Electronic Supplementary Information

A single residue can modulate nanocage assembly in salt dependent ferritin

Mantu Kumar,[a,b†] Joanna Markiewicz-Mizera,[a#] Julian David Janna Olmos,[a#] Piotr Wilk[a], Przemysław Grudnik[a], Artur P. Biela[a,c], Małgorzata Jemioła Rzemińska[a,d], Andrzej Górecki[d], Soumyananda Chakraborti[a,e†*], Jonathan G. Heddle[a*]

[a.] Malopolska Centre of Biotechnology, Jagiellonian University, Gronostajowa 7A, 30-392 Krakow, Poland
[b.] Postgraduate School of Molecular Medicine, Żwirki i Wigury 61, 02-091 Warsaw, Poland
[c.] Department of Cell Biology and Imaging, Institute of Zoology and Biomedical Research. Jagiellonian University, Gronostajowa 9,30-387 Krakow
[d.] Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Poland.
[e.] National Institute of Malaria Research, Delhi, India
† These authors contributed equally
# These authors contributed equally
*Authors for Correspondence: Dr. Soumyananda Chakraborti; *email: soumyabiochem@gmail.com, Prof. Jonathan Heddle; *email: Jonathan.heddle@uj.edu.pl

Table of contents:

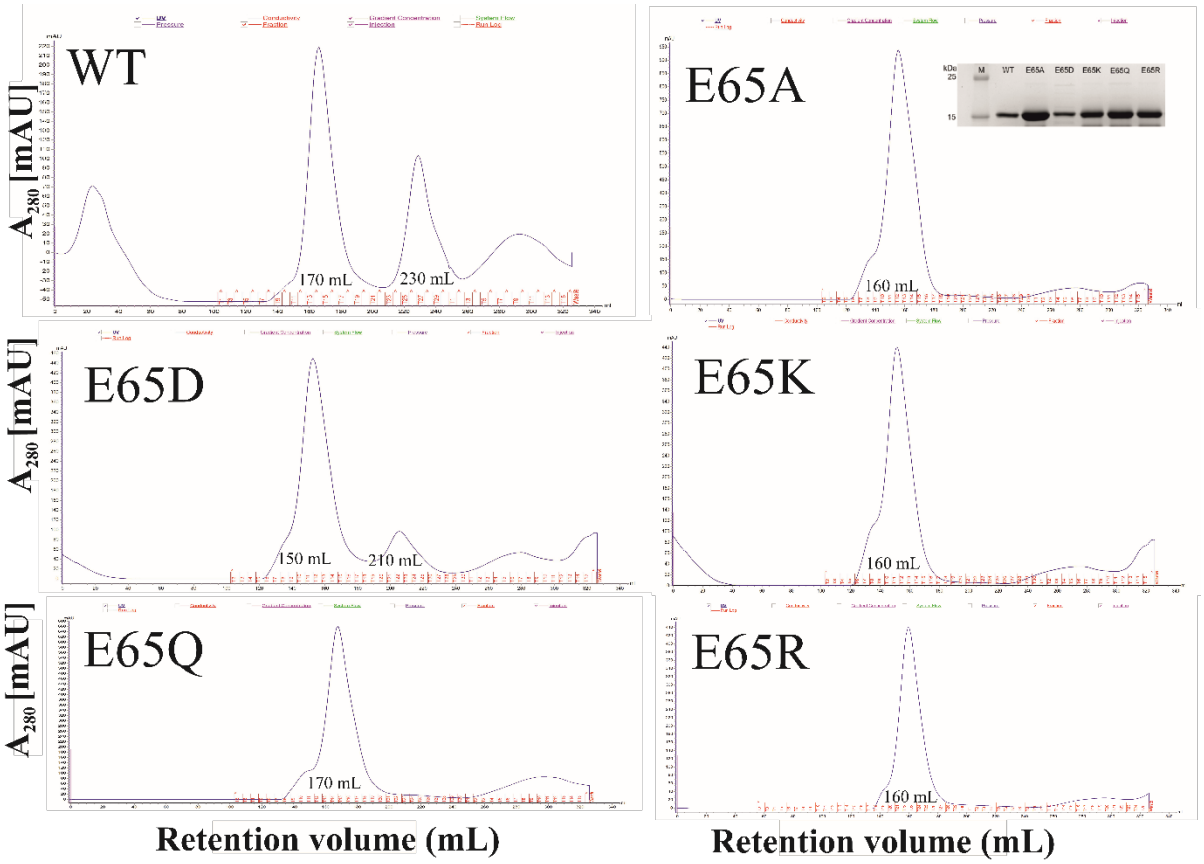


Figure S1. Protein purification and characterisation. SEC elution profiles and corresponding SDS-PAGE gel showing the purity of proteins (inset in E65A panel). WT TmFtn and mutant variants were purified using a HiLoad 26/600 Superdex 200 pg column as a second and final step in the purification. As indicated, cage retention volume was found to be ~ 160 mL while dimer elution can be seen at ~ 230 ml. SDS-PAGE inset displays the position of protein bands corresponding to the 19.4 kDa TmFtn monomer.

