Supporting Information for

Improved repeat protein stability by combined consensus and computational protein design

Erich Michel[1], Stefano Cucuzza[2†], Peer R.E. Mittl[1], Oliver Zerbe[2], Andreas Plückthun[1]**

[1]Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

[2]Department of Chemistry, University of Zurich, CH-8057 Zurich, Switzerland

*Email: e.michel@bioc.uzh.ch

*Email: plueckthun@bioc.uzh.ch

[†]Present address: Novartis Pharma AG, CH-4057 Basel, Switzerland

Table of Contents

Experimental details

a. Cloning of target genes

dArmRPs are composed of N- and C-terminal capping repeats, where in our nomenclature the subscripts refer to the version of the cap, and internal M repeats, where the subscript indicates the number of M repeats. Cloning of target genes and protein expression and purification is described in the supplementary information. All genes encoding dArmRPs were PCR-amplified from a codon-optimized $N_{YIII}M_3C_{AII}$ gene[1] using the oligonucleotide primer and template DNA combinations listed in Tables S1 and S2. PCR products encoding dArmRPs with one internal repeat were cloned into the expression vector pEM3BT2[1] using the SapI/BamHI restriction sites. Genes encoding dArmRPs with four internal M repeats were assembled by ligation of a 5'- and a 3'-PCR product, separately digested with XbaI/SapI and SapI/BamHI, respectively, into XbaI/BamHI-digested pEM3BT2. All constructs were cloned as fusion constructs to an N-terminal $(His)_6$-tagged GB1 domain, which is separated by a flexible linker encoding a TEV-protease cleavage site for facile proteolytic removal of the N-terminal $(His)_6$-GB1.[2] The expression plasmid pEM3BTC, which encodes a HRV 3C-protease cleavage site in the linker between $(His)_6$-GB1 and the target gene, was generated by mutagenesis PCR of the pEM3BT2 plasmid[1] using the 3BTC_Fwd and 3BTC_Rev oligonucleotide primers. The MNG-3BTC plasmid for expression of target peptides fused to mNeonGreen (MNG) was prepared by ligation of the SapI/BglII-digested PCR product encoding mNeonGreen into SapI/BamHI-digested pEM3BTC. Complementary oligonucleotides encoding the $(KR)_5$-target peptide were annealed after heating to 95°C by passive cooling to 25°C and were subsequently introduced into MNG-3BTC using the BamHI/BsaI restriction sites. The single Cys-variants E15C, Q92C and S221C of $N_{A4}MC_{AII}$, required for the site-specific attachment of dia- and paramagnetic tags, were prepared by mutagenesis as previously described.[3]

b. Protein expression and purification

All proteins were expressed in *E. coli* BL21-Gold (DE3) cells (Agilent Technologies) growing at 37°C with shaking in 200 mL 2YT medium. Expression was induced with 1 mM IPTG at an OD_{600} of ca. 0.6–0.8 for ca. 16 h at 30°C. [^{13}C,^{15}N]-labeled proteins for NMR analysis were also expressed using *E. coli* BL21-Gold (DE3) cells but grown in minimal medium.[4] After harvesting by centrifugation, the obtained cell pellets were resuspended in 15 mL buffer A (50 mM sodium phosphate at pH 7.7, 500 mM sodium chloride, 20 mM imidazole, 30 µM sodium

azide) supplemented with 5 mM magnesium sulfate, 1 mg/mL hen egg white lysozyme (Sigma-Aldrich) and 0.05 mg/mL DNaseI (Roche). Cells were lysed with a Branson Ultrasonics 250 Sonifier (Branson Ultrasonics) for 3 min on ice using a duty cycle of 70% and an output power of 4. Insoluble debris was subsequently removed by centrifugation and the supernatant was filtered through a 0.2 µm sterile syringe filter unit (Sartorius) before purification on a 5 mL HisTrap HP column as previously described.[5] The N-terminal $(His)_6$-GB1 fusion was then removed by proteolytic cleavage with 2 mg TEV protease in case of dArmRPs and with 1 mg HRV 3C protease for the $(KR)_5$-mNeonGreen fusion.[5] After separation of the target protein from $(His)_6$-tagged species by re-application on a 5 mL HisTrap HP column (GE Healthcare), the purified proteins were dialyzed against NMR buffer (20 mM sodium phosphate, 50 mM sodium chloride, 30 µM sodium azide) and concentrated in 3 kDa MWCO ultrafiltration devices (Merck Millipore). Proteins intended for affinity measurements by fluorescence anisotropy were dialyzed against PBS (50 mM sodium phosphate at pH 7.4, 150 mM sodium chloride, 30 µM sodium azide). The $N_{A4}M_4C_{AII}$ construct prepared for crystallization was additionally purified by size exclusion chromatography on a HiLoad 26/60 Superdex 75 column (GE Healthcare), equilibrated in 10 mM Tris-HCl at pH 7.6, prior to concentration in a 10 kDa MWCO ultrafiltration device (Merck Millipore).

