Supplementary Material for

Computationally Designed Armadillo Repeat Proteins for Modular Peptide Recognition.

Christian Reichen[a], Simon Hansen[a], Cristina Forzani[a], Annemarie Honegger[a], Sarel J. Fleishman[b,c], Ting Zhou[a], Fabio Parmeggiani[a,e], Patrick Ernst[a], Chaithanya Madhurantakam[a,f], Christina Ewald[d], Peer R. E. Mittl[a], Oliver Zerbe[d], David Baker[b], Amedeo Caflisch[a] and Andreas Plückthun[a,*]

[a] Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland

[b] Department of Biochemistry, University of Washington, Seattle, USA

[c] Current address: Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

[d] Institute of Organic Chemistry, University of Zurich, 8057 Zurich, Switzerland

[e] Current Address: Life Science Building, University of Bristol, Bristol BS8 1TH, United Kingdom

[f] Current Address: Department of Biotechnology, TERI University, 10, Institutional Area, New Delhi 110070, India

* Corresponding author.
Address: Department of Biochemistry,
University of Zurich, Winterthurerstrasse 190,
8057 Zurich, Switzerland.
Phone: +41 44 635 55 70.
Fax: +41 44 635 57 12
E-mail: plueckthun@bioc.uzh.ch

Supplementary Tables

Supplementary Table ST1 Biophysical properties of dArmRPs

Constructs[a]	Short name	Residues (repeats)[b]	pI[c]	MW_{cal} (kDa)[d]	MW_{obs} (kDa)[e]	Oligomeric state[f]	$MW_{obs/cal}$ [g]	CD_{222} (MRE)[h]	T_m (°C)[i]	CD_{GdnHCl} (M)[j]
$Y_I(D_{PPSA})_4A_I$		253 (6)	5.4	27.2	53.1	monomer	1.0	-14,637	n.a.	~1.0
$Y_I(D_{PASA})_4A_I$		253 (6)	5.4	26.9	48.2	monomer	1.0	-14,933	n.a.	n.d.
$Y_I(D_{PAVA})_4A_I$		253 (6)	5.4	26.9	55.7	intermediate[e]	1.3	-13,107	n.a.	~2.9
$Y_I(D_{PAAA})_4A_I$		253 (6)	5.4	26.8	45.6	monomer	1.0	-15,615	n.a.	~2.2
$Y_I(D_{PAAL})_4A_I$		253 (6)	5.4	27.0	46.5	monomer	0.9	-14,479	n.a.	~2.2
$Y_I(D_{PAAC})_4A_I$		253 (6)	5.4	26.9	48.1	monomer/dimer	1.0 / 1.9	-14,834	n.a.	n.d.
$Y_I(D_{SPAA})_4A_I$		253 (6)	5.3	26.9	53.3	monomer	1.0	-14,123	n.a.	n.d.
$Y_I(D_{SPVA})_4A_I$		253 (6)	5.4	27.0	72.1	intermediate[e]	1.4	-12,964	81 ± 3	~2.5
$N_S(D_{SPVA})_4C_{PAF}$		251 (6)	5.3	26.3	32.4	monomer	1.0	-14,108	92 ± 2	~2.4
$N_S(D_{SPVA})_4C_{PAI}$		251 (6)	5.3	26.3	32.4	monomer	1.0	-13,458	91 ± 1	~2.4
$N_S(D_{SPVA})_4C_{SPI}$		251 (6)	5.3	26.3	33.2	monomer	1.0	-13,813	88 ± 2	~2.3
$N_V(D_{SPVA})_4C_{PAF}$	CAR0	251 (6)	5.3	26.3	30.4	monomer	1.0	-15,121	93 ± 2	2.2
$N_V(D_{SPVA})_4C_{PAI}$		251 (6)	5.3	26.3	30.5	monomer	0.9	-14,348	98 ± 7	~2.5
$N_V(D_{SPVA})_4C_{SPI}$		251 (6)	5.3	26.3	30.4	monomer	1.0	-13,557	~110	~2.4
$N_A(D_{SPVA})_4C_{PAF}$		251 (6)	5.3	26.3	31.8	monomer	1.0	-13,560	~129	~2.0
$N_A(D_{SPVA})_4C_{PAI}$		251 (6)	5.3	26.2	31.7	monomer	1.0	-13,363	99 ± 18	~2.4
$N_A(D_{SPVA})_4C_{SPI}$		251 (6)	5.3	26.3	32.4	monomer	1.0	-13,450	89 ± 3	~2.3
$Nq(D_{SPVA})_4C_{PAF}$		251 (6)	5.3	26.3	31.3	monomer	1.0	-12,470	n.a.	1.6
$N_V(Dq)_4C_{PAF}$	CAR1	251 (6)	5.0	26.3	31.9	monomer	0.9	-16,824	73 ± 0.2	2.8
$Y_{III}(Dq)_4C_{PAF}$	CAR2	251 (6)	5.1	26.3	31.9	monomer	0.9	-14,997	73 ± 0.5	3.0
$Y_{III}(D_{SPVA})_4C_{PAF}$		251 (6)	5.4	26.3	30.6	monomer	0.9	-13,325	92 ± 3	2.0
$Y_{III}(D_{SPVA})_4A_{II}$		252 (6)	5.5	26.6	31.7	monomer	0.9	-16,316	84 ± 2	1.6
$Y_IM_4A_I$ [k]		253 (6)	4.7	27.1	34.4	monomer	1.0	-15,474	73 ± 0.5	~3.4
$Y_{III}M_4A_{II}$ [k]		252 (6)	4.8	26.8	56.9	intermediate[e]	1.4	-17,056	87 ± 1.5	3.9

