SUPPORTING INFORMATION

Charge Engineering Controls Cooperative Assembly and Loading in Protein Host-Guest Complexes

Zhiheng Wang,[1] Dai-Bei Yang,[1] Joshua A. Bulos,[1] Rui Guo,[1] Thomas Troxler,[1,2] Sergei Vinogradov,[1,2] Jeffery G. Saven,[1] and Ivan J. Dmochowski[1, *]

[1]Department of Chemistry, University of Pennsylvania, 231 S. 34th Street, Philadelphia, Pennsylvania 19104-6323, U.S.A.

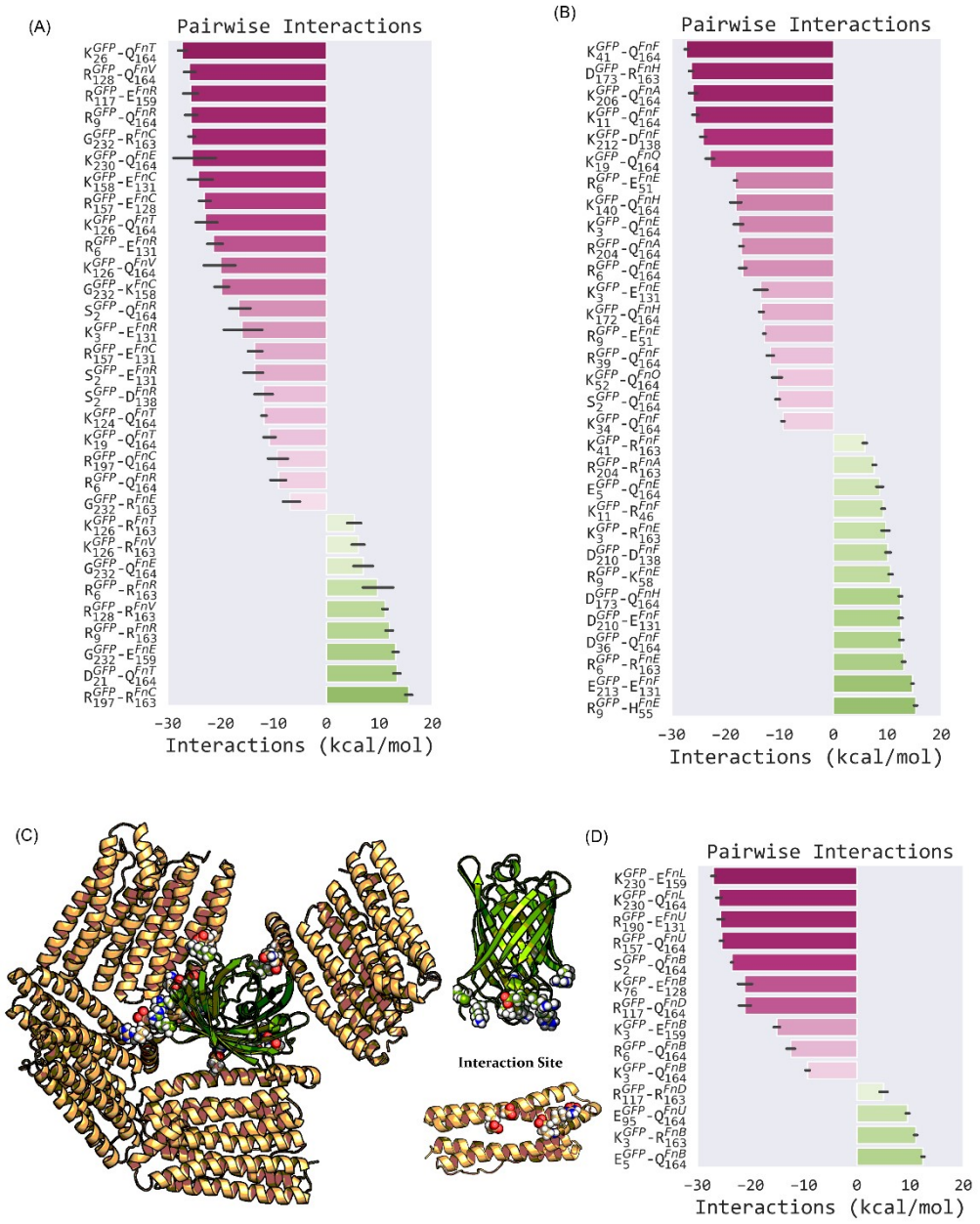
[2]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Corresponding author email: ivandmo@sas.upenn.edu

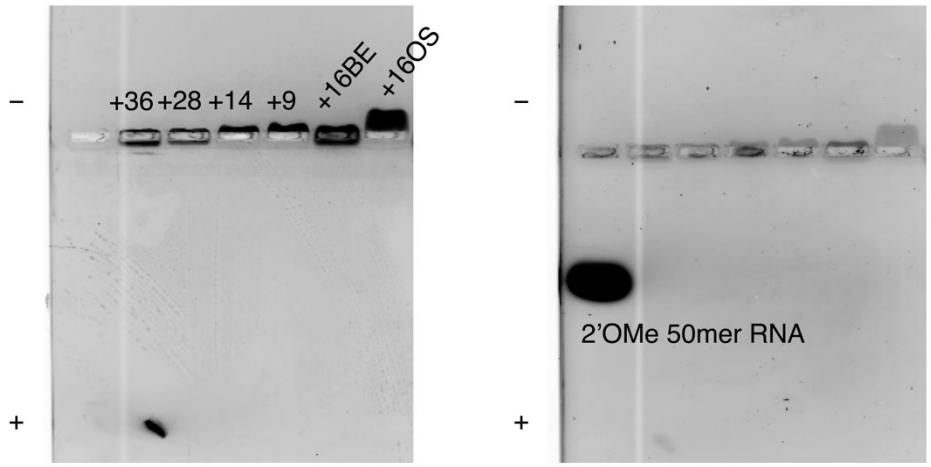
Table of Contents

Supplementary Fig. S1. A third configuration of GFP(+36) binding with AfFtn ……..…..….. 3
Supplementary Fig. S2. Confirmation of nucleic acid-free protein samples ……...…….. 4
Supplementary Fig. S3. Effects of protein concentration on AfFtn-GFP host-guest assembly 5
Supplementary Fig. S4. Native agarose gels of AfFtn mixed with GFP(+36), GFP(+16)BE ……..……...6
Supplementary Fig. S5. Native agarose gels of AfFtn mixed with GFP(+9), GFP(+14), or GFP(+16)OS ……..…...7
Supplementary Fig. S6. Confocal micrographs of free GFP(+28) and AfFtn-GFP(+28) assemblies ……..……8
Supplementary Fig. S7. DLS measurements of mixtures of $AfFtn_2$ and GFP(+9) or GFP(+14) ……..……..…….……....9
Supplementary Fig. S8. TEM images of empty and GFP loaded AfFtn cages at pH 7.6 …..... 10
Supplementary Fig. S9. SEC of AfFtn and AfFtn-GFPs in 20 mM Phos, pH 5.8 buffer .....….. 11
Supplementary Fig. S10. TEM image of AfFtn-GFP(+9) assembly formed at pH 5.8. …...….... 12
Supplementary Fig. S11. Time evolution of the GFP Cα RMSD for monomer (1GFP), dimer (2GFP), and trimer (3GFP) inside the ferritin cage ……. 13
Supplementary Fig. S12. Effects of ionic strength on guest loading density ……...…….… 14
Supplementary Fig. S13. Stability of 1:1AfFtn-GFP(+36) assemblies in buffers with different ionic strengths …….. 15
Supplementary Fig. S14. TEM image of 12-to-1 AfFtn-GFP(+36) complex …….…… 16
Supplementary Fig. S15. Sizes of AfFtn-GFP(+36) assemblies with different host-guest ratios measured by DLS …….……..... 17
Supplementary Fig. S16 Assembly efficiencies with varied GFP variants assessed by SEC …. 18
Supplementary Table 1. Time-resolved fluorescence anisotropy of GFP variants and AfFtn-GFP assemblies under different conditions ……..……..... 18
Additional methods ……..……..……..……..19
Supplementary Fig. S17. Principal axis of GFP ….……..……..……..... 19
Supplementary Fig. S18. Sequence table of GFP wildtype, GFP(+16)OS, and GFP(+16)BE ... 20
Supplementary Fig. S19. Convergence checks for all simulations ……...……..……..…… 22
References ……..……..……. 23

