118 with R125 however varies in different clusters between hydrogen bonding, ionic and vdW interactions depending upon the relative positioning of their side chains. As would be expected, these interactions all converge to vdW upon mutation of D118 to any of the aromatic amino acids. There is no interaction between Asp118 side-chains with each other in the WT structure, however D118X mutations (X=F, W and Y) build aromatic interactions that connect the three interfacial interaction clusters into one extended interaction network across the threefold interface. The D118Y mutation also forms C3-symmetrical aromatic interactions along with phenolic hydrogen bond which have been shown to make a significant contribution to protein stability.[7] *Key:* Side-chain–side-chain interactions appear as lines and backbone–side-chain interactions are depicted as arrows pointing from the backbone to the side-chain. Blue, green and black lines or arrows represent hydrogen bonding, salt bridges and van der Waals interactions respectively. Grey squares, black triangles and circles depict the residues from protein molecules **A**, **I** and **P** respectively. The contact map was generated using Aquaprot[8] and the minimized crystal structure of *Ec*Bfr symmetric trimer.

Construction of expression plasmids

The constructs were generated through site directed mutagenesis. Polymerase (*pfu* turbo, 2.0µl, 2.5units/µl) was added to the polymerase chain reaction (PCR) mixture (20µl, 1X *pfu* reaction buffer, 1µM forward primer, 1µM reverse primer, 0.2mM dNTP mix, 2µl pET-32 vector carrying WT *Ec*Bfr[9]) and the solution was subjected to a PCR cycle composed of a pre-heating step (95°C, 1 min), followed by 18 cycles of denaturation (95°C, 30 s, annealing (55°C, 1 min), and extension (68°C, 7 mins), followed by an elongation step (72°C, 10 mins).

The enzyme, *DpnI* (New England Biolabs, 0.5µl, 20units/µl) was added to the digestion mixture composed of PCR product (5µl) and NEB buffer 4 (New England Biolabs, 1µl, 10x) and ddH_20 (3.5µl) and the solution was incubated (37°C, 2 hours). After electroporation, the resulting transformants were confirmed by sequencing (Fig.s S3-S5).

```
Query  282  GATGACGACGACAAGATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTG  341
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    GATGACGACGACAAGATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTG  60

Query  342  GGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGG  401
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGG  120

Query  402  GGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCA  461
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  GGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCA  180

Query  462  GATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGT  521
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGT  240

Query  522  AAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTG  581
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  AAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTG  300

Query  582  GATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTG  641
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  GATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTG  360

Query  642  AGCCGTTTCATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAA  701
            ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  AGCCGTGATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAA  420

Query  702  ACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGT  761
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGT  480

Query  762  GAAGAAGGT  770
            |||||||||
Sbjct  481  GAAGAAGGT  489
```

Fig. S3. Sequencing results of the *Ec*Bfr D118F construct ("Query") aligned[10] to the wild type sequence ("Sbjct"). Red indicates the primer sequence used for site directed mutagenesis and blue indicates the mutated codon. The letters in green represent the DNA sequence coding for the enterokinase cleavage site.

```
Query  603   GATGACGACGACAAGATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTG  662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     GATGACGACGACAAGATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTG  60

Query  663   GGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGG  722
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    GGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGG  120

Query  723   GGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCA  782
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   GGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCA  180

Query  783   GATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGT  842
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   GATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGT  240

Query  843   AAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTG  902
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   AAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTG  300

Query  903   GATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTG  962
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   GATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTG  360

Query  963   AGCCGTTGGATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAA  1022
             ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361   AGCCGTGATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAA  420

Query  1023  ACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGT  1082
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   ACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGT  480

Query  1083  GAAGAAGGT  1091
             |||||||||
Sbjct  481   GAAGAAGGT  489
```