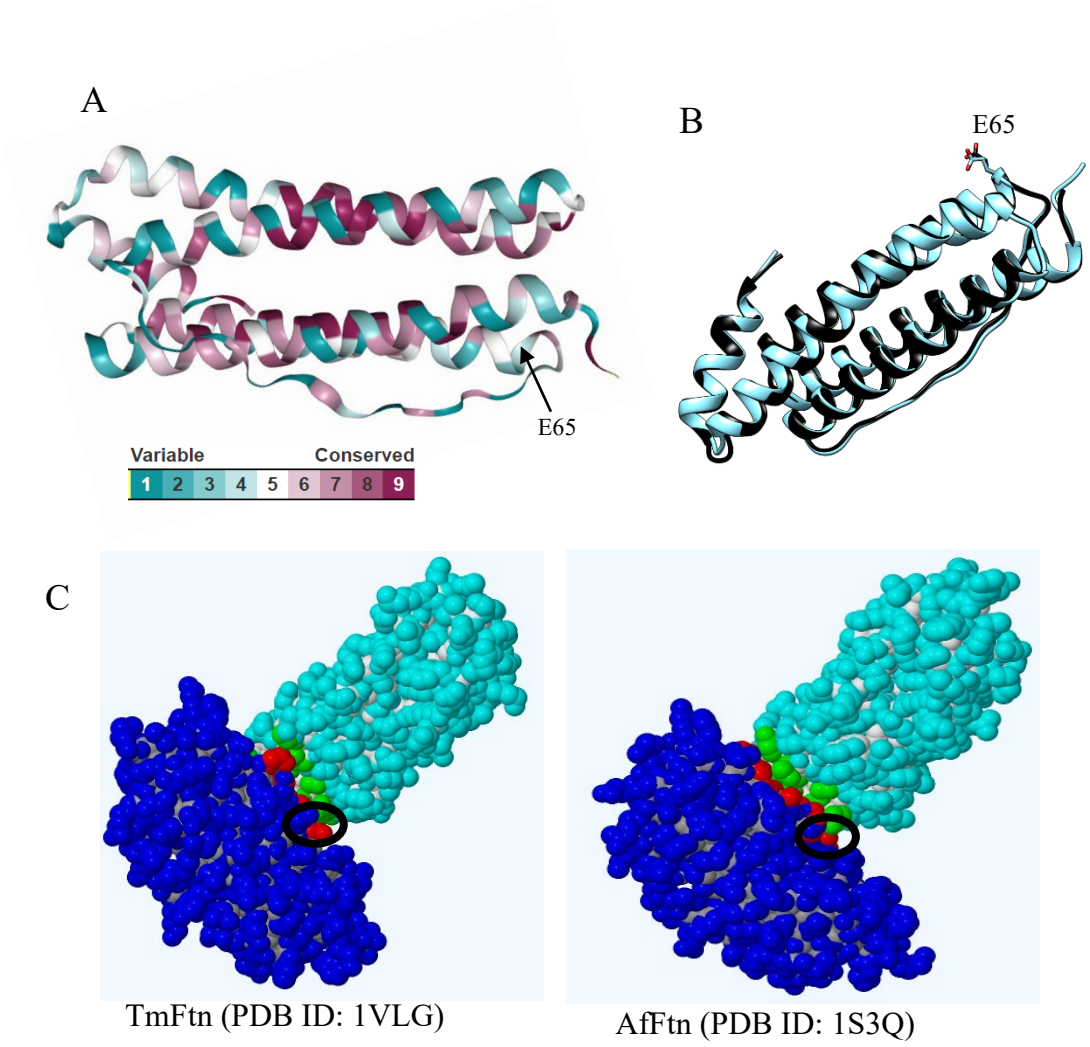


Figure S2. **Bioinformatic analysis of TmFtn and AfFtn** A) Structure based sequence conservation analysis of ferritin monomers comparing 150 sequences from the ferritin family performed using the ConSurf server. The analysis confirms that residue E65 is not conserved. Extent of conservation is shown by the coloured scale bar. B) Structure alignment using the DALI server showing that the position of E65 (stick representation) in AfFtn and TmFtn is conserved, while aligned monomers are shown in black (AfFtn) and cyan (TmFtn). C) Dimeric interface calculated using the PDBePISA server for TmFtn (PDB 1VLG) and AfFtn (PDB 1S3Q). The interface constitutes two monomers (blue and cyan) in perpendicular orientation, and interface residues represented in red and green. In this interface E65 is marked by the black circle.

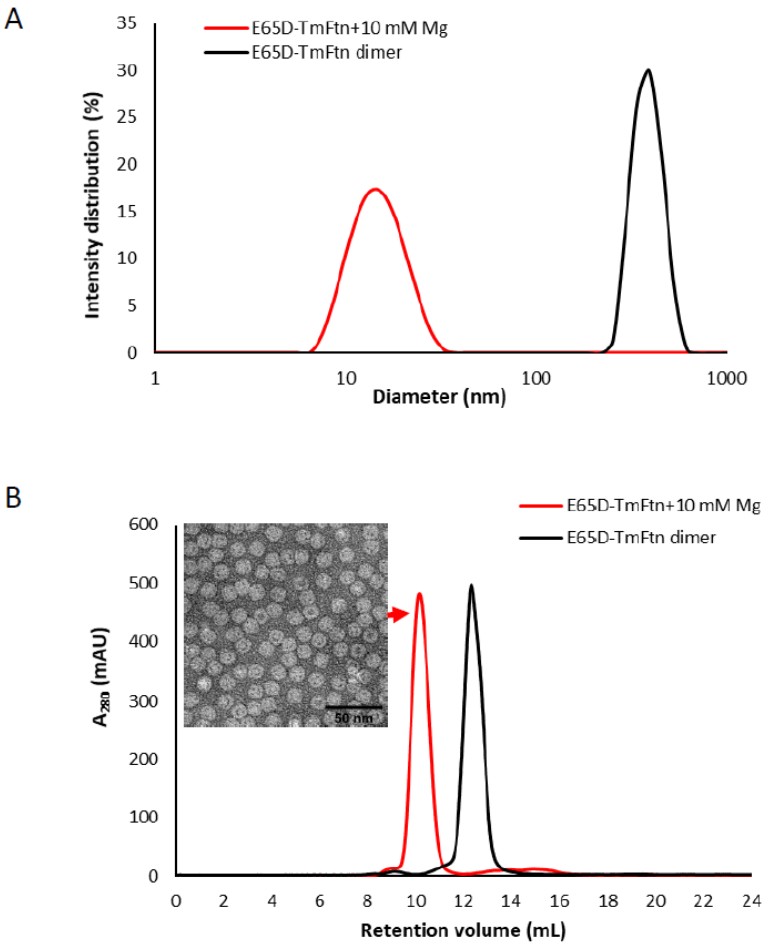


Figure S3. **Dimer to cage transition in the E65D mutant** A) Dimer to cage transition in the E65D mutant measured by DLS in the presence and absence of Mg^{2+}. B) SEC shows similar results: The formation of cage structures was further verified by TEM (inset).