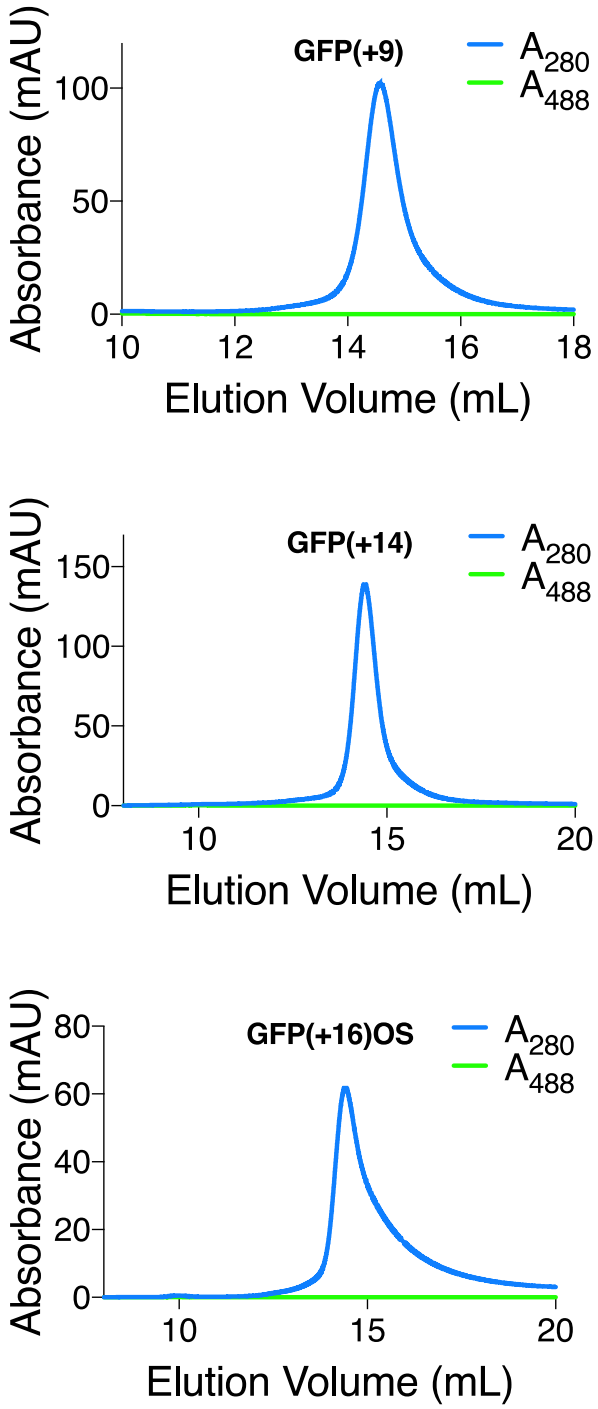
1. Supporting Figures



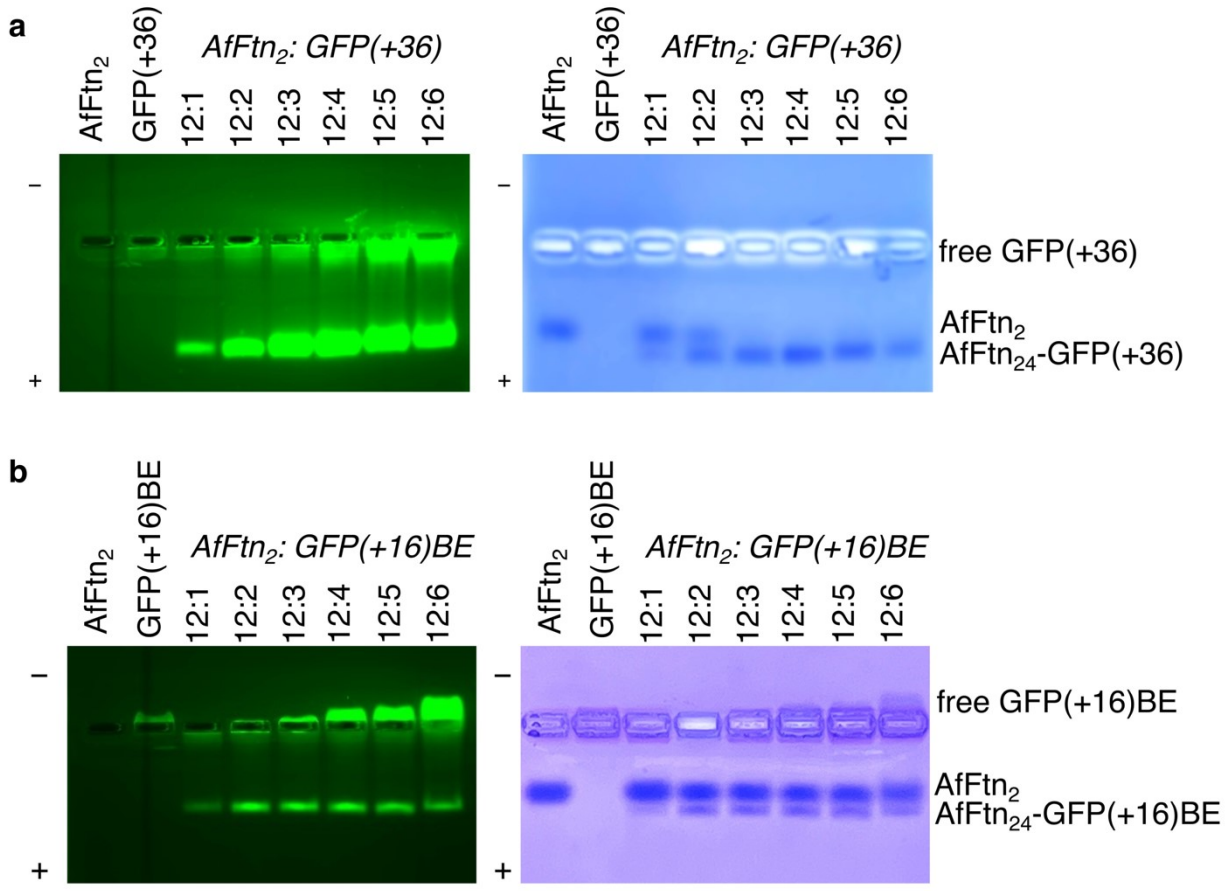
Supplementary Fig S1. Residue contact interaction pairs for configurations shown in Fig. 1. (A) Contact pairs from configuration rendered in Fig. 1a. (B) Contact pairs from configuration rendered in Fig. 1b. Contact pairs are labeled by residue, residue position on GFP or ferritin (FnX, were X labels the subunit chain). Complementary interactions between oppositely charged sidechains are prevalent. (C) Simulations identified a third configuration of GFP(+36) within $AfFtn_{24}$. One side of GFP(+36) is involved in the binding. (D) Residue contact interaction pairs for the configuration rendered in (C). The number of low energy contacts is smaller than for the two configurations, (A) and (B), presented above and in Fig. 1.



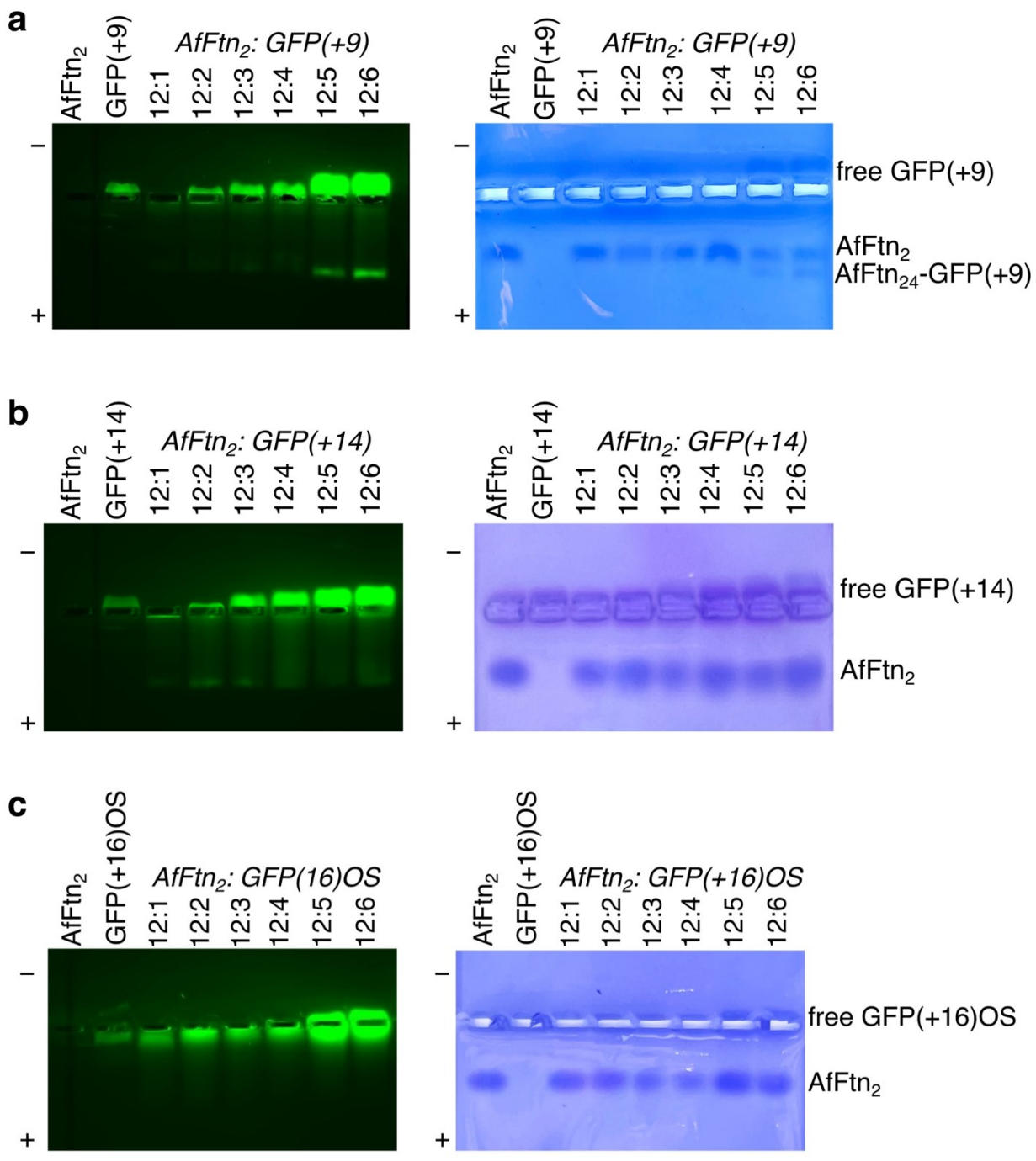
Supplementary Fig S2. Nucleic acid contamination check using native agarose gel. Left: fluorescence imaging before staining. Right: post-stained with SYBR GOLD followed by fluorescence imaging. All GFP samples were determined to be free of oligonucleotide impurities. The excitation wavelength is the same for GFP and SYBR GOLD; hence, GFP was also seen on the right gel. [GFP] = 1 μM, control oligo is 100 picomole 2′-OMe 50mer RNA.