Fig. S4. Sequencing results of the *Ec*Bfr D118W construct ("Query") aligned[10] to the wild type sequence ("Sbjct"). Letters in red indicate the primer used for site directed mutagenesis and blue indicates the mutated codon. The letters in green represent the DNA sequence coding for the Enterokinase cleavage site.

```
Query  512  GATGACGACGACAAGATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTG  571
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    GATGACGACGACAAGATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTG  60

Query  572  GGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGG  631
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGG  120

Query  632  GGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCA  691
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  GGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCA  180

Query  692  GATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGT  751
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGT  240

Query  752  AAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTG  811
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  AAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTG  300

Query  812  GATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTG  871
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  GATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTG  360

Query  872  AGCCGTTATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAA  931
            ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  AGCCGTGATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAA  420

Query  932  ACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGT  991
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGT  480

Query  992  GAAGAAGGT  1000
            |||||||||
Sbjct  481  GAAGAAGGT  489
```

Fig. S5. Sequencing results of the *Ec*Bfr D118Y construct ("Query") aligned[10] to the wild type sequence ("Sbjct"). Letters in red indicate the primer used for site directed mutagenesis and blue indicates the mutated codon. The letters in green represent the DNA sequence coding for the Enterokinase cleavage site.

Gene Expression and Protein Purification

Plasmids from miniprep (Qiagen) with the correct mutated sequence (0.5µl) were electroporated (1.8kV, 1mm path length) into electrocompetent BL21 *E. coli* (75µl) and incubated immediately in LB broth (1ml) with shaking (170-200rpm, 1 hour). The culture was streaked into quadrants on a LB agar plate (50µg/ml carbenicillin) and incubated overnight at 37°C.

A single colony was picked and inoculated into LB growth medium (3ml, 50µg/ml carbenicillin) and incubated with shaking (170-200rpm, 37°C) overnight as a pre-culture. The antibiotic selected pre-culture (2ml) was then added to LB (500ml, 50µg/ml carbenicillin) and incubated with shaking (200rpm, 37°C, 2-3 hours) to allow for cell growth until the OD_{600} (optical density at 600nm) of approximately 0.4 was reached.

The culture was incubated with shaking (170-200rpm, 20°C) overnight (20hrs) after IPTG (200µl, 1M) was added for protein expression. The culture was pelleted (10,000rcf, 4°C, 20 mins) and then resuspended in lysis binding buffer (15ml, 10mM Tris.Cl, 150mM NaCl, 10mM imidazole, pH 8.0) by shaking for at least an hour on ice to achieve homogeneity before it was sonicated (60% amplitude, 5 mins). The cell lysate was separated into soluble and insoluble fractions by centrifugation (10,000rpm, 4°C, 20 mins). The supernatant was clarified by passing it through a syringe filter (0.2µm pore size).

Affinity chromatography. The soluble protein fraction was injected onto an affinity chromatography column (GE His-trap FF crude, 5ml, flow rate of 0.5ml/min), which was equilibrated with lysis binding buffer beforehand. Prior to elution, the column was washed with running buffer (1 volume of elution buffer to 20 volume of lysis binding buffer) to remove weakly bound proteins. The running buffer was then changed to elution buffer (20mM $NaH_2PO_4.H_2O$, 500mM NaCl, 500mM imidazole, pH 7.4) to elute all proteins specifically bound to the column. Eluate fractions (5ml) were collected at the beginning of the next peak.