TEV protease was prepared as previously described.[2] HRV 3C protease in pET24b was expressed in *E. coli* BL21-Gold (DE3) cells growing in 1 L 2YT medium with shaking at 25°C. Protein expression was induced at OD_{600} of 0.6 with 0.5 mM IPTG for 16 h. Cells were harvested as described above and were resuspended in 40 mL buffer A-3C (40 mM HEPES-NaOH at pH 8, 300 mM sodium chloride, 20 mM imidazole, 1 mM DTT, 10% (v/v) glycerol) and lysed with a Branson Ultrasonics Sonifier 250 for 10 min on ice with a duty cycle of 30% and an output level of 4. Clearing of the sample was performed as described above and the filtered sample was applied on a 5 mL HisTrap HP column in buffer A-3C. After washing with 15 column volumes of buffer A-3C, the HRV 3C protease was eluted with a 100 mL linear gradient of buffer A-3C to buffer B-3C (same as buffer A-3C but containing 300 mM imidazole) and dialyzed overnight in a 12–14 kDa MWCO dialysis membrane (Spectrum Labs) at 4°C against 2 L of buffer 3C (10 mM HEPES-NaOH at pH 8, 150 mM sodium chloride, 5 mM EDTA, 1 mM DTT, 10% (v/v) glycerol). The protein solution was then further supplemented with glycerol to a final concentration of 20% (v/v) glycerol, and aliquots containing 2 mg HRV 3C protease were flash-frozen in liquid nitrogen and stored at -80°C.