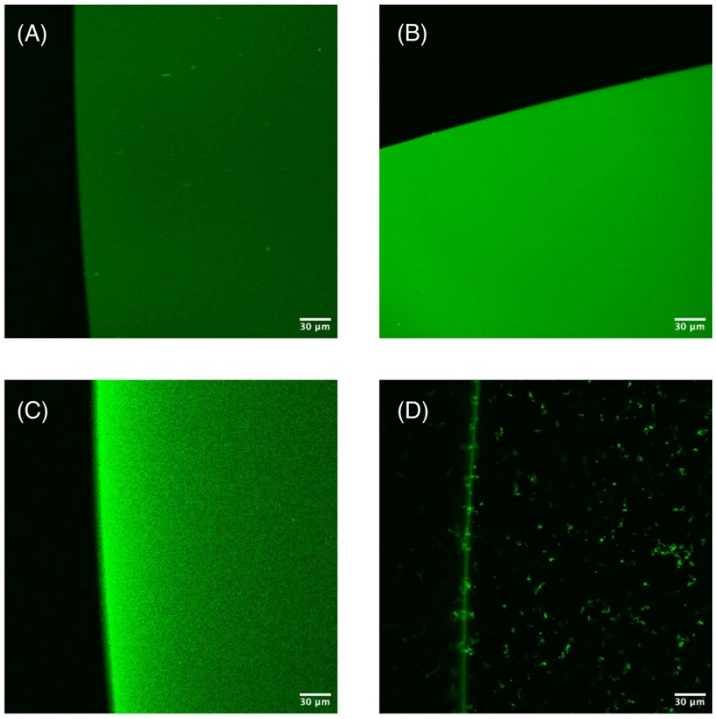
Supplementary Fig S3. Effects of protein concentration on AfFtn-GFP host-guest assembly. SEC traces of AfFtn mixed with GFP(+9), GFP(+14), and GFP(+16)OS at a 12-to-1 $AfFtn_2$:GFP loading ratio. Despite a high concentration of GFP (10 μM), no assembly was induced by all three variants.



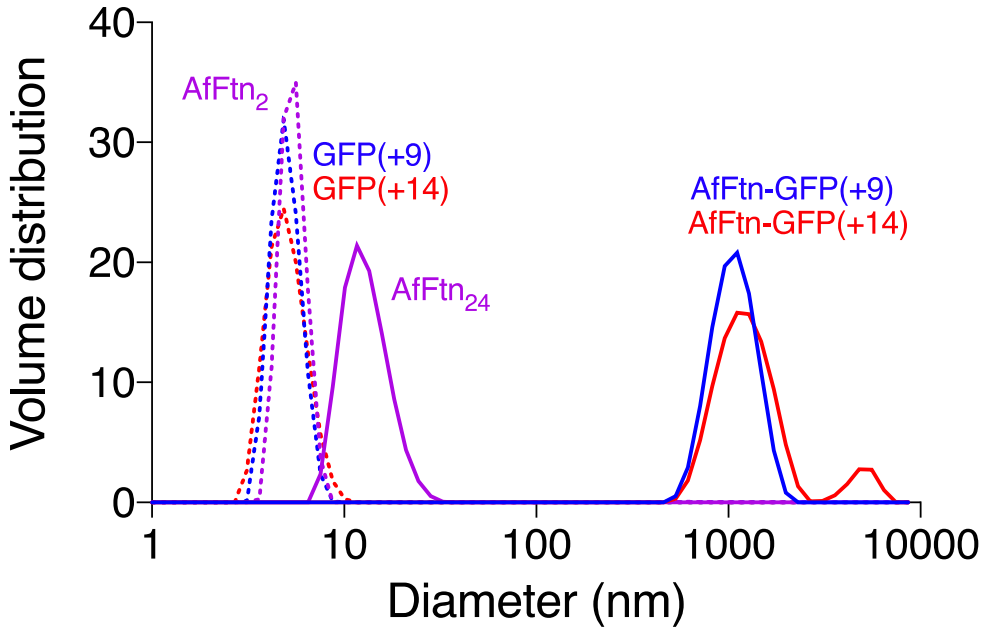
Supplementary Fig S4. Effects of varying stoichiometry on AfFtn-GFP host-guest assembly. $AfFtn_2$ is mixed with (A) GFP(+36) and (B) GFP(+16)BE at different ratios. The overlap between the green fluorescent band (GFP, left) and the blue band (ferritin, right) supports GFP encapsulation within the AfFtn assembly. Both GFP variants are encapsulated at all tested stoichiometries.



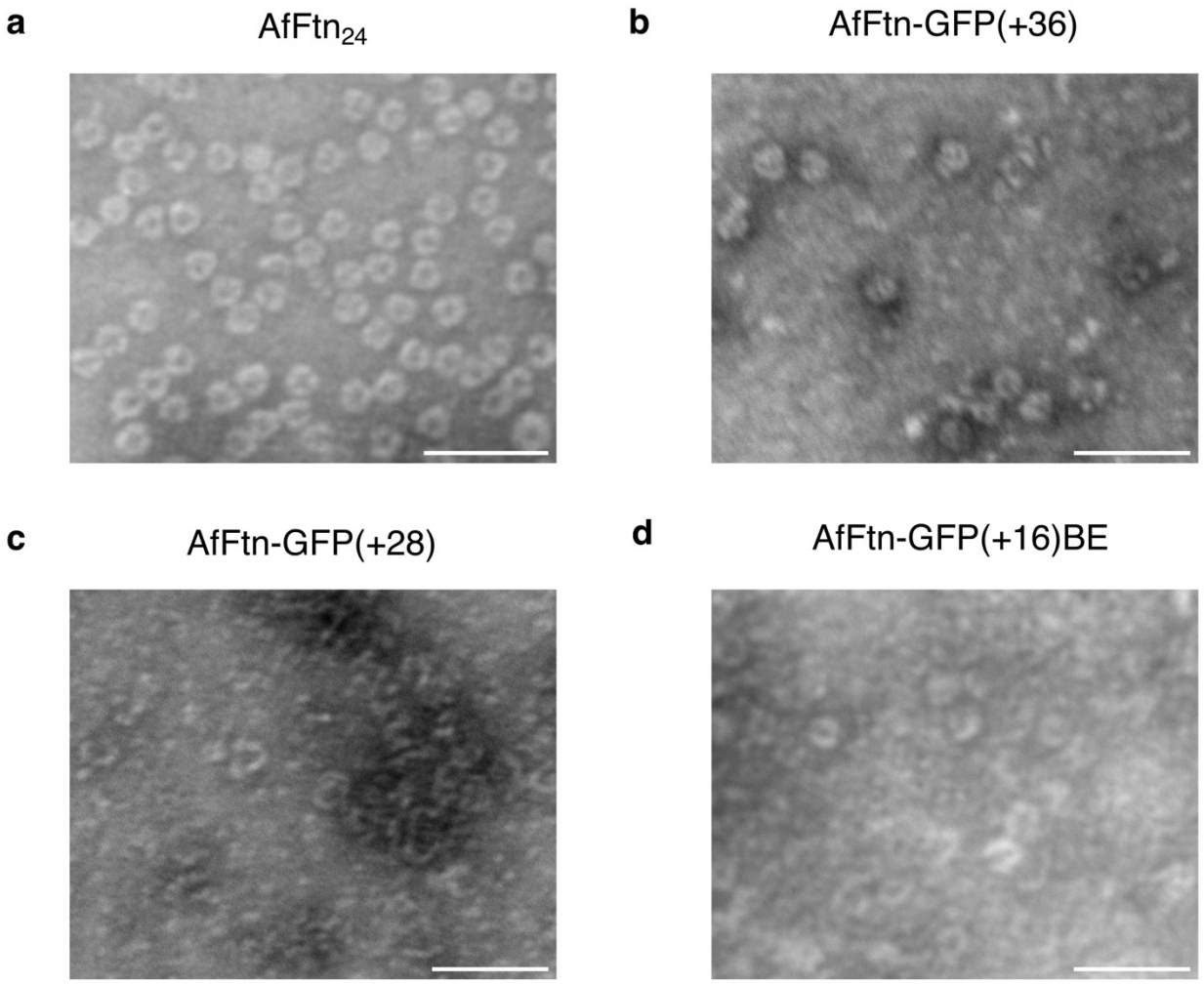
Supplementary Fig S5. Effects of varying stoichiometry on AfFtn-GFP host-guest assembly. $AfFtn_2$ is mixed with (A) GFP(+9), (B) GFP(+14), and (C) GFP(+16)OS at different ratios. Very modest encapsulation was observed for GFP(+9) and GFP(+14), while GFP(+16)OS showed no sign of encapsulation at any ratios tested. It was learned that charge distribution plays a role in assembly formation. The localization of charges on the surface of GFP(+16)OS might prefer a specific binding configuration with $AfFtn_2$, which prevents the formation of $AfFtn_{24}$. Compared to GFP(+16)OS, charges are more spread out on the surfaces of GFP(+9) and GFP(+14), which may help to explain the partial encapsulation of these variants at the highest loading ratios.