The proteins were expressed with a number of fusion tags (Poly-Histidine, thioredoxin and s-tag) to permit affinity purification and to improve solubility. Once the tagged proteins were isolated on affinity resin, the fusion tags were to then be removed by enzymatic digestion. Digestion could be employed in two approaches: on resin and off resin. Initial digestion attempts were made on resin following the same procedure we had successfully used in our previous work.[9,11,12] Due to their steric bulk, these fusion tags presumably prevent cage formation when the tagged proteins are bound on the resin. When the tagged proteins are enzymatically digested, it was assumed that they then will self-assemble as they are released from the resin. While these assumptions worked for wildtype *Ec*Bfr and the destabilized mutants, the highly stabilized mutants behave differently with this approach. We hypothesize that the stabilizing mutations are encouraging formation of cages even if the tags are present. Therefore, when the tags are removed from the some of the monomers, they quickly associate to form protein cages with monomers that still have tags, and thus still associate with the resin and are not eluted. Digestion of the tags on the fully assembled, resin-bound cages is inhibited by steric hindrance and thus these partially tagged cages are cannot release from the resin. (Fig. S6) Hence, little or no cleaved protein was initially obtained after digestion. To optimize the enzymatic tag removal, the digestion was performed off-resin. When the tagged proteins are not bound to the affinity resin, it is presumed that all the fusion tags will be exposed to solution. Hence, digestion of these tags on the fully assembled cage will be less sterically hindered.

Tag Removal by Off Resin Enterokinase digestion. Prior to digestion, the eluate was equilibrated in digestion buffer (lysis binding buffer with 2mM $CaCl_2$) by ultra-filtration and concentrated to a volume of approximately 5ml. Enterokinase (NEB, 5µl, 2 µg/ml) was then added to the concentrated eluate to initiate the digestion. Digestion was carried out on a rotator (4°C, 40 hours). The cleaved tags and undigested protein were removed by a second affinity purification (GE His-trap HP, 5ml). Unbound proteins collected from affinity chromatography were concentrated to about 600µl via ultra-filtration.