[a] dArmR: N-cap (e.g., N_{VA} [pos. 17 and 32] and Y_{III} [k]), C-cap (e.g., C_{PAF} and A_{II} [k]), and internal repeats (e.g., D_{SPVA} and Dq).
[b] The number of residues includes the $MRGSH_6GS$ tag; the number of repeats includes capping repeats.
[c] Isoelectric point (pI).
[d] Molecular weight calculated from the sequence
[e] Observed molecular weight as determined by SEC.
[f] Oligomeric state as indicated by multi-angle static light scattering (MALS). Intermediate state: equilibrium between monomer and dimer.
[g] Ratio between observed (by MALS) and calculated molecular weight ($MW_{obs/calc}$).
[h] Mean residue ellipticity at 222 nm expressed as deg·cm^2/dmol.
[i] Transition midpoint (T_m) observed in thermal denaturation measured by CD. Approximations (~) due to high melting point.
[j] Midpoint of transition in GdnHCl-induced denaturation, measured by CD. Approximations (~) due to limited data points.
[k] Consensus-based ArmRP. M refers to the consensus-based internal repeat $\overline{M}$, reported by Alfarano *et al.*[1]
n.a.: not available
n.d.: not determined

Supplementary Table ST2 Curvature parameters of model and crystal structures of ArmRPs

Name	Origin[a]	Design	Rise h (Å) [b]	Radius r (Å) [b]	Angle 2·Ω (°) [b]	Cα(P/P+2) (Å) [b]	RMSD (Å) [b,c]
GH repeat	Structure	nArmRP (1EE4) [d]	6.2 ± 0.0	15.7 ± 0.1	29.3 ± 0.1	6.5	n.a.
Multirepeat (GH)	Model	Template for Rosetta Design [e]	5.9 ± 0.0	16.0 ± 0.0	28.6 ± 0.0	6.2	1.0
CAR0 [f]	Model	Model Rosetta Design	5.8 ± 0.1	17.2 ± 0.5	27.4 ± 0.5	6.4 ± 0.1	Ref [c]
CAR0 [f]	Model	Model after MDS	5.8 ± 0.3	17.4 ± 1.3	27.4 ± 1.0	6.4 ± 0.1	0.4
CAR1' [g]	Model	Model after MDS	7.1 ± 0.5	15.4 ± 4.5	30.0 ± 7.0	7.6 ± 0.3	2.2
CAR2 [h]	Structure	dArmRP	8.0 ± 0.7	11.9 ± 1.9	30.2 ± 0.9	8.2 ± 0.7	1.8
CAR2.V1 [i]	Structure	dArmRP	7.4 ± 0.3	14.4 ± 0.4	28.6 ± 0.5	7.5 ± 0.2	1.3