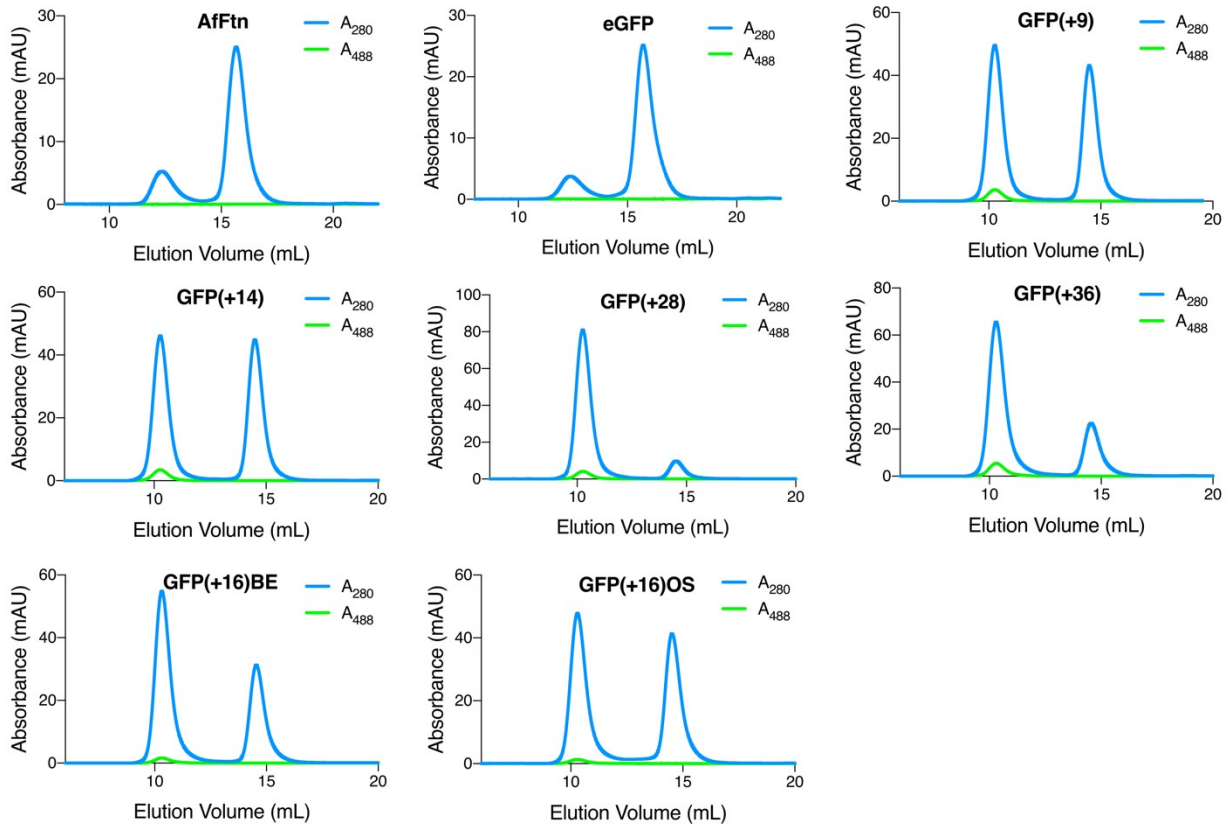
Supplementary Fig S6. Confocal fluorescence micrographs of free GFP(+28) in PBS (A) and no-salt buffer (B). (C) SEC-purified 12: 1 $AfFtn_2$: GFP(+28) assembly. (D) Micrometer-sized puncta were formed when mixing $AfFtn_2$ with GFP(+28) at a 12:6 ratio, demonstrating the effect of stoichiometry on the structure of the assembly. GFP(+28) concentrations were 0.5 – 1 μM in all experiments.



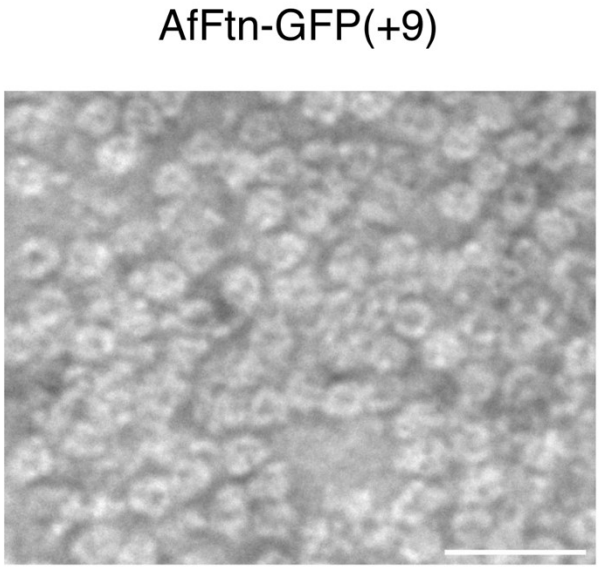
Supplementary Fig S7. Dynamic light-scattering measurements of diameters of $AfFtn_2$ and GFP(+9) and GFP(+14) variants before (dotted line) and after mixing at a 12-to-1 ratio (solid line). Particles with diameters greater than 1000 nm were formed when mixing $AfFtn_2$ with either variant.



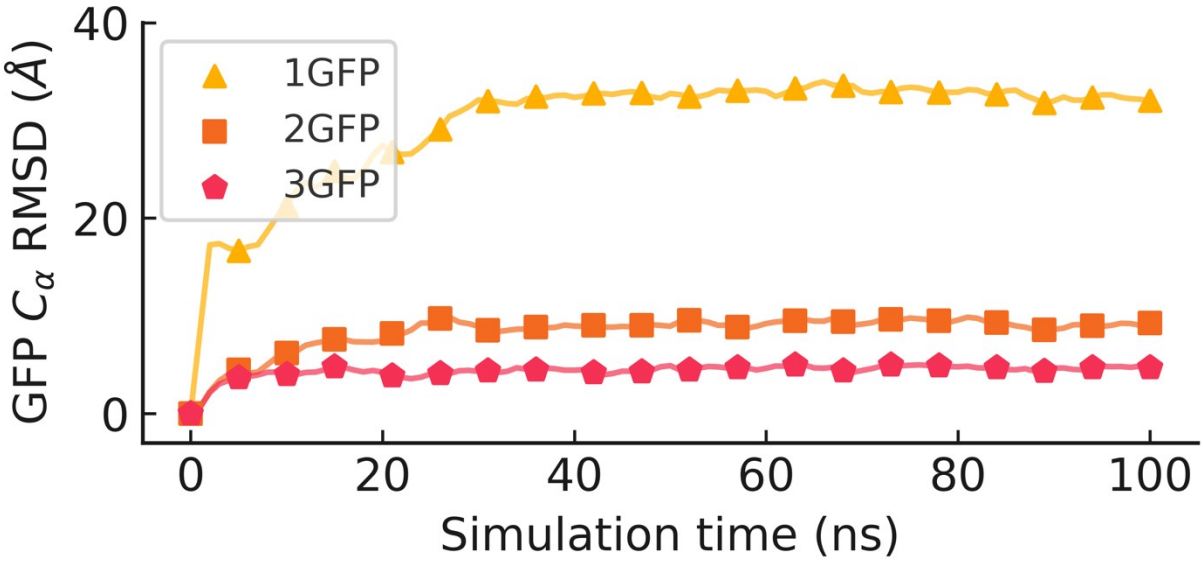
Supplementary Fig S8. TEM images of (a) native $AfFtn_{24}$ formed in the high ionic strength buffer (20 mM sodium phosphate, 800 mM NaCl, pH 7.6) with a hollow core stained by uranyl acetate, (b) AfFtn-GFP(+36), (c) AfFtn-GFP(+28), and (d) AfFtn-GFP(+16)BE complexes show a whiter core, indicating that the guest molecules expel the negative stain. Scale bar: 50 nm.



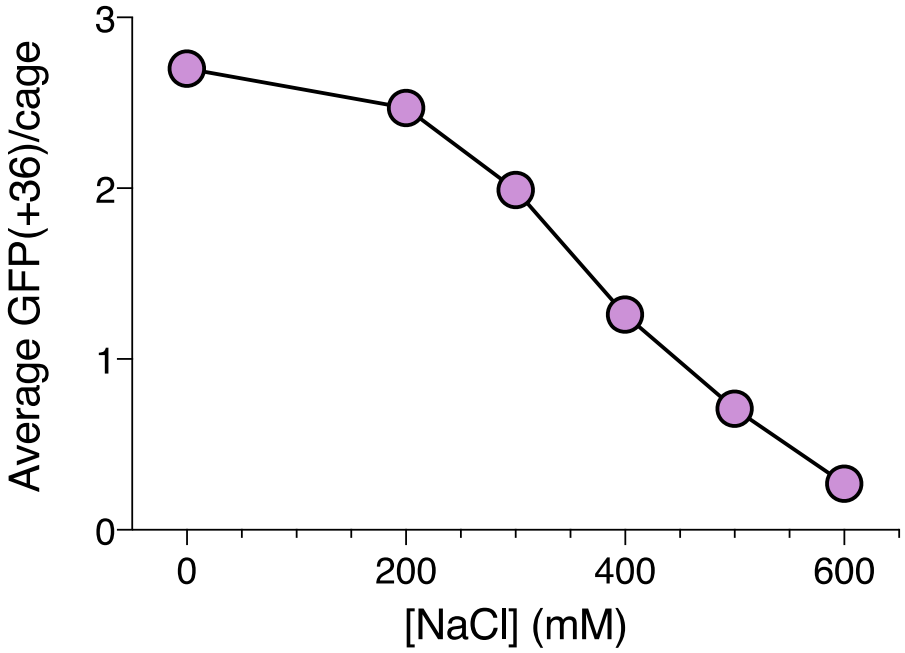
Supplementary Fig S9. SEC of AfFtn and AfFtn-GFP assemblies in 20 mM Phos, pH 5.8 buffer. AfFtn (top left) sample contains only AfFtn; all others contain AfFtn and the indicated GFP variant. In the absence of any GFP, roughly 20% AfFtn remains assembled at pH 5.8 (top left). All the superpositively charged GFPs template AfFtn 24mer formation in these buffer conditions, while negatively charged eGFP (-7) does not (Top middle). Fractions of AfFtn 24mer and AfFtn dimer are obtained using areas under the corresponding peaks (blue).