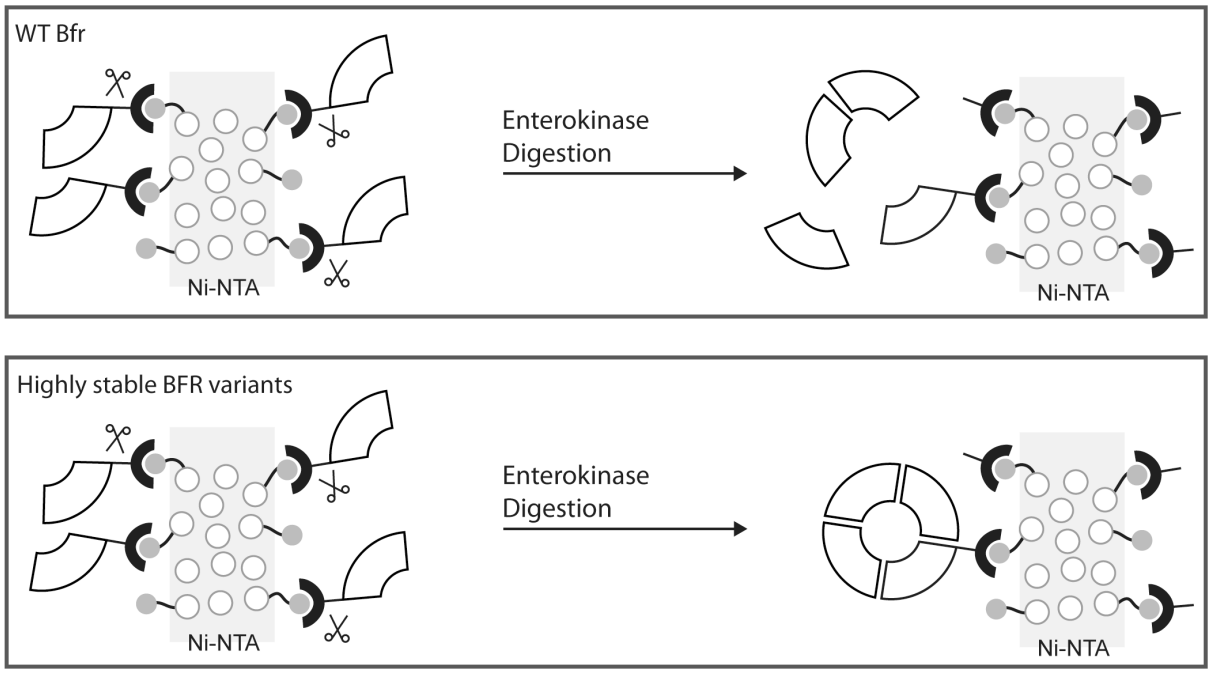

Fig. S6. Schematic diagram depicting postulated partial digestion in highly stabilized *Ec*Bfr mutants resulting in resin-associated, assembled cages that are resistant to further digestion by sterics, and thus resin release.

Size Exclusion Chromatography (SEC). The concentrated protein sample (600µl) was subjected to size exclusion chromatography. The column (Superdex 75, 24ml) was equilibrated with gel filtration chromatography (GFC) running buffer (50mM NaH_2PO_4, 150mM NaCl, pH 7). Protein samples were run in the same buffer at a flow rate of 0.5ml/min. Fractions (0.5ml) were collected from peaks at the retention volume corresponding to 24-mer (8ml) and dimer (10ml) (as mentioned above WT *Ec*Bfr exist in both oligomerization states). The purity of the proteins was determined by SDS-PAGE (Fig. S7). The identity of the protein monomers was confirmed by MALDI TOF mass spectrometry (Fig. S8 and Table S2).

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE). Each sample (10µl) was mixed with loading buffer (10µl, 5.6mg/100µl dithiothreitol (DTT)). The mixture was denatured (97°C, 10 mins) before loading (10µl) into the wells of the SDS-PAGE gel (4% stacking, 15% resolving). The gel was run at 80V (20 mins) and then at 140V (1 hr). The gel was stained with Coomassie Brillant Blue dye for visualization.

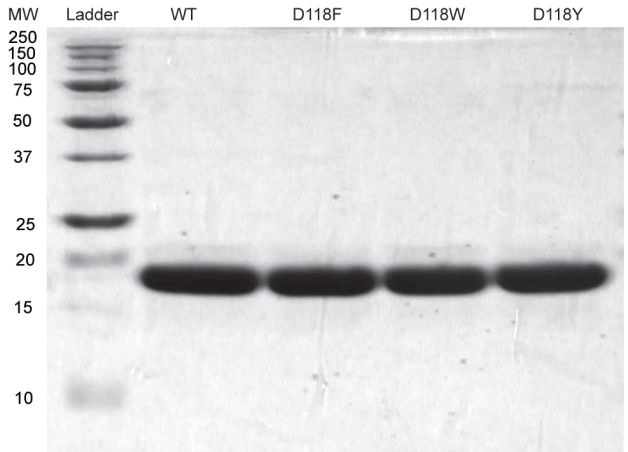

Fig. S7. The purity of WT *Ec*Bfr and D118X mutants was determined by SDS-PAGE electrophoresis. About 10µg of each protein (18.5 KD) was loaded into each well. The ladder lane is the protein molecular weight standard (Biorad, Precision Plus Protein, 10µl) and is in kDa.