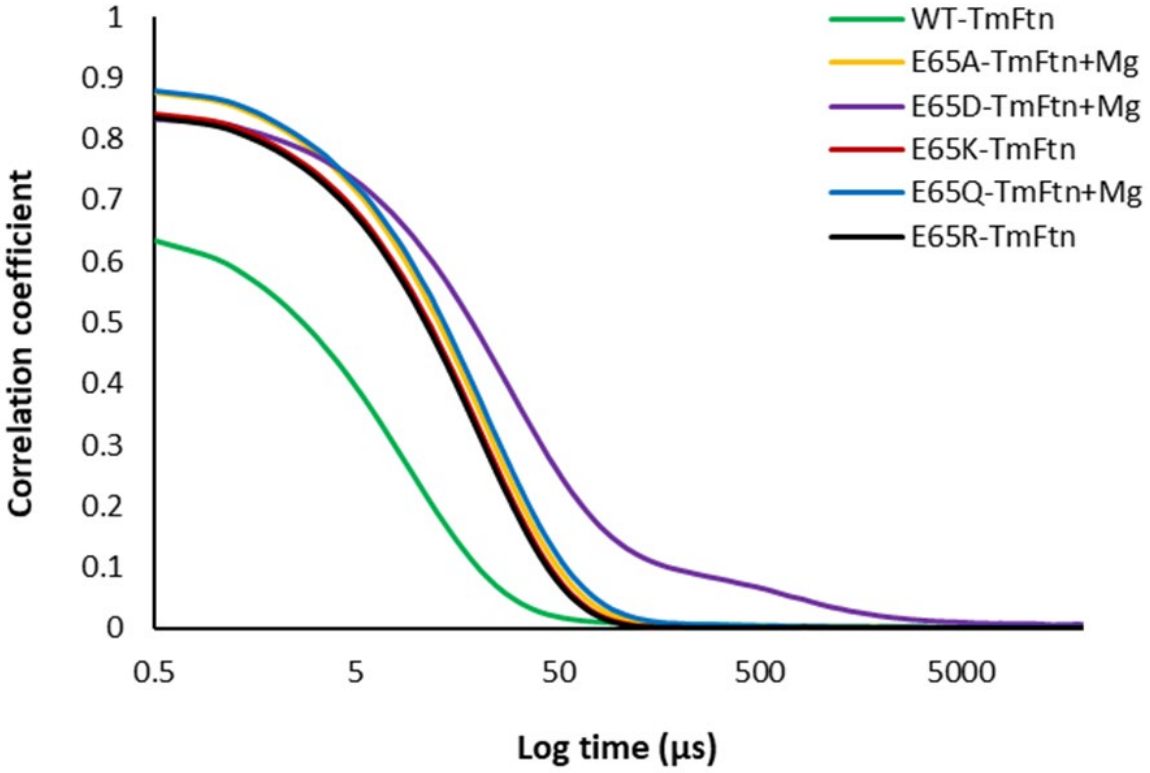
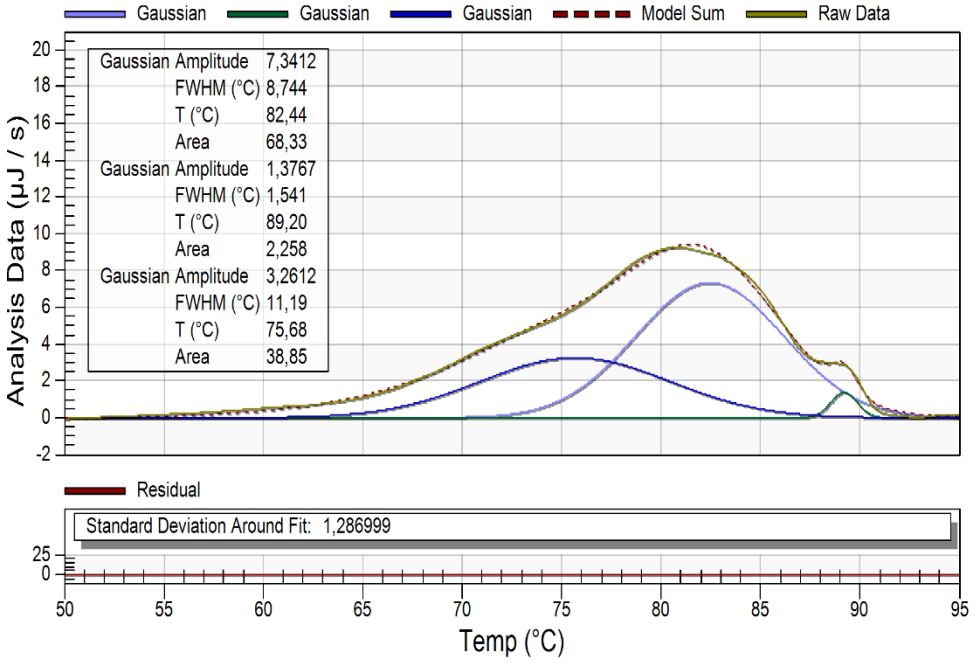


Figure S4. Autocorrelation curves of DLS for different TmFtn mutants. Autocorrelation co-efficient versus log(time) curves of DLS are shown for WT-TmFtn and various mutants. Most TmFtn mutant variant cages show similar autocorrelation curves while dimer (WT-TmFtn) shows a distinct pattern.