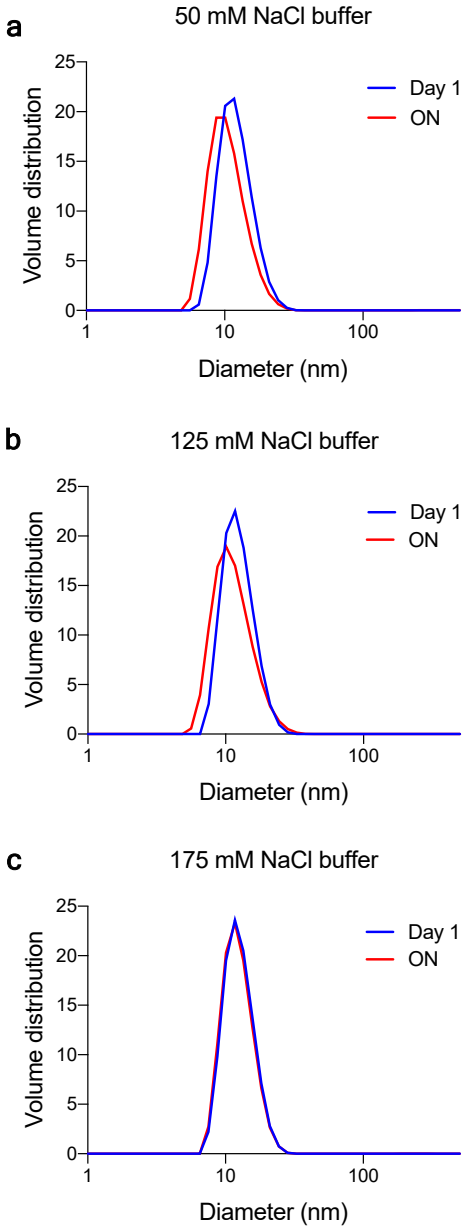
Supplementary Fig S10. TEM images of AfFtn-GFP(+9) assemblies formed in pH 5.8 buffer. The sample was purified with analytical SEC before grid preparation. Scale bar: 50 nm.



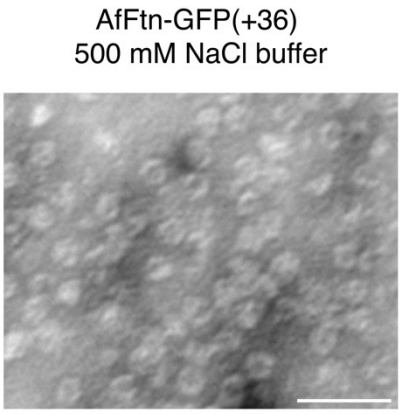
Supplementary Fig S11. Time evolution of the GFP Cα RMSD for monomer (1GFP), dimer (2GFP), and trimer (3GFP) inside the ferritin cage. Trajectories are first aligned to the cage, and RMSD values of GFPs are computed relative to the initial frame. The results show that the monomer exhibits the highest RMSD, suggesting greater mobility. In contrast, the dimer and trimer exhibit lower RMSD values, indicating increased confinement as the number of cargo proteins increases.



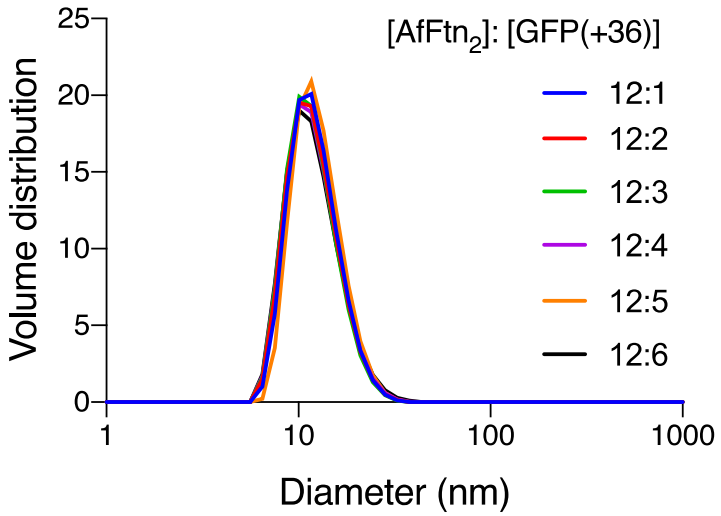
Supplementary Fig S12. Effect of ionic strength (20 mM Phos, pH 7.6) on the average number of GFP(+36) proteins encapsulated per $AfFtn_{24}$ assembly. An input loading ratio of 12 $AfFtn_2$ to 1 GFP(+36) was used for all tested ionic strengths.



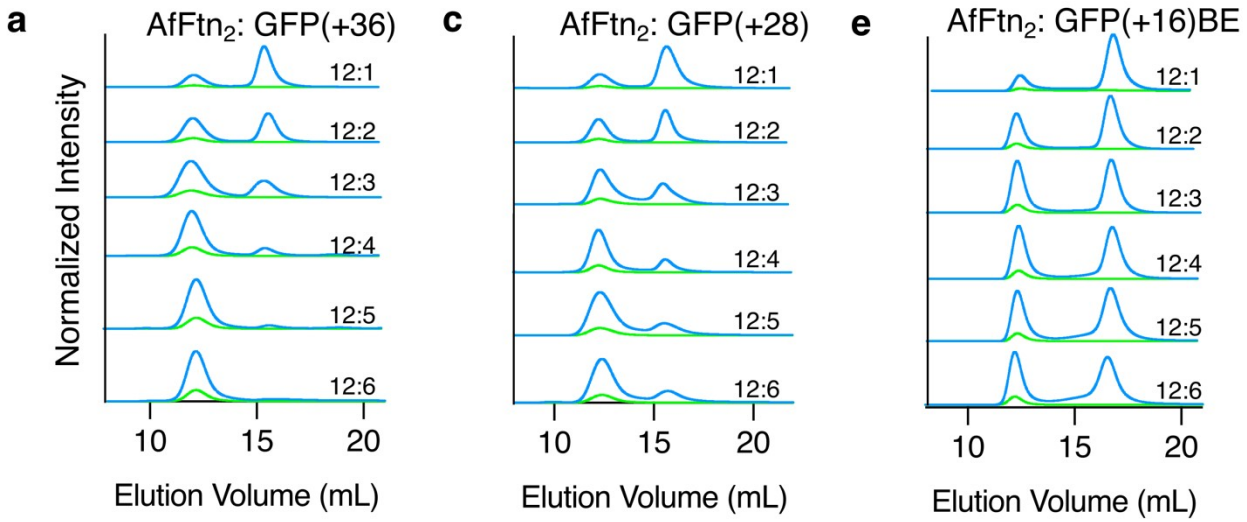
Supplementary Fig S13. Stability of 1:1 AfFtn$_{24}$-GFP(+36) assembly, formed in 500 mM NaCl buffer and buffer exchanged in (a) 50, (b) 125, and (c) 175 mM NaCl buffer solutions measured by DLS. The blue trace represents the diameter of the assembly right after FPLC purification, and the size is measured again after overnight incubation, shown in red. No discernible change in size is observed when 175 mM NaCl concentration is used, indicating this ionic strength is high enough to stabilize the assembly when fewer cargo are encapsulated.



Supplementary Fig S14. TEM image of the 12-to-1 AfFtn-GFP(+36) complexes formed in 20 mM phosphate, 500 mM NaCl buffer. Some cages show a darker core, while others show a whiter core, consistent with a mixture of empty and filled cages. Scale bar: 50 nm.



Supplementary Fig S15. Hydrodynamic diameters of AfFtn-GFP(+36) assemblies formed at different $AfFtn_2$: GFP input loading ratios. All assemblies were purified by analytical SEC, and only the peak corresponding to $AfFtn_{24}$ (the first protein sample to elute) was taken for DLS analysis. The average assembly diameters and distributions are similar despite a different number of GFP(+36) proteins encapsulated at each ratio. The average diameter is 13 nm.



Supplementary Fig S16. Varying loading densities and assembly efficiencies with supercharged GFP guests mixed at different ratios in pH 7.6 buffer. SEC traces (normalized to the highest intensity) with increasing (a) GFP(+36), (c) GFP(+28), and (e) GFP(+16)BE to $AfFtn_2$ ratios. Blue: absorbance at 280 nm; Green: absorbance at 488 nm.