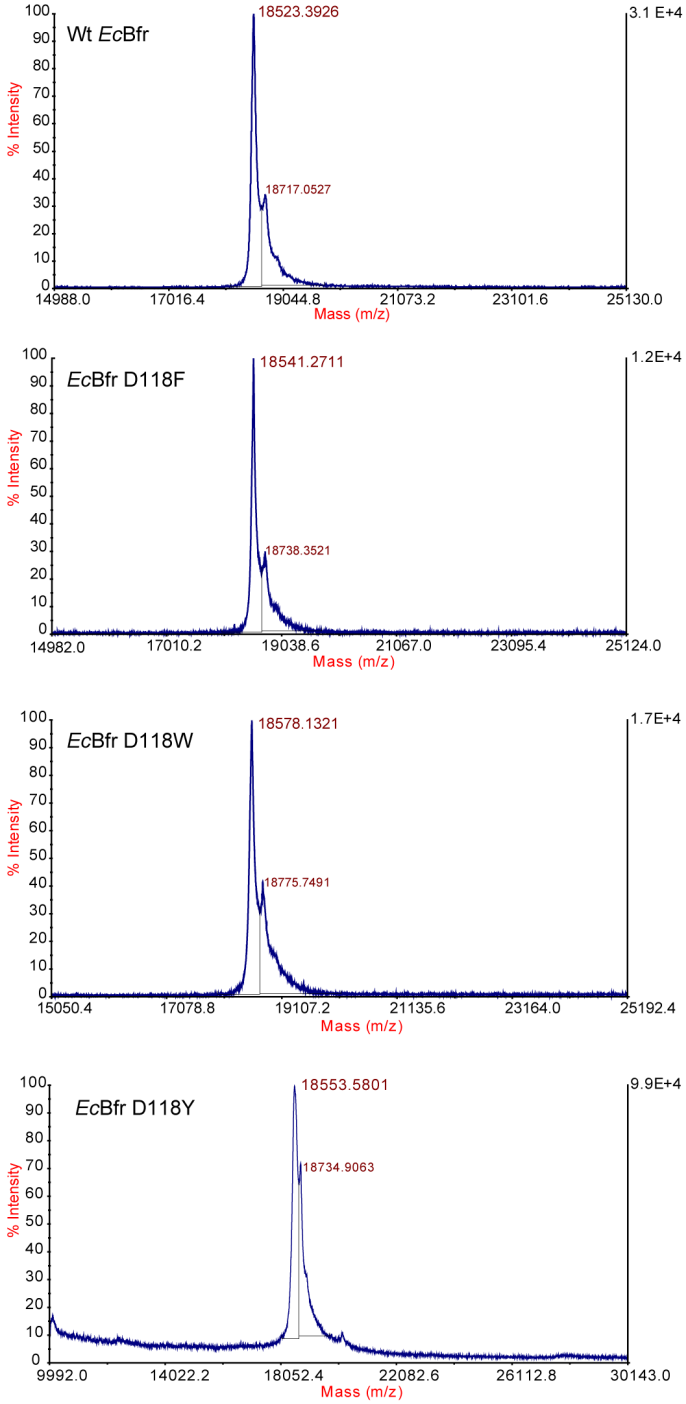

Fig. S8. MALDI-TOF mass spectrographs of purified monomeric proteins.

Table S2. The molecular weight of WT and mutated proteins as determined by MALDI-TOF MS. Theoretical molecular weights of the proteins were calculated using ExPASy server.[13]

Protein	MW (calculated)	MW (MS)	Percentage error*
WT *Ec*Bfr	18495.03	18523.39	0.15
*Ec*Bfr **D118F**	18527.12	18541.27	0.08
*Ec*Bfr **D118W**	18566.15	18578.13	0.06
*Ec*Bfr **D118Y**	18543.12	18553.58	0.06

*0.1 percent error is the normal error for the accuracy of protein mass measured by MALDI-TOF[14]

Biophysical characterization:

Analytical SEC. The SEC column (GE, Superdex 200 10/300 GL) was calibrated using six proteins as standards (GE Biosystems Calibration Kit).[12] Analytical SEC was performed (AktaFPLC, GFC buffer flow rate of 0.5 mL/min, 4°C) at least five times for each protein using protein samples (20μl, 300 μg/mL) equilibrated in GFC buffer (3 days, 4°C).

Native PAGE Electrophoresis. Each sample (10μl, 300 μg/ml) was mixed with native gel buffer (20μl) before loading (30μl) into the wells of the native PAGE gel (4% stacking, 7% resolving). The gel was run at 80V (20 mins) and then at 140V (1 hr). The gel was stained overnight with Coomassie Brillant Blue dye for visualisation.