Figure S5. A) (top panel) Circular Dichroism of TmFtn protein at different pHs. TmFtn WT and mutants (conc. 1 mg/ml) were incubated overnight in different pH buffers and CD profile checked. There is no significant difference found for pH 2-10 but a drastic change is observed for pH 13, where the CD spectra shows complete denaturation, depicted in the red chromatogram. This is in good agreement with the pH stability studies, as shown in Figure 3 in the main text. **B) (bottom panel) Graphical representation of secondary structure content of different mutants of TmFtn at varying pHs.** The graph was generated using CONTINLL algorithm from the CDPro spectrum decomposition software and SDP48 reference dataset.[1]

A



B

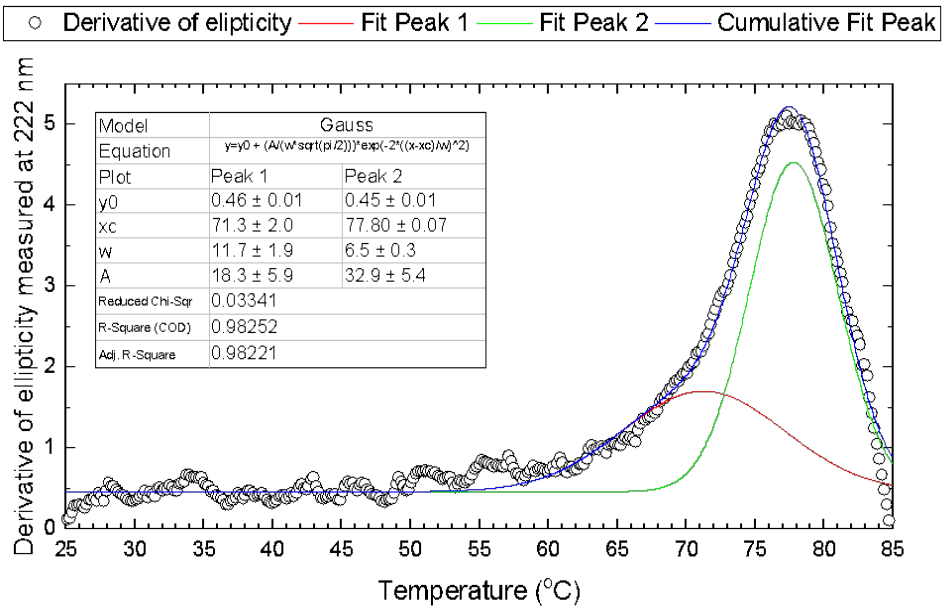
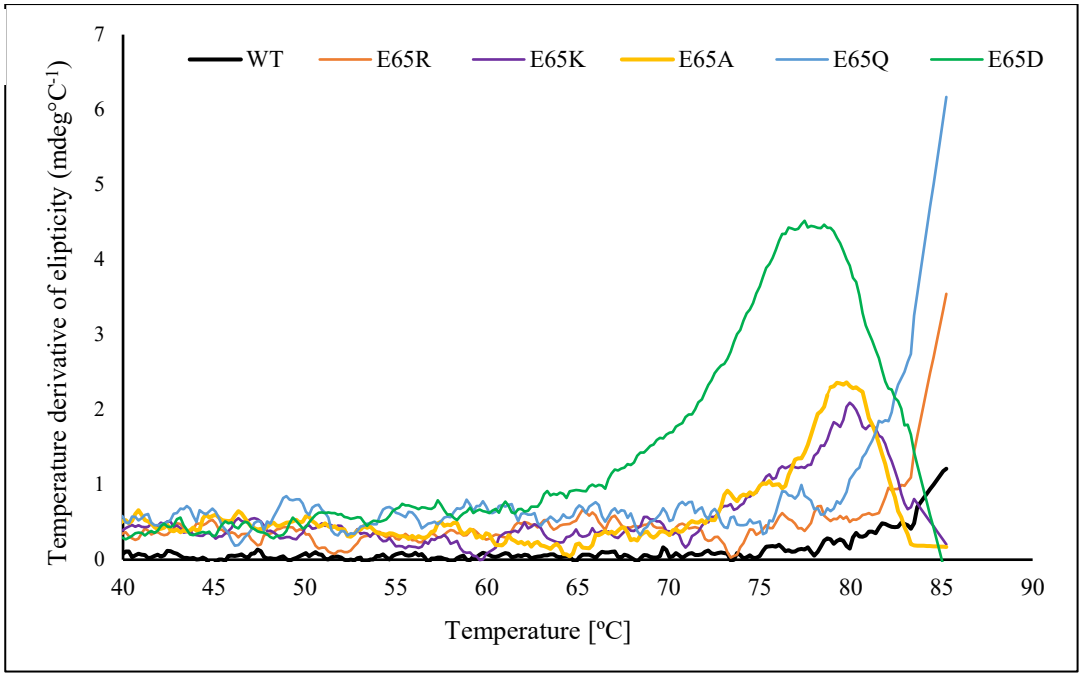


Figure S6. E65D TmFtn mutant stability measurement by DSC and CD. A) DSC thermogram of E65D mutant of TmFtn. Deconvolution of the DSC profile resulted in three Gaussian peaks centred at 75.7, 82.4 and 89.2 ºC. B) CD thermogram of E65D mutant of TmFtn. Deconvolution of the CD profile resulted in two Gaussian peaks centred at 71.3, 77.8 ºC, which closely resembling with the peaks of DSC.