Supplementary Table 1. Time-resolved fluorescence anisotropy of GFP variants and AfFtn-GFP assemblies under different conditions

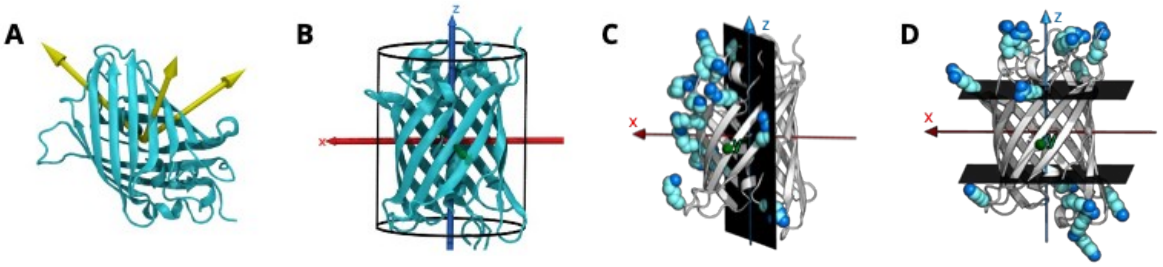
Complex	Buffer pH	[NaCl[(mM)	Lifetime (ns)	Anisotropy Decay Times (ns)		
				τ_1 (R_1)	τ_2 (R_2)	τ_{inf} (R_0)
AfFtn-GFP(+36)	7.6	0	2.5	1.4 (0.091)	0.26 (0.166)	(0.134)
AfFtn-GFP(+36)	5.8	0	2.6	2.3 (0.043)	0.19 (0.126)	(0.235)
AfFtn-GFP(+36)	7.6	175	2.6	*na* (0.312)	0.36 (0.057)	
AfFtn-GFP(+36)	7.6	500	2.5	*na* (0.314)	0.19 (0.048)	
GFP(+36)	7.6	0	2.6	12.1 (0.291)	0.20 (0.031)	
GFP(+36)	5.8	0	2.6	12.0 (0.294)	0.11 (0.098)	

AfFtn-GFP(+36) assemblies were purified by SEC. *R* represents the amplitude of anisotropy. τ stands for the timescale of measured anisotropy, τ2 is instrument response, and τ_{inf} is the infinite component that is much slower than the fluorescence lifetime of GFP.

2. Additional methods

GFP(+16)OS and GFP(+16)BE design

The structure of the superfolder GFP (PDB 2B3P)[1] was used for the design of GFP(+16)OS and BE variants. To define the top, bottom, and side regions of the GFP barrel, the GFP structure was centered at the origin, and its principal axes (of the moment of inertia tensor) were aligned with a Cartesian coordinate system (Suppl. Fig. S17a), in which the "barrel axis" aligned with the z-axis (Suppl. Fig. S17b). For the "one side" [(+16)OS] variant, after inspection of the structure the two sides of the barrel are divided by a plane that formed a 120° angle with the positive direction of the x-axis (Suppl. Fig. S17c); residues on one side were selected from positioning of positively charged residues. For the "both ends" [(+16)BE] variant, residues at the two ends were defined as those whose alpha carbons have z-coordinate values of $z < -10$ Å or $z > 10$ Å, indicated by planes in Suppl. Fig. S17d. A pool of candidate mutations was gathered from known positively charged GFP variants (GFP+9, GFP+14, GFP+28, GFP+36, and GFP+48) previously reported.[2] From this pool, amino acid residue identities were selected as either positive (known lysine or arginine substitutions) or wild type superfolder GFP so as to achieve the targeted patterning of charge at the selected residues. For the GFP(+16)OS variant, mutations were selected so that their alpha carbons all fell on the same side of the barrel (Suppl. Fig. S17c). Whereas, for the GFP(+16)BE variant, all positive mutations whose alpha carbon was at the top or bottom of the barrel were selected (Suppl. Fig. S17d). The mutated residues are highlighted in the sequences presented in Supplementary Fig. S18.



Supplementary Fig S17. Template structure of superfolder GFP (PDB 2B3P).[1] (A) The principal axes of the GFP barrel are shown in yellow. (B) GFP is centered at the origin, with its major principal axis aligned to the z-axis. (C) A plane divides the two sides of the GFP barrel. The plane is perpendicular to the xy plane and goes through the z-axis while forming a 120° angle with the positive direction of the x-axis. Positively charged residues in GFP(+16)OS are rendered as space-filling. (D) Residues with positive charges are above and below indicated planes, with the z-coordinates of their alpha carbons having $z < -10$ Å and $z > 10$ Å. Positively charged residues applied in GFP(+16)BE are rendered as space-filling.

Supplementary Fig S18. Sequences of wild-type superfolder GFP (PDB 2B3P),[1] GFP(+16)BE, and GFP(+16)OS, denoted as WT, BE, and OS, respectively. Mutated residues are highlighted: positive mutations in blue and neutral mutations in green.

New GFP sequences

All GFP(+16) designs have the pJ411 backbone. The insert sequences are as follows:

GFP(+16)BE:
AAGGAGGTAAAAAATGGGTCATCATCATCACCATCATGGCGGTGCCTCGAAGGGCG
AGCGTCTTTTCACTGGTGTAGTCCCAATCCTGGTGGAATTGGACGGCGATGTGAACG
GTCACAAATTCTCAGTGCGTGGGGAGGGCGAAGGCGACGCTACCCGTGGCAAGCTG
ACGCTTAAGTTCATTTGCACCACAGGGAAACTGCCGGTGCCGTGGCCTACGTTGGTA
ACGACTCTTACGTATGGAGTTCAGTGTTTTTCCCGTTACCCGAAACACATGAAGCGT
CACGATTTTTTTAAATCCGCCATGCCCGAAGGGTACGTACAAGAACGTACGATCTCA
TTTAAAAAGGACGGAACTTACAAAACTCGCGCGGAAGTTAAGTTCGAGGGACGTAC
GTTGGTTAACCGCATTGAATTGAAGGGCCGTGATTTTAAGGAAAAAGGTAACATCTT
GGGGCACAAATTGCGTTATAATTTCAACTCCCACAACGTCTACATTACAGCGGATAA
ACGTAAAAATGGCATCAAAGCCAATTTTAAAATTCGCCACAATGTAAAAGACGGAT
CTGTTCAGTTAGCAGATCACTATCAACAAAATACTCCGATTGGCCGCGGTCCGGTGT
TACTGCCTCGTAACCACTACTTAAGTACCCAGTCTGTCTTATCAAAAGACCCAAAAG
AGAAACGCGACCACATGGTACTTCTTGAATTTGTCACGGCTGCTGGTATTACCCACG
GGATGGACGAGCTGTACAAA

GFP(+16)OS:
AAGGAGGTAAAAAATGGGTCATCATCATCATCATCACGGCGGCGCGTCTAAAGGTG
AACGTCTGTTTCGCGGTAAGGTCCCTATTTTGGTTAAGCTGAAGGGTGATGTGAATG
GCCACAAGTTTAGCGTTCGTGGTAAAGGCAAGGGTGATGCGACTCGTGGTAAACTG
ACCTTGAAATTCATCTGCACGACGGGTAAGCTGCCGGTCCCGTGGCCGACCCTGGTA
ACGACCCTGACCTATGGTGTTCAATGTTTCAGCCGTTACCCGGACCACATGAAACAG
CACGATTTCTTCAAAAGCGCGATGCCGAAAGGCTACGTGCAAGAGCGTACCATCTC

GTTTAAAGACGACGGTACGTATAAGACCCGTGCAGAAGTGAAGTTCGAGGGTCGTA
CTCTGGTCAACCGCATTGAGTTAAAGGGCATCGACTTTAAGGAAGATGGCAATATTC
TGGGTCACAAACTGCGCTACAACTTCAACTCCCACAATGTGTACATTACCGCGGACA
AACAGAAAAACGGTATCAAGGCTAACTTTAAGATCCGTCACAATGTTGAGGATGGC
AGCGTTCAACTGGCCGACCACTATCAGCAGAACACCCCGATTGGTCGTGGCCCGGT
GTTGCTGCCGGACAATCACTATCTGAGCACGCGCAGCGCGCTGAGCAAAGATCCAA
AAGAAAAGCGCGACCACATGGTTCTGCTGGAATTCGTCACGGCCGCAGGCATCACC
CATGGTATGGATGAACTCTACAAATGATAA