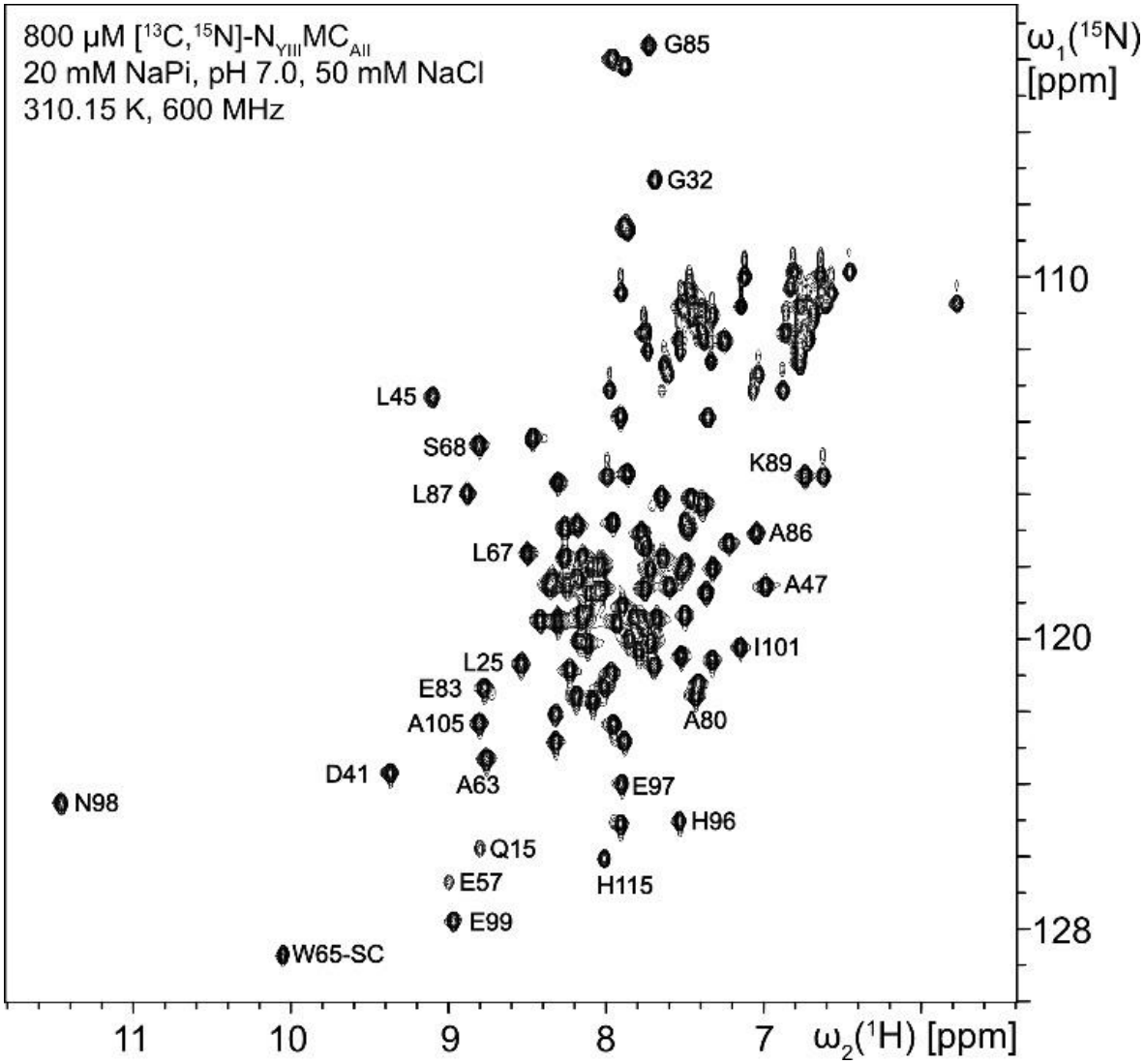

Figure S1. 2D [^{1}H,^{15}N]-HSQC spectrum of [^{13}C,^{15}N]-$N_{YIII}MC_{AII}$ indicates a unique and well-folded population. The data were recorded at 37°C on a 600 MHz spectrometer using 800 µM dArmRP in 20 mM sodium phosphate at pH 7 containing 50 mM sodium chloride.

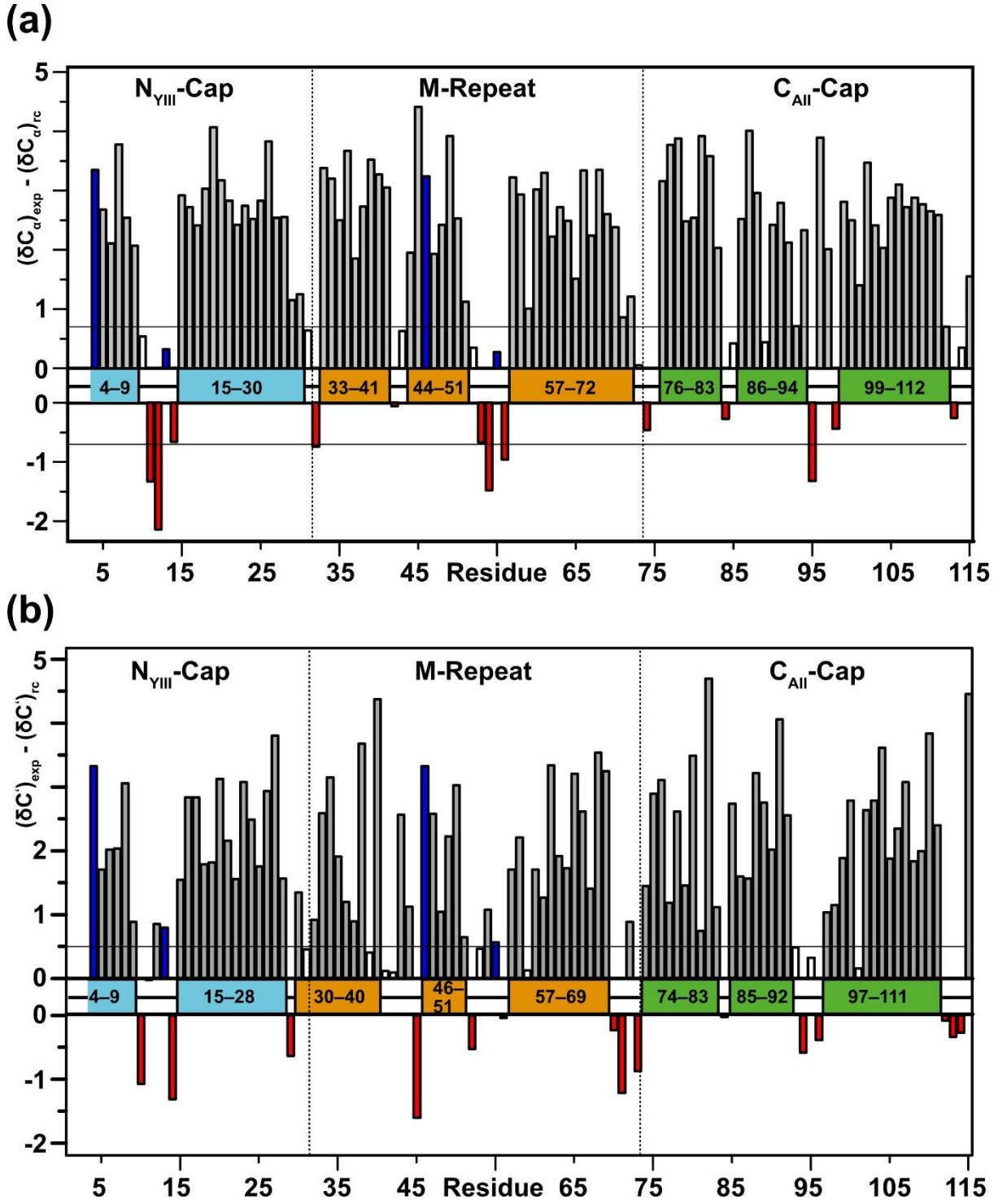

Figure S2. Secondary structure of $N_{YIII}MC_{AII}$ from chemical shift indices. Secondary chemical shifts derived from assigned C_α (**a**) and C' (**b**) spins of $N_{YIII}MC_{AII}$. Red bars indicate residues with secondary shift values that oppose α-helix formation while blue bars indicate proline residues. The lines at ordinate values of 0.7 (a) or 0.5 (b) indicate thresholds to define helical residues from C_α and C' chemical shifts, respectively. Segments forming regular α-helices are schematically shown as colored boxes.