Transmission Electron Microscopy (TEM). Purified protein (10 μl, 100 μg/ml) was prepared in phosphate buffer (15.67 mM NaH_2PO_4, 34.33 mM K_2HPO_4, pH 7.2) and dropped on a piece of Parafilm. A circular carbon-coated copper grid was gently placed onto the drop and left for one minute. The excess solution was then removed by wicking with a piece of filter paper. Once the surface of the grid was dry, it was placed onto a droplet of uranyl acetate solution (10μl, 1% w/v) for one minute. The excess stain was then removed by wicking and the grid was left to dry. TEM data was obtained using a Jeol JEM-1400 transmission electron microscope operating at 100keV. TEM micrographs were analyzed using ImageJ (NIH) to quantify the inner and outer diameters of the cages.[15] At least 100 measurements were performed for each variant (Fig. S9A, B, C, and D).

Dynamic Light Scattering (DLS). A solution of the freshly purified protein (500 μl, 300 μg/ml) was prepared in phosphate buffer (50 mM NaH_2PO_4, 150 mM NaCl, pH 7.0). DLS analysis was performed using a Brookhaven 90Plus Particle Size Analyzer and a quartz cell (1 cm path length). Each data set was collected over 3 min analysis, and five data sets were collected for each protein. The average hydrodynamic diameter of the protein particles (Fig. 2A) was calculated.

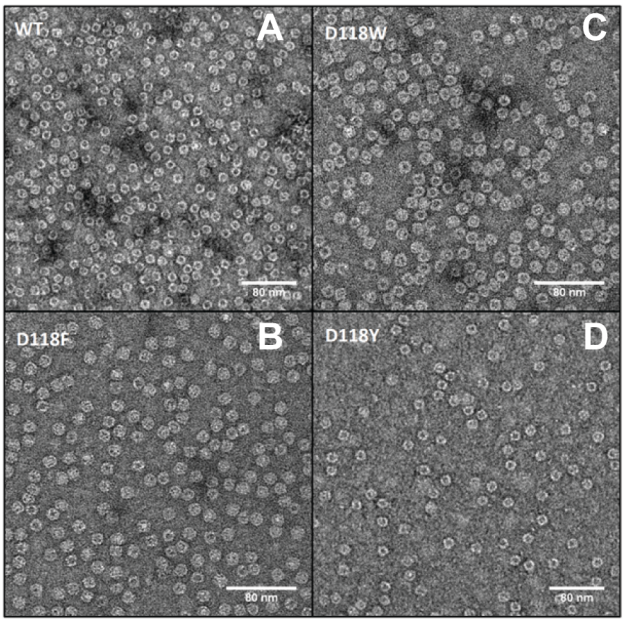

Fig. S9. TEM micrographs of WT (A) and the redesigned variants (B, C and D) were analyzed to determine cage and cavity size of the proteins (Fig. 2A of the main text).

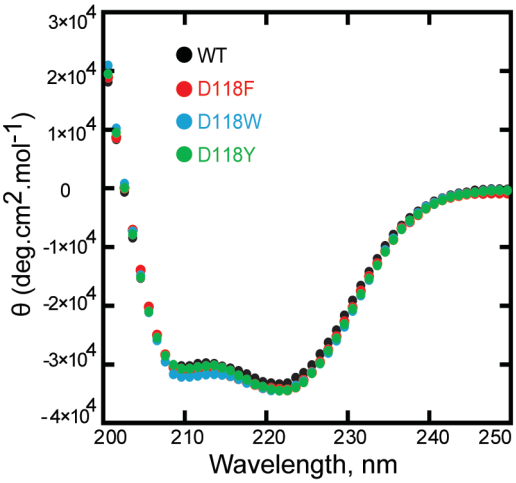

Fig. S10. The designed mutations do not alter the tertiary structure. Circular dichroism spectra of the WT and D118X mutants at 20 °C show that the redesigned proteins possess a native-like fold. The color-convention follows that of Fig. 1 of the main text.