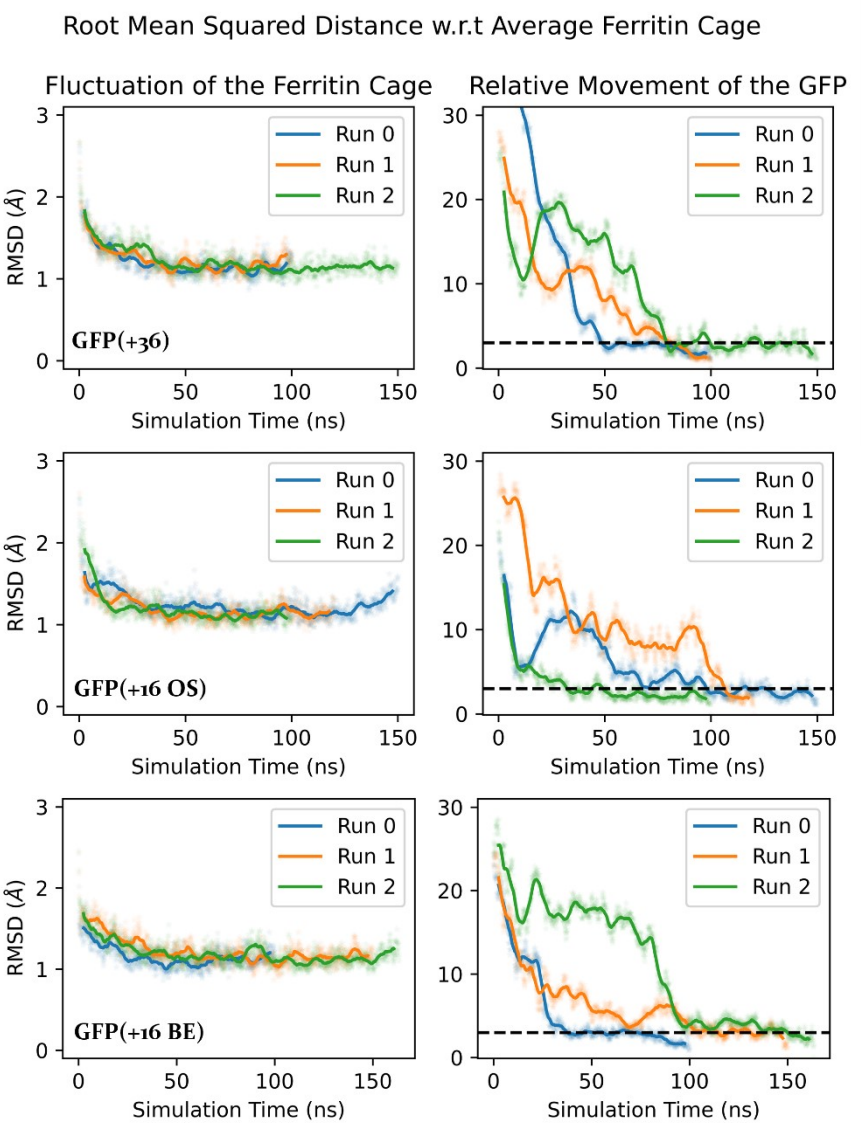
Molecular dynamics simulations

Simulations consisted of one GFP molecule inside one AfFtn cage composed of 24 subunits. GFP is initially positioned at the center of AfFtn and arranged to avoid interaction with the cage. Periodic boundary conditions were adopted in all directions. The minimum distance between any atoms of the protein and the edge of the box is 20 Å. To achieve charge neutrality, Na^+ and Cl^- counterions were added consistent with [NaCl] = 0.15 mol/L were added. For each GFP, three independent runs are carried out from the same initial configuration. Overall, the system contains ~316,000 atoms, and configurations are sampled at 10-ps intervals.

These simulations were carried out first under an NPT ensemble at 310 K and 1.0 atm to mimic the experimental conditions. The CHARMM36 force field was used for all standard amino acids and the fluorophore in GFP.[3] NAMD 3.0 software was used for GPU simulations.[4,5] The control of the temperature is achieved with a Langevin thermostat with a 1.0 ps^{-1} damping coefficient. Pressure control is established with the Nose-Hoover Langevin piston barostat, with a period set to 100 fs and decay set to 50 fs. A time step was chosen to be 2 fs. Water molecules are initially equilibrated for 10 ns with all proteins constrained, followed by a simulation of the system until GFP persistently associates with the AfFtn cage. The minimum simulation duration is 100 ns.

To evaluate whether we obtain a stable binding configuration, a criterion based on the Root Mean Squared Displacement (RMSD) of GFP within AfFtn is used.[6] The AfFtn is aligned first, and later, the RMSD of GFP within the cage is evaluated with reference to the last sampled configuration. A binding is considered to be tight if the Cα-RMSD is below 3 angstroms over at least 10 ns (Supplementary Fig. S19).

Only the last 10 ns (1000 sampled configurations) after tight binding were used in further analysis. Waters and ions are removed. The gRINN software (get Residue Interaction Energies and Networks) with NAMD 2.14 was used to analyze pairwise interactions.[7] The residue pair searching cutoff was set to 12 angstroms, with a dielectric constant of 4 to avoid exaggerating complementary electrostatic interactions. Other parameters in gRINN remained the default. Later, the pairwise interaction plots were rendered with custom Python scripts. Visualization of structures is achieved with pyMOL.[8]



Supplementary Fig S19. Convergence checks for all simulations. In each case, the fluctuation for the cage is minimal, and GFP net displacement is smaller than 3 angstroms for at least 10 ns.

References

(1) Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat Biotechnol* **2006**, *24* (1), 79–88. https://doi.org/10.1038/nbt1172.

(2) Lawrence, M. S.; Phillips, K. J.; Liu, D. R. Supercharging Proteins Can Impart Unusual Resilience. *J. Am. Chem. Soc.* **2007**, *129* (33), 10110–10112. https://doi.org/10.1021/ja071641y.

(3) Huang, J.; MacKerell Jr, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *Journal of Computational Chemistry* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(4) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781–1802. https://doi.org/10.1002/jcc.20289.

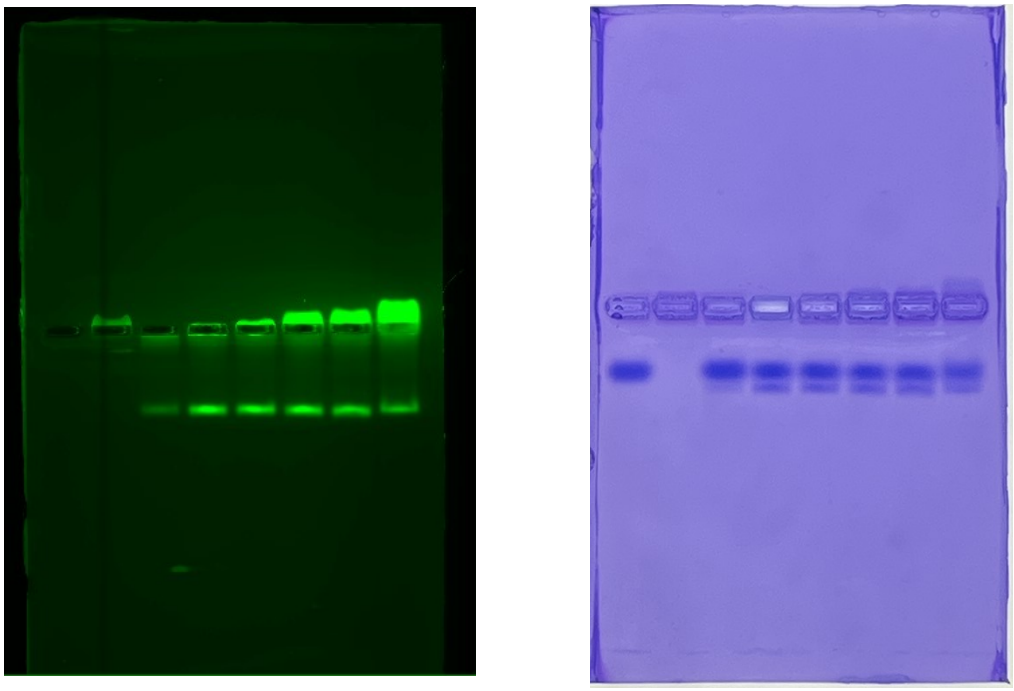
(5) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *The Journal of Chemical Physics* **2020**, *153* (4), 044130. https://doi.org/10.1063/5.0014475.

(6) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

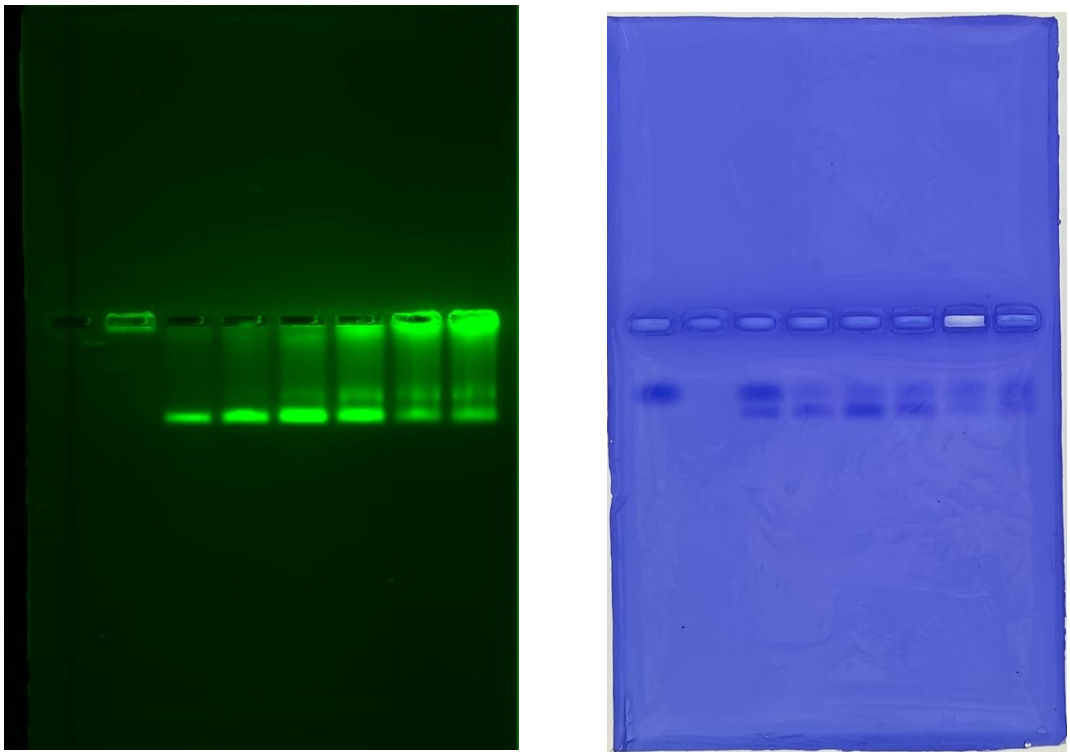
(7) Serçinoğlu, O.; Ozbek, P. gRINN: A Tool for Calculation of Residue Interaction Energies and Protein Energy Network Analysis of Molecular Dynamics Simulations. *Nucleic Acids Research* **2018**, *46* (Web Server issue), W554. https://doi.org/10.1093/nar/gky381.