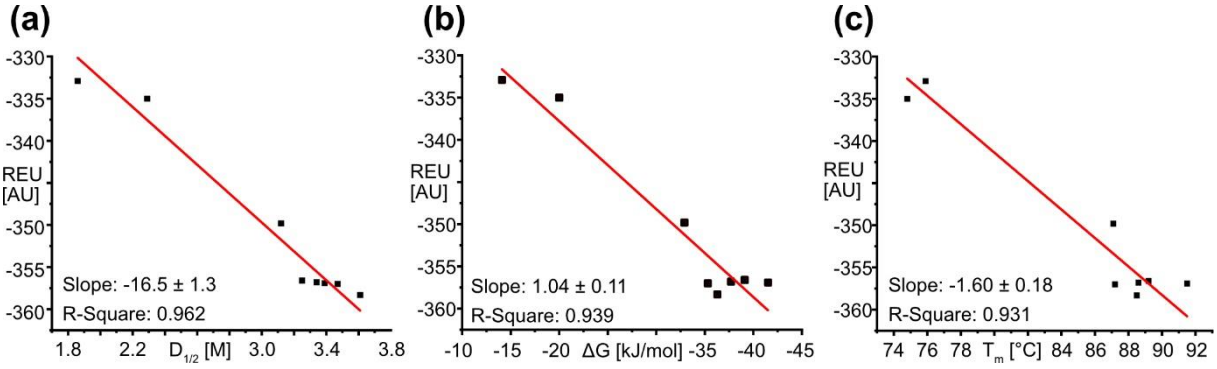

Figure S3. Correlation of Rosetta energy units with experimental NMC protein stabilities. Correlation of the Rosetta energy units (REU) with the experimentally determined (**a**) D_m and (**b**) ΔG values obtained from denaturant-induced unfolding and (**c**) T_m values from thermal unfolding of NMC proteins with different N-caps.