A



B

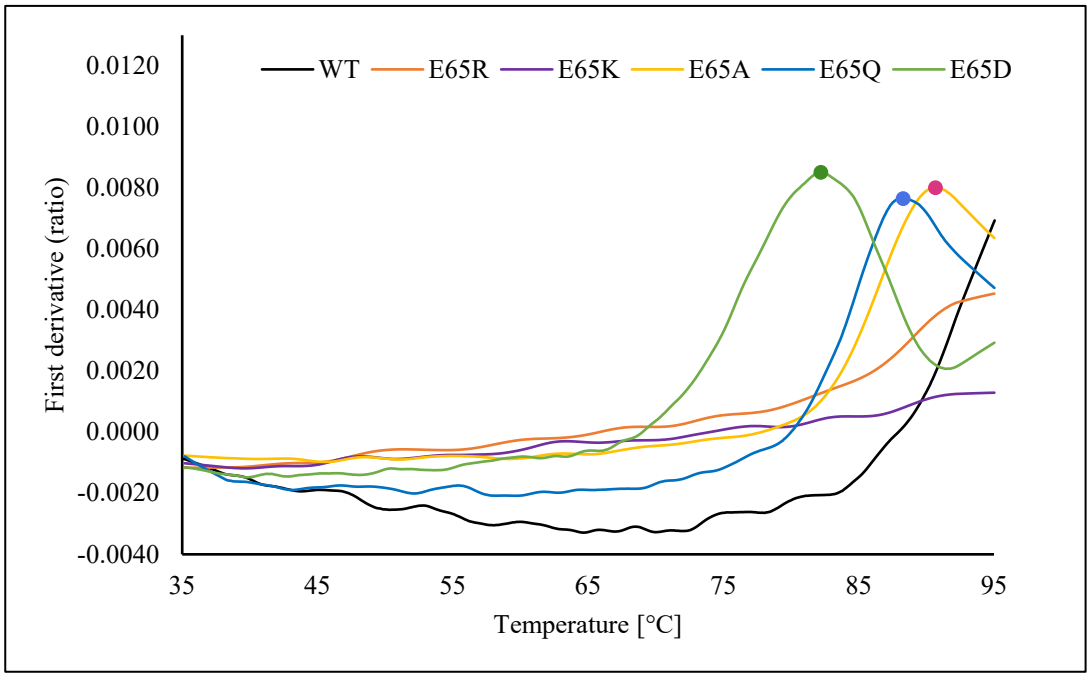
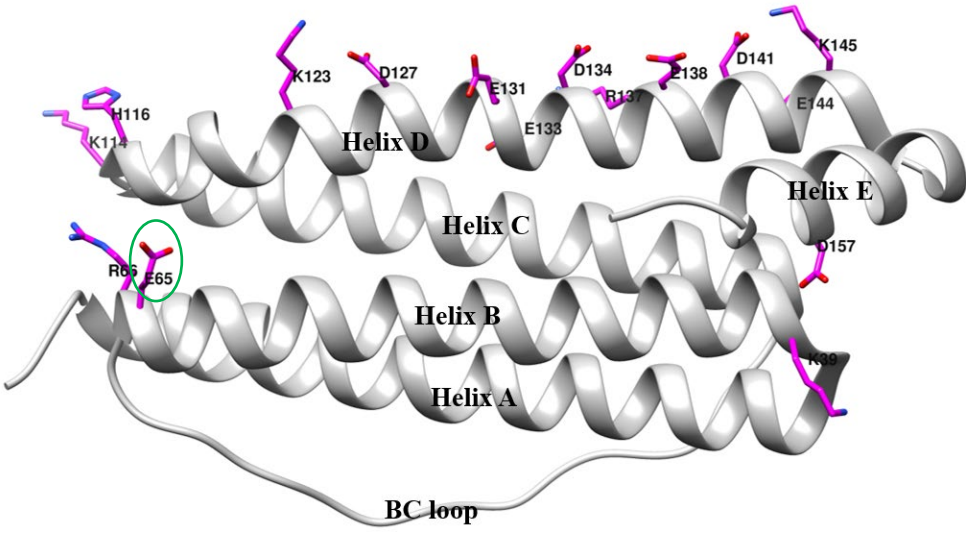


Figure S7. Thermal denaturation profile of different TmFtn mutants. A) Temperature derivative ellipticity measured at 222 nm using CD. B) Thermal denaturation profile measured, intrinsic fluorescence derivative of 330 and 350 nm shown, measured using a NanoTemper Tyco instrument. As observed and in agreement with previous results, E65D, E65A and E65Q result in the weakest and lowest quality cages. This is due to due to the partial dependence on cationic species for these three nanocage variants. The additional mutation at residue 42 may also be a contributing factor in the case of E65D.



Position in monomer	Interface residue
Helix B	K39, E65
Loop BC	R66
Loop CD	K114
Helix D	H116, K123, D127, E131, E133, D134, R137, E138, D141, E144, K145
Helix E	D157

Figure S8. Charged amino acid residues at dimeric interface. Cartoon representation of TmFtn subunit (monomer) along with possible dimeric interface comprising negatively and positively charged residues shown in purple stick representation, where E65 residue marked in green circle. Table shows residue numbers and position in the secondary structure (Helix and loop). Image prepared from PDB ID: 1VLG.

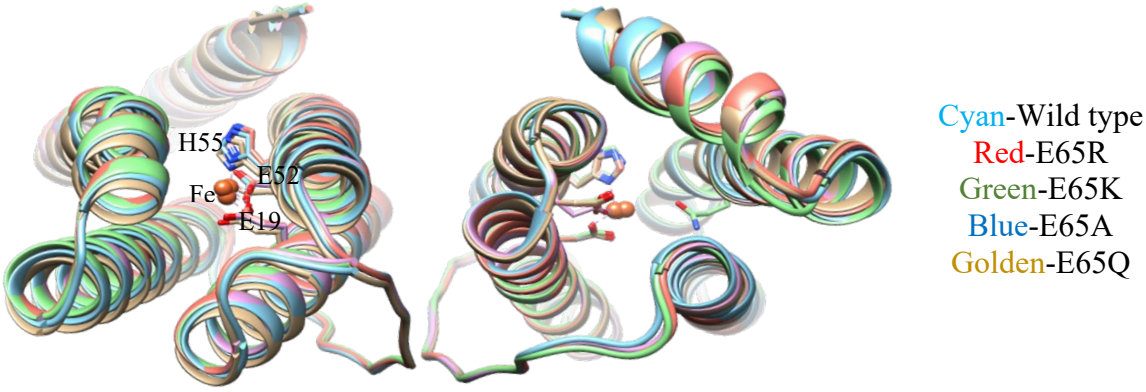


Figure S9 TmFtn mutants core iron binding site superimposition. The figure shows superimposed iron (orange spheres) binding sites of wild type and different TmFtn mutants. The core iron binding site is highly conserved among different mutants and is found to be very similar to WT-TmFtn. Iron binding residues are represented as sticks and labeled in one monomer.