(8) DeLano, W. L. PyMOL: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr* **2002**, *40* (1), 82–92.

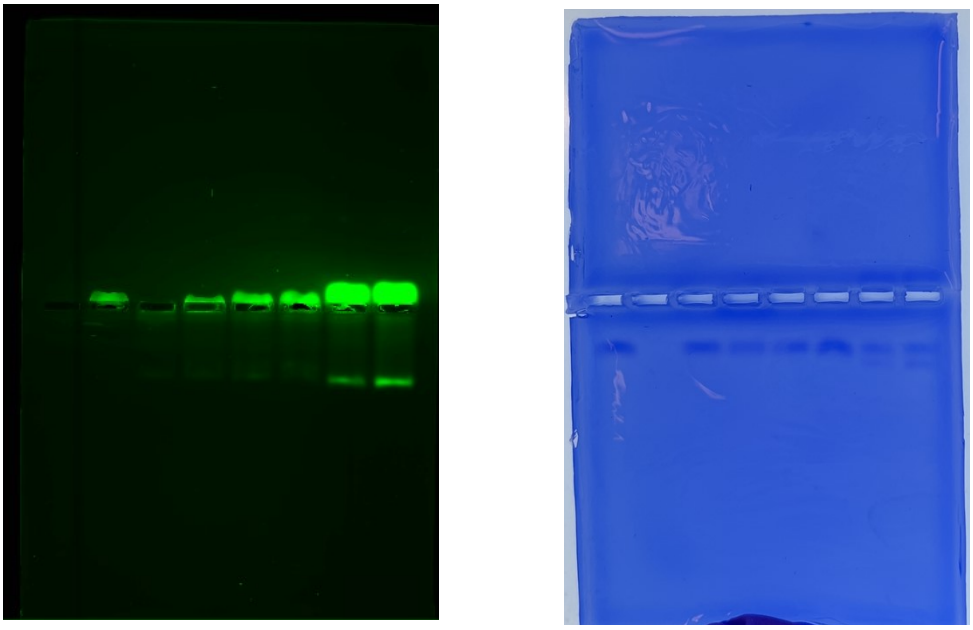
Raw Gel Electrophoresis Images



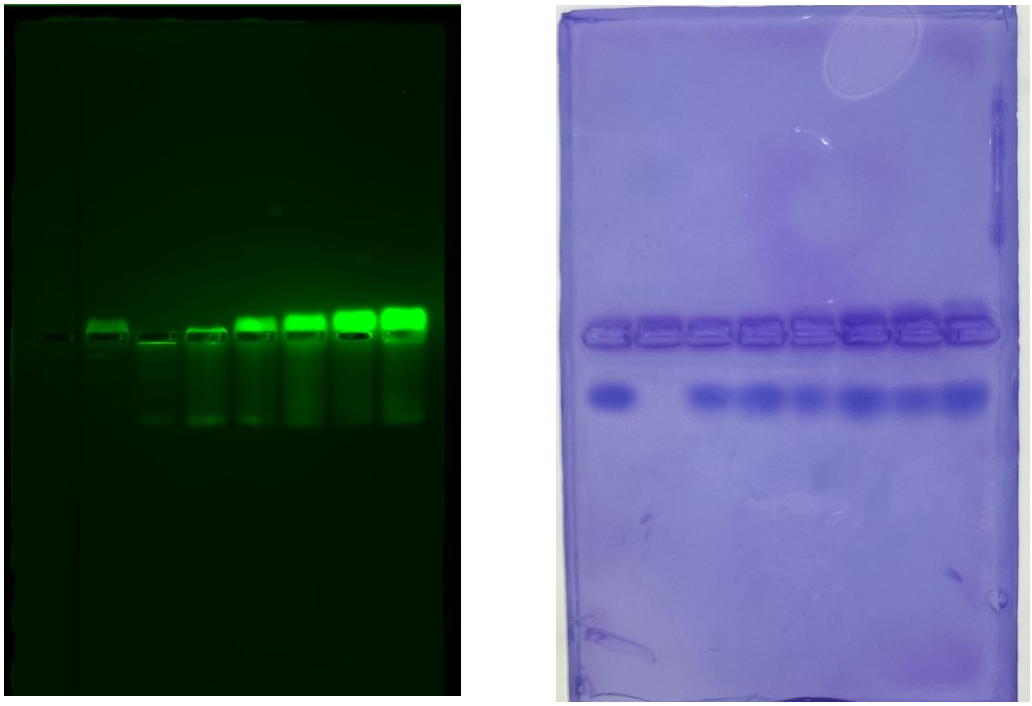
GFP(+16)BE



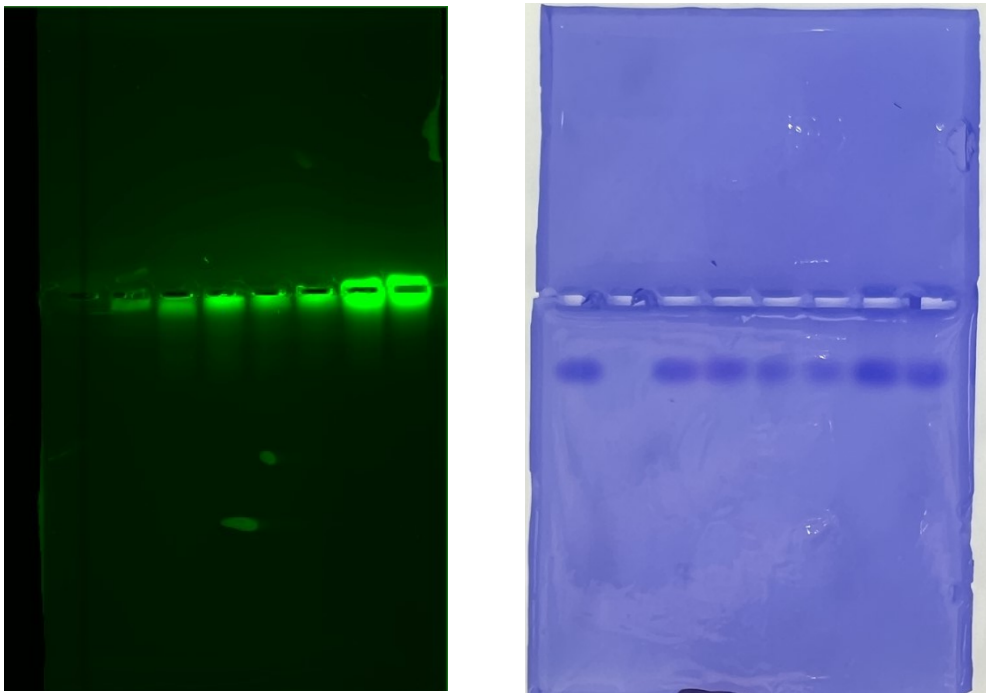
GFP(+28)



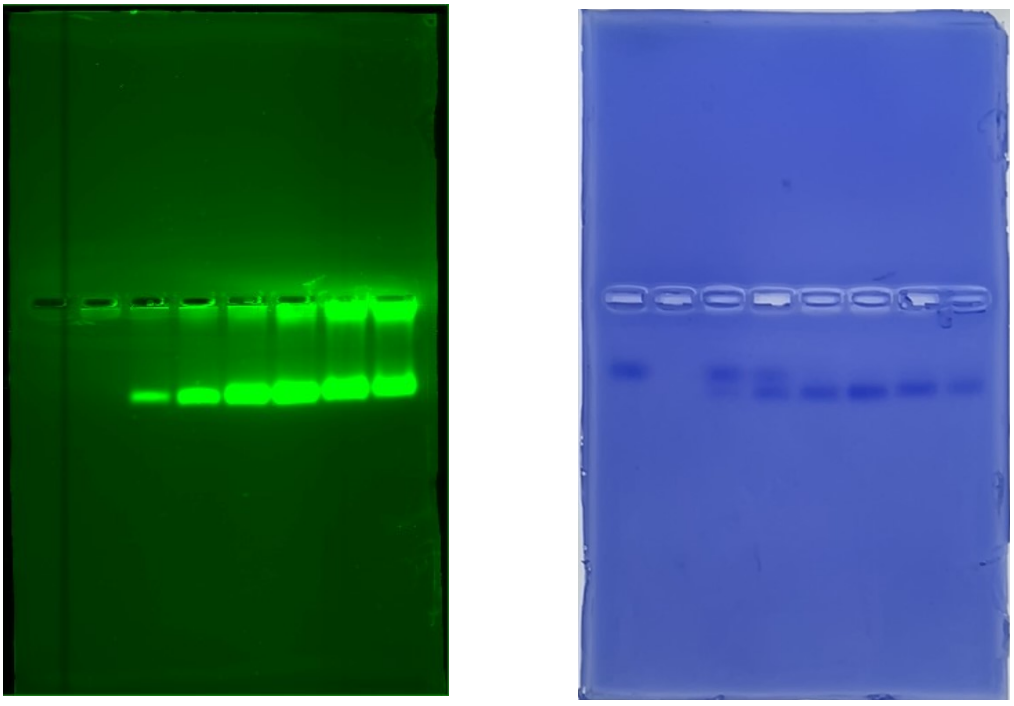
GFP(+9)



GFP(+14)



GFP(+16)OS



GFP(+36)