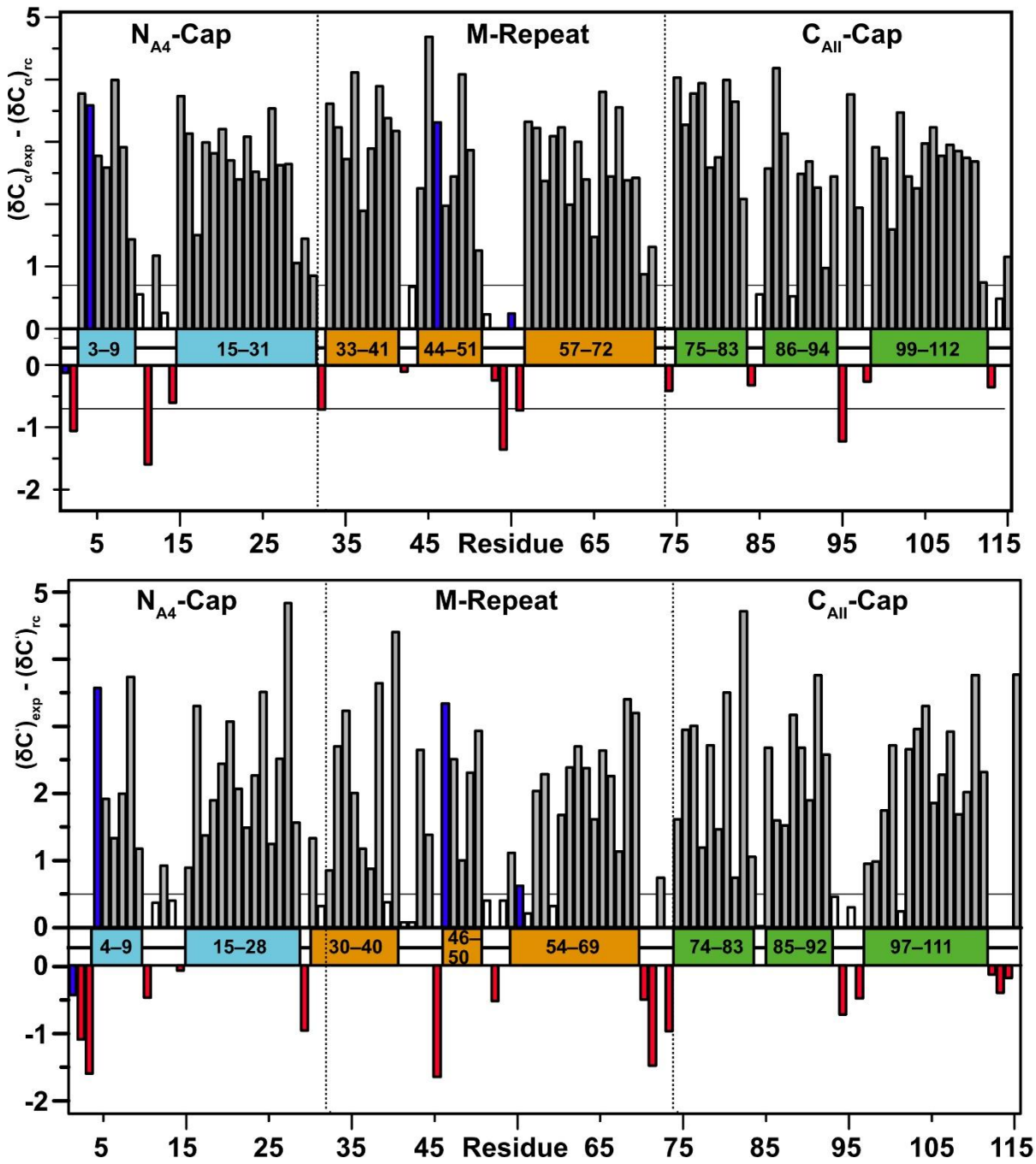

Figure S4. Secondary structure of $N_{A4}MC_{AII}$ from chemical shift indices. Secondary chemical shifts derived from assigned C_α (**a**) and C' (**b**) spins of $N_{A4}MC_{AII}$. Red bars indicate residues with secondary shift values that oppose α-helix formation while blue bars indicate proline residues. The lines at ordinate values of 0.7 (a) or 0.5 (b) indicate thresholds to define helical residues from C_α and C' chemical shifts, respectively. Segments forming regular α-helices are schematically shown as colored boxes.

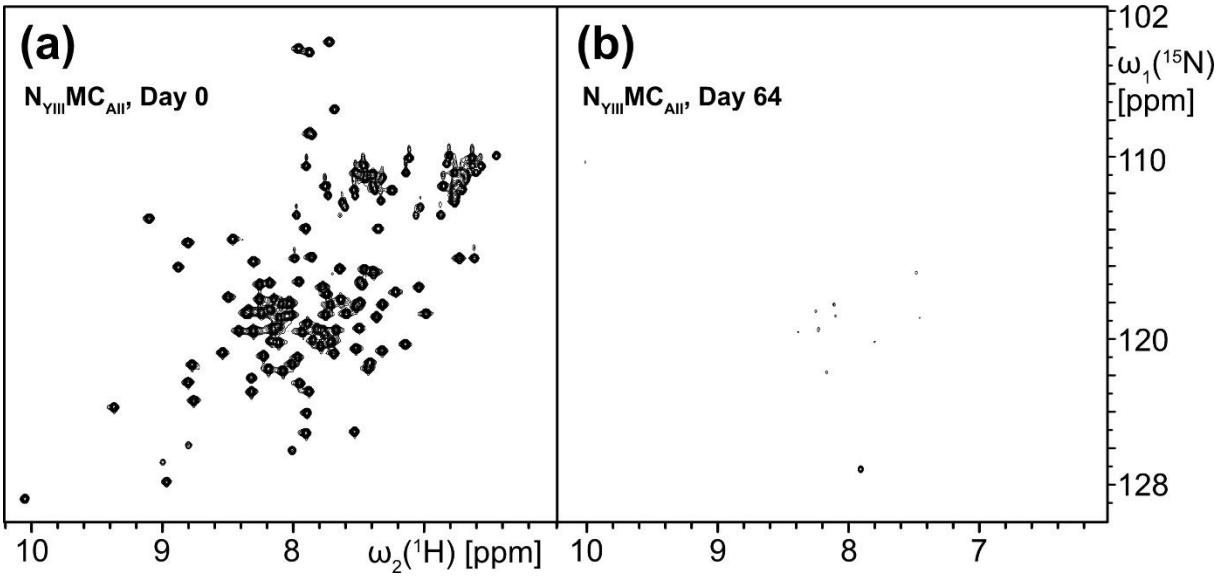

Figure S5. [^{15}N,^{1}H]-HSQC spectra of 100 µM $N_{YIII}MC_{AII}$ in PBS buffer at pH 7 recorded at day 0 (**a**) and at day 64 (**b**) after incubation at 37°C. Both spectra were recorded at 37°C and 600 MHz using identical measurement and processing parameters.