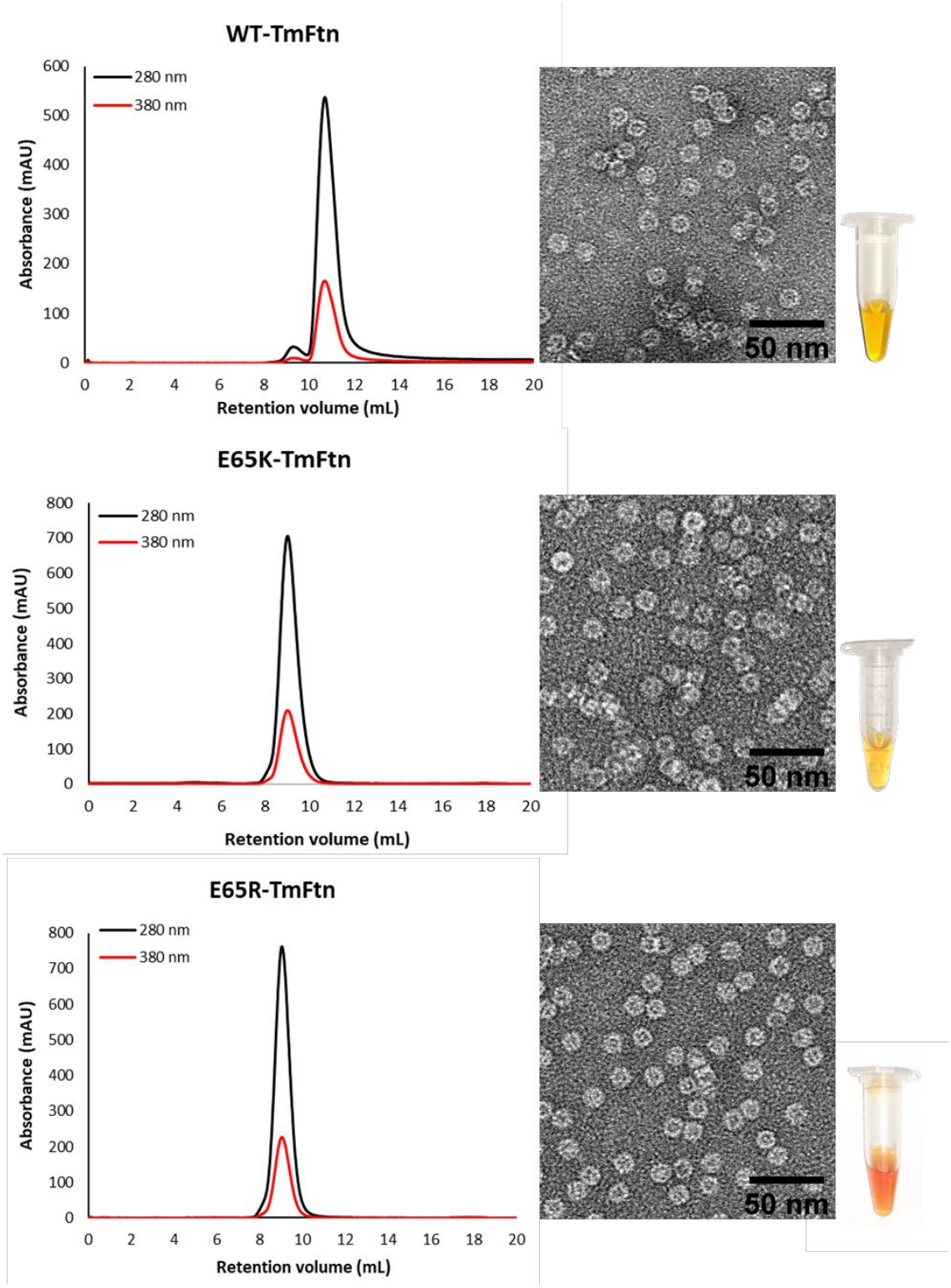


Figure S10. Mineralisation of WT-TmFtn, E65K and E65R mutants. Mineralisation of WT-TmFtn, E65R and E65K was characterised by SEC (graphs) and TEM (insets). In SEC, co-elution of iron can be seen at 380 nm absorbance along with protein absorbance at 280 nm, indicating mineral core formation inside the cage which is seen as a light brown colour solution due to iron oxidation and mineralisation in the cage. Inset shows corresponding TEM images of mineralised TmFtn cages.