[a] Models were obtained by Rosetta design or MD simulations. Structures were solved by X-ray crystallography.
[b] Protein curvature parameters were averaged over each internal repeat pair. For structures with several molecules within the asymmetric unit, parameters were additionally averaged over all molecules.
[c] RMSD: root mean square deviation of all Cα of four internal repeats in comparison to the CAR0 model (Ref).
[d] Repeat pair (GH) of yeast Importin-α (PDB ID: 1EE4)
[e] Multi-repeat model based on several copies of superimposed repeat pairs GH.
[f] CAR0 = $N_V(D_{SPVA})_4C_{PAF}$
[g] CAR1 derivative (differences in the C-cap)
[h] CAR2 = $Y_{III}(Dq)_4C_{PAF}$
[i] CAR2.V1 = $Y_{III}(Dq.V1)_4C_{PAF}$ in complex with peptide $(RR)_5$

Supplementary Table ST3 Oligonucleotides used for the assembly and cloning of dArmRPs

name	sequence 5'-3'	Description [a]
pQE30 for	CGGATAACAATTTCACACAG	for assembly of modules
pQE30 rev	GTTCTGAGGTCATTACTG	rev assembly of modules
pQE30_short_Bam_f	ATCGCATCACCATCACCATCACG	for assembly of modules short
pQE30_short_KpnI_r	TAAGCTTGGCTGCAGGTCGACC	rev assembly of modules short
N1f	CCAGGGATCCTAGGAAGACCTCG	for assembly of D type internal repeat
N2r_S	GCATCTTCCAGCCATTTGCTCAGCTGCGGCAG	rev assembly of D type internal repeat
N2r_V	GCATCTTCCAGCCATTTAACCAGCTGCGGCAG	rev assembly of D type internal repeat
N2r_A	GCATCTTCCAGCCATTTCGCCAGCTGCGGCAG	rev assembly of D type internal repeat
N3f	TGGAAGATGCGAGCGAAGAAACCCAGAAAAACG	for assembly of D type internal repeat
N4r	GCAATTTTCGCAATCTGCTGCGCCGCGTTTTTCTG	rev assembly of D type internal repeat
N5f	CGAAAATTGCGGCGGGCAACAACGAATGAGACC	for assembly of D type internal repeat
N6r	TTCCTGGTACCCTAAGGTCTCATTC	rev assembly of D type internal repeat
D-rev2 43	GAGCACCAGCTTCCACCAGTTTCTGCAGCATTTCGAGGTCTTC	rev assembly of D type internal repeat
D-for3-V 49	GCTGGTGCTCTGYCTSCACTGGTTAAACTGCTGGAAGATGCGTCCGAAG	for assembly of D type internal repeat
D-for3-SA 49	GCTGGTGCTCTGYCTSCACTGKCTAAACTGCTGGAAGATGCGTCCGAAG	for assembly of D type internal repeat
D-rev4-C 45	GTTCGCAATCGCAGCGCAAGCGTTCTTCTTCACCTCTTCGGACGC	rev assembly of D type internal repeat
D-rev4-A 45	GTTCGCAATCGCAGCAGCAGCGTTCTTCTTCACCTCTTCGGACGC	rev assembly of D type internal repeat
D-rev4-L 45	GTTCGCAATCGCAGCCAGAGCGTTCTTCTTCACCTCTTCGGACGC	rev assembly of D type internal repeat
D-for5 39	CGATTGCGAACATTGCGGCGGGCAACAACGAATGAGACC	for assembly of D type internal repeat
D-rev2-QII	GAGCACCAGCTTCTATCAGTTTCTGGATCTGTTCGAGGTCTTC	rev assembly of Dq internal repeat
D-for3-SPVD	GCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAG	for assembly of Dq internal repeat
D-rev4-IV	GTTCGCAATCGCAGCAACAGCGTTCTTGATCACCTCTTCGGACGC	rev assembly of Dq internal repeat
N1_RD_f	CCAGGGATCCGAACTGCCGCAGC	for assembly of N_D-cap
N2r_A	GCATCTTCCAGCCATTTCGCCAGCTGCGGCAG	rev assembly of N_D-cap
N2r_S	GCATCTTCCAGCCATTTGCTCAGCTGCGGCAG	rev assembly of N_D-cap
N2r_V	GCATCTTCCAGCCATTTAACCAGCTGCGGCAG	rev assembly of N_D-cap
N3f	TGGAAGATGCGAGCGAAGAAACCCAGAAAAACG	for assembly of N_D-cap
N4r	GCAATTTTCGCAATCTGCTGCGCCGCGTTTTTCTG	rev assembly of N_D-cap
N5f	CGAAAATTGCGGCGGGCAACAACGAATGAGACC	for assembly of N_D-cap
N6r	TTCCTGGTACCCTAAGGTCTCATTC	rev assembly of N_D-cap
C1f	CCAGGGATCCTAGGAAGACCTCG	for assembly of C_D-cap
C2r	GCTTCTTCCAGTTTCTGTTTCATTTCGAGGTCTTC	rev assembly of C_D-cap
C3f_PA	TGGAAGAAGCGGGCGCGCTGCCGGCGCTGGAAAAACTGC	for assembly of C_D-cap
C3f_SP	TGGAAGAAGCGGGCGCGCTGAGCCCGCTGGAAAAACTGC	for assembly of C_D-cap
C4r	CGCGTTTTTCTGCACTTCTTCGTTCGCATGGCTCTGCAGTTTTTC	rev assembly of C_D-cap
C5f_F	GAAAAACGCGCAGGCGGCGCTGGAAGCGTTTAACAGCTAATGAG	for assembly of C_D-cap
C5f_I	GAAAAACGCGCAGGCGGCGCTGGAAGCGATTAACAGCTAATGAG	for assembly of C_D-cap
C6r	CCGTTCCTGGTACCTCATTAGCTG	rev assembly of C_D-cap