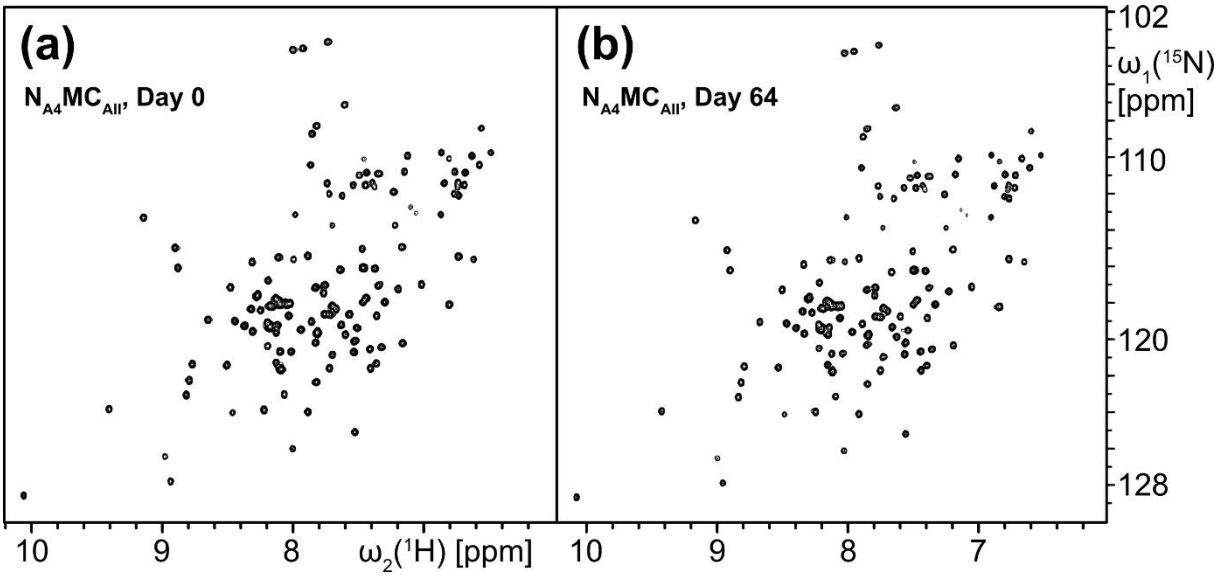

Figure S6. [^{15}N,^{1}H]-HSQC spectra of 100 µM $N_{A4}MC_{AII}$ in PBS buffer at pH 7 recorded at day 0 (**a**) and at day 64 (**b**) after incubation at 37°C. Both spectra were recorded at 37°C and 600 MHz using identical measurement and processing parameters.

Table S1. Cloning of target genes and expression plasmids

Construct name	Oligonucleotides	Template DNA for PCR	Recipient Plasmid
$N_{H23}MC_{AII}$	H23MC_Fwd/H23MC_Rev	$N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{YIII}MC_{AII}$	M3_Fwd/Y_Rev	$N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{SH2}MC_{AII}$	V1_Fwd/V1_Rev	$N_{H23}MC_{AII}$	pEM3BT2
$N_{A4}MC_{AII}$	V41_Fwd/V41_Rev	$N_{SH2}MC_{AII}$	pEM3BT2
$N_{A5}MC_{AII}$	V42_Fwd/V41_Rev	$N_{SH2}MC_{AII}$	pEM3BT2
$N_{A6}MC_{AII}$	V5_Fwd/V5_Rev	$N_{H23}MC_{AII}$	pEM3BT2
$N_{A7}MC_{AII}$	V6_Fwd/V6_Rev	$N_{SH2}MC_{AII}$	pEM3BT2
$N_{A8}MC_{AII}$	V5_Fwd/V6_Rev	$N_{SH2}MC_{AII}$	pEM3BT2
$N_{A9}MC_{AII}$	V5_Fwd/V8_Rev	$N_{SH2}MC_{AII}$	pEM3BT2
$N_{H23}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{H23}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{YIII}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{YIII}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{A4}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{A4}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{A5}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{A5}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{A6}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{A6}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{A7}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{A7}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{A8}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{A8}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
$N_{A9}M_4C_{AII}$	T7/M3_R + M1_F/T7T	$N_{A9}MC_{AII}$ + $N_{YIII}M_3C_{AII}$	pEM3BT2
MNG-3BTC	mNG-3BTC_F/mNG-3BTC_R	mNeonGreen	pEM3BTC
$(KR)_5$-mNeonGreen	KR5_Top/KR5_Bot	-	mNeonGreen-3BTC