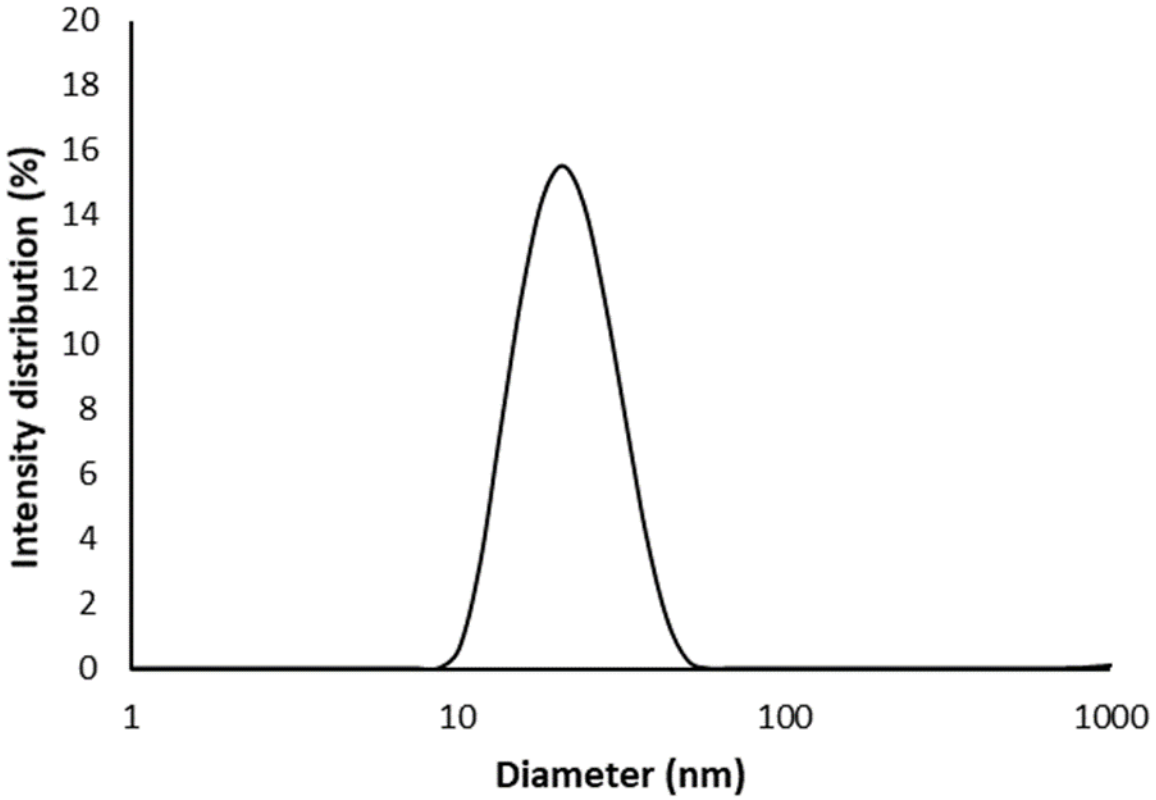


Figure S11. Ammonium sulphate induced assembly of WT-TmFtn. Figure shows results of Dynamic Light Scattering (DLS), measurement carried out in the presence of 1M Ammonium sulphate $(NH_4)_2SO_4$ which induced cage assembly of TmFtn. DLS analysis cage diameter was found to be ~ 15 nm.

Table ST1. Codon optimised (for *E. coli* expression) WT-TmFtn gene sequence with start and stop codons shown in bold.

ATGATGGTGATCAGCGAAAAAGTGCGTAAAGCCCTGAACGATCAACTGAATCGCGAAATTTACAG
CAGCTACCTGTATCTCTCGATGGCGACCTATTTTGATGCGGAAGGTTTTAAAGGCTTTGCCCACTG
GATGAAAAAACAGGCGCAGGAAGAGCTGACCCATGCGATGAAATTTTACGAATATATCTATGAAC
GCGGCGGTCGCGTTGAACTGGAAGCGATTGAAAAACCGCCGAGCAACTGGAACGGCATTAAAGAC
GCGTTTGAAGCCGCGCTGAAACATGAAGAATTTGTCACCCAGAGCATTTATAACATTCTGGAACTG
GCGAGTGAAGAAAAAGATCACGCCACCGTCAGCTTCCTGAAATGGTTTGTTGATGAACAGGTGGA
AGAGGAAGATCAGGTTCGTGAAATCCTGGATCTGCTGGAGAAAGCGAACGGTCAGATGAGCGTGA
TTTTCCAGCTGGACCGTTATCTGGGCCAGCGTGAG**TAA**

Table ST2. Sequence of primers used to generate TmFtn point mutant variants.

TmFtn E65K
TmFtn E65K F5'-CGAATATATCTATAAACGCGGCGGTCGCG-3'
TmFtn E65K R5'-CGCGACCGCCGCGTTTATAGATATATTCG-3'

TmFtn E65R
TmFtn E65R F5'-CGAATATATCTATCGCCGCGGCGGTCGCGTTG-3'
TmFtn E65R R5'-CAACGCGACCGCCGCGGCGATAGATATATTCG-3'

TmFtn E65A
TmFtn E65A F5'-CGAATATATCTATGCACGCGGCGGTCGCG-3'
TmFtn E65A R5'-CGCGACCGCCGCGTGCATAGATATATTCG-3'

TmFtn E65Q
TmFtn E65Q F5'-CGAATATATCTATCAACGCGGCGGTCGCG-3'
TmFtn E65Q R5'-CGCGACCGCCGCGTTGATAGATATATTCG-3'

TmFtn E65D
TmFtn E65D F5'-CGAATATATCTATGACCGCGGCGGTCGCG-3'
TmFtn E65D R5'-CGCGACCGCCGCGGTCATAGATATATTCG-3'

Table ST3. Composition of different pH buffer solutions used in the present study

pH	Buffer composition
2	50 mM glycine-HCl
4	50 mM sodium acetate
6	50 mM potassium phosphate
7.4	20 mM HEPES
8	50 mM Tris-HCl
10	50 mM glycine-NaOH
13	50 mM potassium chloride-NaOH

Table ST4. Physiochemical properties of wild type and different TmFtn mutant variants used in this study

TmFtn Variants	Theoretical mass (Da)	Observed mass (Da)	Estimated Charge at pH 7.4 (monomer)†	Heat denaturation Ti (°C)‡	Interacting partner	$[Mg^{2+}]$ dependence
WT-TmFtn	19378.88	19378.64	-11.8	> 95	E131, D134	Yes, [50 mM] Disassembly to dimer over time observed
E65K	19377.94	19377.73	-9.8	> 95	E131	None
E65R	19405.95	19405.71	-9.8	> 95	E131, D134	None
E65A	19320.84	19320.68	-10.8	90.7	none	Yes [10 mM]
E65Q	19377.89	19377.90	-10.8	88.3	E131, D134	Yes [10 mM]
E65D	19364.85	19392.69*	-11.8	82.2	D134	Yes [10 mM]

† Charge estimated using online program PROTEIN CALCULATOR v3.4 (http://protcalc.sourceforge.net/)
‡ Inflection temperature (Ti) calculated using Tyco nano-temper instrument (nano-DSF) based on protein
*Deviation from theoretical mass due to additional Ala 42 to Val mutation

Table ST5. Secondary structure estimation of different mutant of TmFtn at varying pH. CONTINLL algorithm from the CDPro spectrum decomposition software and the SDP48 reference dataset[1] were used for deconvolution purpose.