Variable Temperature Circular Dichroism (CD). Pure protein fractions from SEC were equilibrated and concentrated in phosphate buffer (15.67mM NaH_2PO_4, 34.33mM K_2HPO_4, pH 7.2) and protein concentrations were determined by BCA colorimetry (Novagen). The concentrated sample (400µl, 300µg/ml) was then transferred to a quartz cuvette (1mm path length) and sealed with a Teflon cap and parafilm. Spectra were recorded using a JASCO J-810 spectropolarimeter from 250nm to 200nm (1 nm bandwidth, three accumulations) at temperatures ranging from 20°C to 110°C at 5°C intervals (470s equilibration time at each temperature point and heated at a rate of 2°C/min). Melting curves at 222nm were obtained from these scans. A final spectrum was collected after the sample was cooled to 20°C over 30 min to determine thermal refoldability.

The melting curves were then fit using KaleidaGraph software assuming a two state unfolding transition:

$$CD_{222} = m1 + \frac{m2 - m1}{1 + \left(\frac{m0}{m3}\right)^{m4}}$$

where, *m0* is the temperature variable, *m1* is the ellipticity value for the 0% folded plateau and *m2* is that for 100% folded plateau, *m3* is the T_m and *m4* gives an approximation of how cooperatively the protein unfolds.