Table S2. Oligonucleotide primers used in this study

Name	Sequence
H23MC_Fwd	5′-AAAGCTCTTCACAGGGCGCCCTTCCAGCCC
H23MC_Rev	5′-GCTTTGTTAGCAGCCGGATC
3BTC_Fwd	5′-CGAAAGCAGCGGCCTGGAAGTGCTGTTTCAGGGTCCGAGAAGAGCCATGGC
3BTC_Rev	5′-GCCATGGCTCTTCTCGGACCCTGAAACAGCACTTCCAGGCCGCTGCTTTCG
mNG-3BTC_F	5′-AAAGCTCTTCACCGGGATCCAAAAGTGGTCTCGGCGCCGGCTCGAAGGGGGAAGAAGATAAC
mNG-3BTC_R	5′-AAAAGATCTTTATTACTTATAAAGCTCATCCATGCCC
Y_Rev	5′-AAAGCTCTTCAACCGCTTGCAATCTGTGAGAG
M3_Fwd	5′-AAAGCTCTTCAGGCGGTAACGAGCAGATTCAGGC
V1_Fwd	5′-AAAGCTCTTCAGTGAAGTTACTGAAAAGCTCTAACGAACAGATTCTCCAAGAGG
V1_Rev	5′-AAAGCTCTTCACACCAGTTTAGGCAGATCCGGACCCTGGAAGTACAGGTTTTCGC
V41_Fwd	5′-AAAGCTCTTCACTGCGTGCACTCGCTGAAATTGCCAGCGGCGGTAACGAGCAGATTC
V41_Rev	5′-AAAGCTCTTCACAGCGCTTTCAGCAGGATTTCCTCGTTAGAGCTTTTCAGTAACTTCACC
V42_Fwd	5′-AAAGCTCTTCACTGAAGGCACTCGCTGAAATTGCCAGCGGCGGTAACGAGCAGATTC
V5_Fwd	5′-AAAGCTCTTCACTGAAGACACTCGCTGAAATTGCCAGCGGCGGTAACGAGCAGATTC
V5_Rev	5′-AAAGCTCTTCACAGCGCTTTCAGCAGGGTTTCCTCATCAGGTGACGAAAGCAATTGGAC
V6_Fwd	5′-AAAGCTCTTCACTGCGTACACTCGCTGAAATTGCCAGCGGCGGTAACGAGCAGATTC
V6_Rev	5′-AAAGCTCTTCACAGCGCTTTCAGCAGGGTTTCCTCATCAGAGCTTTTCAGTAACTTCACC
V8_Rev	5′-AAAGCTCTTCACAGCGCTTCCAGCAGGGTTTTCTCATCAGAGCTTTTCAGTAACTTCACC
M3_R	5′-AAAGCTCTTCACCCACCAGAGGCAATGTTAG
M1_F	5′-AAAGCTCTTCAGGGAATGAGCAAATCCAAGCCGTG
T7	5′-TAATACGACTCACTATAGGG
T7T	5′-GCTAGTTATTGCTCAGCGG

Table S3. Data collection and refinement statistics of $N_{A4}M_4C_{AII}$:lysozyme

	PDB ID: 7QNP
Wavelength	1.000
Resolution range (Å)	41.06–1.59 (1.65–1.59)
Space group	$P2_12_12_1$
Unit cell	
a, b, c (Å)	56.59, 62.66, 108.74
α, β, γ (°)	90, 90, 90
Total reflections	706321 (71586)
Unique reflections	52752 (5191)
Multiplicity	13.4 (13.8)
Completeness (%)	99.96 (99.98)
Mean I/sigma(l)	22.36 (1.48)
Wilson B-factor	28.93
R-merge	0.056 (1.382)
R-meas	0.059 (1.435)
R-pim	0.016 (0.384)
CC1/2	1 (0.702)
CC*	1 (0.908)
ISa	30.57
Reflections used in refinement	52751 (5190)
Reflections used for R-free	2637 (259)
R-work	0.186 (0.437)
R-free	0.214 (0.423)
CC(work)	0.964 (0.811)
CC(free)	0.948 (0.761)
Number of non-hydrogen atoms	3282
Macromolecules	2922
Ligands	34
Solvent	326
Protein residues	369
RMS (bonds)	0.029
RMS (angles)	1.92
Ramachandran favored (%)	99.45
Ramachandran allowed (%)	0.55
Ramachandran outliers (%)	0.00
Rotamer outlier (%)	0.32
Clashscore	6.24
Average B-factor	36.42
Macromolecules	35.01
Ligands	52.50
Solvent	47.43
Number of TLS groups	2

Statistics for the highest-resolution shell are shown in parentheses.

Table S4. Computational stability scanning mutagenesis of individual N_{H23}-cap residues in $N_{H23}MC_{AII}$ using the Rosetta software suite. Rosetta energy unit (REU) differences in NMC proteins resulting from single mutations after energy minimization are shown.