		α-helix	β-strand	turn	Unstructured
WT	pH 2.0	91.1%	1.7%	2.2%	4.9%
	pH 4.0	92.3%	1.4%	2.3%	4.1%
	pH 7.4	88.0%	1.2%	3.9%	6.8%
	pH 10.0	77.6%	2.6%	7.5%	12.3%
	pH 13.0	5.9%	13.0%	7.9%	73.2%
E65R	pH 2.0	88.6%	1.9%	3.0%	6.6%
	pH 4.0	78.6%	3.0%	8.7%	9.6%
	pH 7.4	85.1%	3.0%	4.6%	7.3%
	pH 10.0	85.1%	1.8%	4.9%	8.2%
	pH 13.0	8.4%	18.4%	8.9%	64.2%
E65K	pH 2.0	78.8%	3.0%	5.3%	12.9%
	pH 4.0	93.6%	1.0%	2.6%	2.8%
	pH 7.4	91.1%	2.4%	3.0%	3.5%
	pH 10.0	76.9%	3.9%	7.0%	12.2%
	pH 13.0	6.1%	9.7%	6.3%	77.9%
E65A	pH 2.0	91.9%	0.4%	2.6%	5.1%
	pH 4.0	84.4%	1.4%	5.5%	8.7%
	pH 7.4	89.1%	2.6%	3.1%	5.3%
	pH 10.0	89.5%	2.7%	3.0%	4.8%
	pH 13.0	8.0%	15.7%	8.2%	68.1%
E65Q	pH 2.0	95.1%	1.5%	1.5%	1.9%
	pH 4.0	91.2%	1.1%	3.2%	4.4%
	pH 7.4	92.8%	2.9%	2.3%	2.0%
	pH 10.0	95.8%	2.4%	1.5%	0.3%
	pH 13.0	4.1%	2.8%	1.4%	91.7%
E65D	pH 2.0	87.80%	0.80%	3.70%	7.70%
	pH 4.0	86.60%	3.30%	3.00%	7.20%
	pH 7.4	76.90%	3.50%	8.30%	11.30%
	pH 10.0	86.70%	2.60%	4.00%	6.70%
	pH 13.0	8.20%	13.10%	8.80%	70.00%

Table ST6. Crystallographic data collection and refinement statistics.

Data set	Wild Type_TmFtn	TmFtn_E65K	TmFtn_E65R	TmFtn_E65A	TmFtn_E65Q	TmFtn_E65D
PDB ID	6TXN	6TXK	6TXM	6TXI	6TXL	6TXJ
Data collection						
Wavelength (Å)	0.9184	0.895	0.9184	0.9184	0.9184	0.9184
Resolution (Å)	48.8-2.2 (2.33-2.20)	48.25-2.36 (2.50-2.36)	49.01-2.2 (2.33-2.20)	48.94-1.76 (1.87-1.76)	49.12-2.1 (2.22-2.10)	48.25-2.3 (2.44-2.30)
Unique reflections	106131	86438	107042	206460	123375	93349
Multiplicity	10.19	7.52	10.24	10.19	10.04	10.09
Completeness (%)	99.6 (98.6)	99.8 (99.6)	99.8 (99.3)	99.9 (99.6)	99.6 (98.0)	99.8 (99.3)
Mean I/sigma(I)	13.58 (0.63)	8.72 (1.33)	7.52 (0.83)	19.79 (0.94)	10.27 (0.70)	15.22 (0.80)
Wilson B-factors (Å^2)	61.49	40.58	47.18	38.63	51.25	66.89
$CC_{1/2}$ [b]	99.9 (38.6)	99.2 (46.4)	99.6 (40.3)	100.0 (39.2)	99.8 (33.1)	99.9 (50.4)
Space group	R32 (155)	R32 (155)	R32 (155)	R32 (155)	R32 (155)	R32 (155)
Unit cell a, b, c (Å)	176.17 176.17 351.71	175.81 175.81 354.14 90.0	176.26 176.26 353.78 90.0	175.70 175.70 353.57 90.0	176.22 176.22 355.04 90.0	176.14 176.14 352.64 90.0
α, β, γ (°)	90.0 90.0 120.0	90.0 120.0	90.0 120.0	90.0 120.0	90.0 120.0	90.0 120.0
Refinement						
R-work[c]	0.2044	0.2011	0.2039	0.1801	0.1925	0.2120
R-free[d]	0.2343	0.2365	0.2433	0.2045	0.2218	0.2410
No. of atoms	11598	11850	11671	12379	11921	11410
Macromolecules	11016	11088	10459	10944	11137	10970
Ligands	228	320	278	234	313	305
Solvent	354	442	308	1201	471	135
RMS (bonds) (Å)	0.008	0.003	0.003	0.012	0.006	0.006
RMS (angles) (°)	1.066	0.582	0.626	1.238	0.694	0.855
Ramachandran						
favored (%)	99.54	99.54	99.46	99.61	99.61	99.69
outliers (%)	0	0	0	0	0	0
Average B-factors (Å^2)	77.01	51.41	62.73	50.55	66.90	86.25
Macromolecules	76.64	50.58	61.81	49.05	66.26	85.82
Ligands	98.64	76.57	97.89	79.45	88.77	107.07
Solvent	74.37	54.02	63.87	58.62	67.50	74.21

[a]Numbers in parentheses refer to data in the highest resolution shell.

[b]The CC1/2 is the Pearson correlation coefficient (CC) calculated from each subset containing a random half of the measurements of unique reflection

[c]Rwork = Σ ||Fobs|-|Fcal||/ Σ|Fobs|

[d]Rfree is the same as Robs for a selected subset (5%) of the reflections that was not included in prior refinement calculations

Table ST7. Metal binding core, 3-fold and 4-fold axis residues along with structural differences between them summarised.

TmFtn variants	Metal	Metal binding residues (within 3Å)	3-fold axis residues	4-fold axis residues	Dimer RMSD (Å)
WT-TmFtn[2]	Fe	E19, E52, H55	S111, V119, S120, K123	Q149, S151, Q155	-
E65K	Fe	E19, E52, H55	S111, V119, S120, K123	Q149, S151, Q155	0.26
E65R	Fe	E19, E52, H55	S111, V119, S120, K123	Q149, S151, Q155	0.26
E65A	Fe	E19, E52, H55	S111, V119, S120, K123	Q149, S151, Q155	0.26
E65Q	Fe	E19, E52, H55	S111, V119, S120, K123	Q149, S151, Q155	0.26
E65D	Fe	E19, E52, H55	S111, V119, S120, K123	Q149, S151, Q155	0.31

Supplementary References:

1. Sreerama N., *et al.* Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Inclusion of Denatured Proteins with Native Proteins in the Analysis. *Anal. Biochem.* vol. 287 (**2**) 243-251 (2000)
2. J. Robin Harris, Jon Marles-Wright, in Macromolecular Protein Complexes: Structure and Function, ed. Alejandro Yevenes, Springer, (Springer Nature Switzerland), Mar 7, 2017, Chapter 3, The Ferritin Superfamily (pp: 75-102).