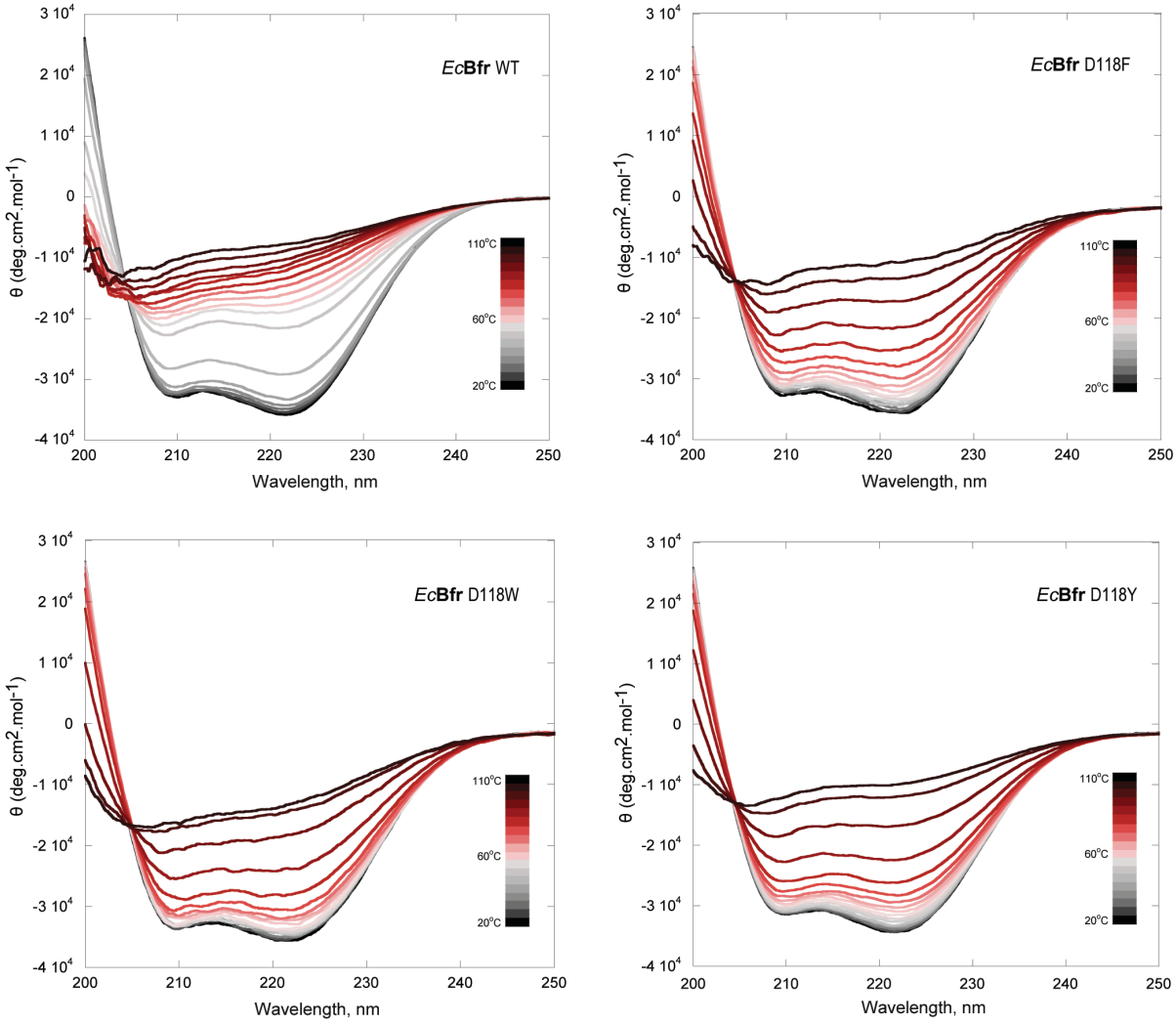

Fig. S11. The D118X mutants fold into alpha helical structure and this structure is temperature dependent. This data was used to construct the 222nm melting curves in Fig. 3 of the main text.

Table S3. Thermodynamic and structural analysis for the WT and redesigned protein variants

Variant	Pocket volume (Å^3)	Tm (°C)	ΔTm (°C)
Wild type	100.2	58.9±1.3	-
D118F	15.5	83.6±1.3	24.7
D118W	4.4	85.2±1.1	26.3
D118Y	16.3	82.0±1.7	23.1

References:

(1) Guerois, R.; Nielsen, J. E.; Serrano, L. *Journal of Molecular Biology* **2002**, *320*, 369.

(2) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. *Nucleic Acids Research* **2006**, *34*, W116.

(3) Dunbrack Jr, R. L. *Current Opinion in Structural Biology* **2002**, *12*, 431.

(4) Smith, C. A.; Kortemme, T. *Journal of Molecular Biology* **2008**, *380*, 742.

(5) Livingstone, C. D.; Barton, G. J. *Computer applications in the biosciences : CABIOS* **1993**, *9*, 745.

(6) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. *Bioinformatics* **2009**, *25*, 1189.

(7) Pace, C. N.; Horn, G.; Hebert, E. J.; Bechert, J.; Shaw, K.; Urbanikova, L.; Scholtz, J. M.; Sevcik, J. *Journal of Molecular Biology* **2001**, *312*, 393.

(8) Reichmann, D.; Cohen, M.; Abramovich, R.; Dym, O.; Lim, D.; Strynadka, N. C. J.; Schreiber, G. *Journal of Molecular Biology* **2007**, *365*, 663.

(9) Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P. *Biochemistry* **2009**, *48*, 5623.

(10) Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. *Journal of Computational Biology* **2000**, *7*, 203.

(11) Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P. *Journal of Biological Chemistry* **2010**, *285*, 12078.

(12) Ardejani, M. S.; Li, N. X.; Orner, B. P. *Biochemistry* **2011**, *50*, 4029.

(13) Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J.-C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F. In *2-D Proteome Analysis Protocols*; Link, A. J., Ed.; Humana Press: 1999; Vol. 112, p 531.

(14) Creighton, T. E. *Protein structure: a practical approach*; IRL Press at Oxford University Press: Oxford; New York, 1997.

(15) Gallagher, S. R. In *Current Protocols Essential Laboratory Techniques*; John Wiley & Sons, Inc.: 2010, p 3:A.3C.1.