$N_{H23}MC_{AII}$ Residue	Rosetta Suggestion	ΔREU (Rosetta-Original)
Gly	Pro	-0.663
Ala	Asp	1.179
Leu	Leu	-
Pro	Pro	-
Ala	Lys	-0.51
Leu	Leu	-
Val	Val	-
Gln	Lys	-0.438
Leu	Leu	-
Leu	Leu	-
Ser	Lys	-0.707
Ser	Ser	-
Pro	Asn	0.15
Asn	Asp	0.987
Glu	Glu	-
Gln	Lys	-0.229
Ile	Glu	-0.289
Leu	Leu	-
Gln	Leu	-1.801
Glu	Glu	-
Ala	Ala	-
Leu	Leu	-
Trp	Arg	-3.487
Ala	Thr	0.178
Leu	Leu	-
Ser	Ala	-1.926
Asn	Val	-0.805
Ile	Ile	-
Ala	Ala	0.004
Ser	Ser	-0.007

Table S5. Rosetta energy differences at individual N_{YIII}- and N_{A4}-cap positions.

	N_{YIII}-Cap		N_{A4}-Cap		ΔREU
Position	Residue	[REU]	Residue	[REU]	N_{A4}-N_{YIII}
1	Gly	-0.12	Pro	1.74	1.86
2	Glu	-1.13	Asp	-0.35	0.78
3	Leu	-1.85	Leu	-3.26	-1.41
4	Pro	-0.60	Pro	0.28	0.88
5	Gln	1.78	Lys	0.93	-0.85
6	Met	-1.72	Leu	-5.14	-3.42
7	Val	-5.72	Val	-6.02	-0.30
8	Gln	1.60	Lys	1.27	-0.33
9	Gln	-1.14	Leu	-3.90	-2.76
10	Leu	-5.52	Leu	-5.22	0.30
11	Asn	0.10	Lys	0.96	0.86
12	Ser	-1.08	Ser	-0.47	0.61
13	Pro	0.90	Ser	-0.85	-1.75
14	Asp	-1.12	Asn	-1.80	-0.68
15	Gln	0.41	Glu	0.20	-0.21
16	Gln	1.29	Glu	1.53	0.24
17	Glu	-1.08	Ile	-2.90	-1.82
18	Leu	-4.69	Leu	-4.48	0.21
19	Gln	0.21	Leu	-3.32	-3.53
20	Ser	0.55	Lys	-0.10	-0.65
21	Ala	-4.35	Ala	-4.98	-0.63
22	Leu	-6.30	Leu	-6.70	-0.40
23	Arg	0.81	Arg	0.18	-0.63
24	Lys	-0.09	Ala	-3.81	-3.72
25	Leu	-6.67	Leu	-6.84	-0.17
26	Ser	-1.73	Ala	-6.97	-5.24
27	Gln	0.79	Glu	1.17	0.38
28	Ile	-2.00	Ile	-3.20	-1.20
29	Ala	-4.04	Ala	-4.07	-0.03
30	Ser	0.45	Ser	0.64	0.19

Table S6. D_m, ΔG and m values from denaturant-induced unfolding

Construct	D_m [M]	ΔG [kJ/mol]	m [kJ/mol×M]
N_{YIII}MC	1.86	-14.1 ± 0.9	7.6 ± 0.5
N_{H23}MC	2.29	-20.0 ± 2.0	8.8 ± 0.9
N_{A4}MC	3.61	-32.9 ± 4.0	9.1 ± 1.1
N_{A5}MC	3.47	-39.1 ± 4.7	11.3 ± 1.4
N_{A6}MC	3.12	-37.7 ± 5.1	12.1 ± 1.6
N_{A7}MC	3.39	-41.5 ± 8.3	12.2 ± 2.4
N_{A8}MC	3.25	-35.3 ± 5.3	10.9 ± 1.6
N_{A9}MC	3.34	-36.3 ± 7.1	10.9 ± 2.1

Table S7. Affinities of NM_4C proteins to $(KR)_5$-peptides

NM_4C variant	K_d ± St.Dev. [nM]
N_{YIII}-M_4C	36.1 ± 2.9
N_{A4}-M_4C	30.5 ± 2.3
N_{A5}-M_4C	48.6 ± 10.7
N_{A6}-M_4C	29.9 ± 5.6
N_{A7}-M_4C	28.7 ± 6.4
N_{A8}-M_4C	22.9 ± 5.1
N_{A9}-M_4C	45.1 ± 3

References

[1] Michel, E., Plückthun, A., and Zerbe, O. (2018) Peptide-guided assembly of repeat protein fragments, *Angew. Chem. Int. Ed. 57*, 4576–4579.

[2] Michel, E., and Wüthrich, K. (2012) High-yield *Escherichia coli*-based cell-free expression of human proteins, *J. Biomol. NMR 53*, 43–51.

[3] Cucuzza, S., Güntert, P., Plückthun, A., and Zerbe, O. (2021) An automated iterative approach for protein structure refinement using pseudocontact shifts, *J. Biomol. NMR 75*, 319-334.

[4] Michel, E., and Allain, F. H. (2015) Selective amino acid segmental labeling of multi-domain proteins, *Methods Enzymol. 565*, 389–422.

[5] Michel, E., Plückthun, A., and Zerbe, O. (2019) Peptide binding affinity redistributes preassembled repeat protein fragments, *Biol. Chem. 400*, 395-404.