N2_SPV_rev.CR	CATTTCGTTGCCACCAGAAGCGATCTGAGACAG	rev Y_{III}-cap insertion in $(D_{SPVA})_4$-constructs
SPV_N2_for.CR	GCTTCTGGTGGCAACGAAATGCTGCAGAAACTGG	for Y_{III}-cap insertion in $(D_{SPVA})_4$-constructs
C1_SPV_for.CR	GCAACAACGAACAGAAACAGGCTGTTAAAG	for A_{II}-cap insertion in $(D_{SPVA})_4$-constructs
SPV_C1_rev.CR	CTGTTTCTGTTCGTTGTTGCCCGCCGCAATGTTC	rev A_{II} -cap insertion in $(D_{SPVA})_4$-constructs
Dq_N2_for.CR	GCTTCTGGTGGCAACGAACAGATCCAGAAACTGATAG	for Y_{III}-cap insertion in $(Dq)_4$-constructs
N2_Dq_rev.CR	CTGTTCGTTGCCACCAGAAGCGATCTGAGACAG	rev Y_{III}-cap insertion in $(Dq)_4$-constructs
Gene V1	GGATCCGAACTGCCGCAGATGGTTCAGCAGCTGAACTCTCCGGACCAGCAGGAACTGCAGTCTGCTCTTCGAAAACTGTCTCAGATCGCTTCTGGTGGTAACGAA CAGATCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCAAGGAAGCTGTTTGGGCGATTGCGAAC ATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGATAGAAGCCGGCGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCAAGGAA GCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCG TCCGAAGAGGTGATCAAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTG GTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCAAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGGAAGAA GCGGGCGCGGAACCGGCGCTGGAAAAACTGCAGAGCTCCCCGAACGAAGAAGTGCAGAAAAACGCGCAGGCGGCGCTGGAAGCGCTGAACAGCTAATGATGAGGT ACCCCGGGTCGACCTGCAGCCAAGCTT
Gene V2	GGATCCGAACTGCCGCAGATGGTTCAGCAGCTGAACTCTCCGGACCAGCAGGAACTGCAGTCTGCTCTTCGAAAACTGTCTCAGATCGCTTCTGGTGGTAACGAA CAGATCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCCAGGAAGCTGTTTGGGCGATTGCGAAC ATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGATAGAAGCCGGCGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCAGGAA GCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCG TCCGAAGAGGTGATCCAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTG GTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCCAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGGAAGAA GCGGGCGCGGAACCGGCGCTGGAAAAACTGCAGAGCTCCCCGAACGAAGAAGTGCAGAAAAACGCGCAGGCGGCGCTGGAAGCGCTGAACAGCTAATGATGAGGT ACCCCGGGTCGACCTGCAGCCAAGCTT
Gene V3	GGATCCGAACTGCCGCAGATGGTTCAGCAGCTGAACTCTCCGGACCAGCAGGAACTGCAGTCTGCTCTTCGAAAACTGTCTCAGATCGCTTCTGGTGGTGACGAA CAGACCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCAAGGAAGCTGTTTGGGCGATTGCGAAC ATTGCGTCCGGCAACGACGAACAGACCCAGAAACTGATAGAAGCCGGCGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCAAGGAA GCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACGACGAACAGACCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCG TCCGAAGAGGTGATCAAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACGACGAACAGACCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTG GTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCAAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGGAAGAA GCGGGCGCGGAACCGGCGCTGGAAAAACTGCAGAGCTCCCCGAACGAAGAAGTGCAGAAAAACGCGCAGGCGGCGCTGGAAGCGCTGAACAGCTAATGATGAGGT ACCCCGGGTCGACCTGCAGCCAAGCTT
Gene V3	GGATCCGAACTGCCGCAGATGGTTCAGCAGCTGAACTCTCCGGACCAGCAGGAACTGCAGTCTGCTCTTCGAAAACTGTCTCAGATCGCTTCTGGTGGTGACGAA CAGACCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCCAGGAAGCTGTTTGGGCGATTGCGAAC ATTGCGTCCGGCAACGACGAACAGACCCAGAAACTGATAGAAGCCGGCGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCCAGGAA GCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACGACGAACAGACCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTGGTTAAACTGCTGGACGATGCG TCCGAAGAGGTGATCCAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACGACGAACAGACCCAGAAACTGATAGAAGCTGGTGCTCTGTCTCCACTG GTTAAACTGCTGGACGATGCGTCCGAAGAGGTGATCCAGGAAGCTGTTTGGGCGATTGCGAACATTGCGTCCGGCAACAACGAACAGATCCAGAAACTGGAAGAA GCGGGCGCGGAACCGGCGCTGGAAAAACTGCAGAGCTCCCCGAACGAAGAAGTGCAGAAAAACGCGCAGGCGGCGCTGGAAGCGCTGAACAGCTAATGATGAGGT ACCCCGGGTCGACCTGCAGCCAAGCTT

[a] for=forward; rev=reverse

Supplementary Table ST4 List of target peptides used in ELISAs

Name	Biotinylation	Linker	Fusion partner	Sequence	C-terminal end
$(KR)_5$	Yes	Ttds[a]	—	KRKRKRKRKR	carboxyl
$(KR)_4$	Yes	Ttds[a]	—	KRKRKRKR	carboxyl
$(ER)_5$	Yes	Ttds[a]	—	DRDRDRDRDR	amide
$(DR)_5$	Yes	Ttds[a]	—	DRDRDRDRDR	amide
NLS[b]	Yes	Ttds[a]	—	KKKRKV	amide
H1[c]	Yes	Ttds[a] - GGSGG	—	DSNFYRAL	amide
H2[d]	Yes	Ttds[a] - GGSGG	—	NPEYLGLD	amide
H3[e]	Yes	Ttds[a] - GGSGG	—	DEEYEYMNR	amide
H4[f]	Yes	Ttds[a] - GGSGG	—	KAEYLKNN	amide
pD-H0[g]	Yes	—	pD[g] - GSGGSGG	AAAAVVVV	carboxyl

[a] Ttds: 4,7,10-trioxa-1,13-tridecanediamine succinimic acid connected to biotin

[b] NLS: nuclear localization sequence from the SV40 large T antigen (GenBank accession number: AAB59924; residue: 126-133)

[b] pD: bacteriophage lambda protein D from vector pAT223 (GenBank accession number: AY327138) with a glycine-serine linker (GSG)

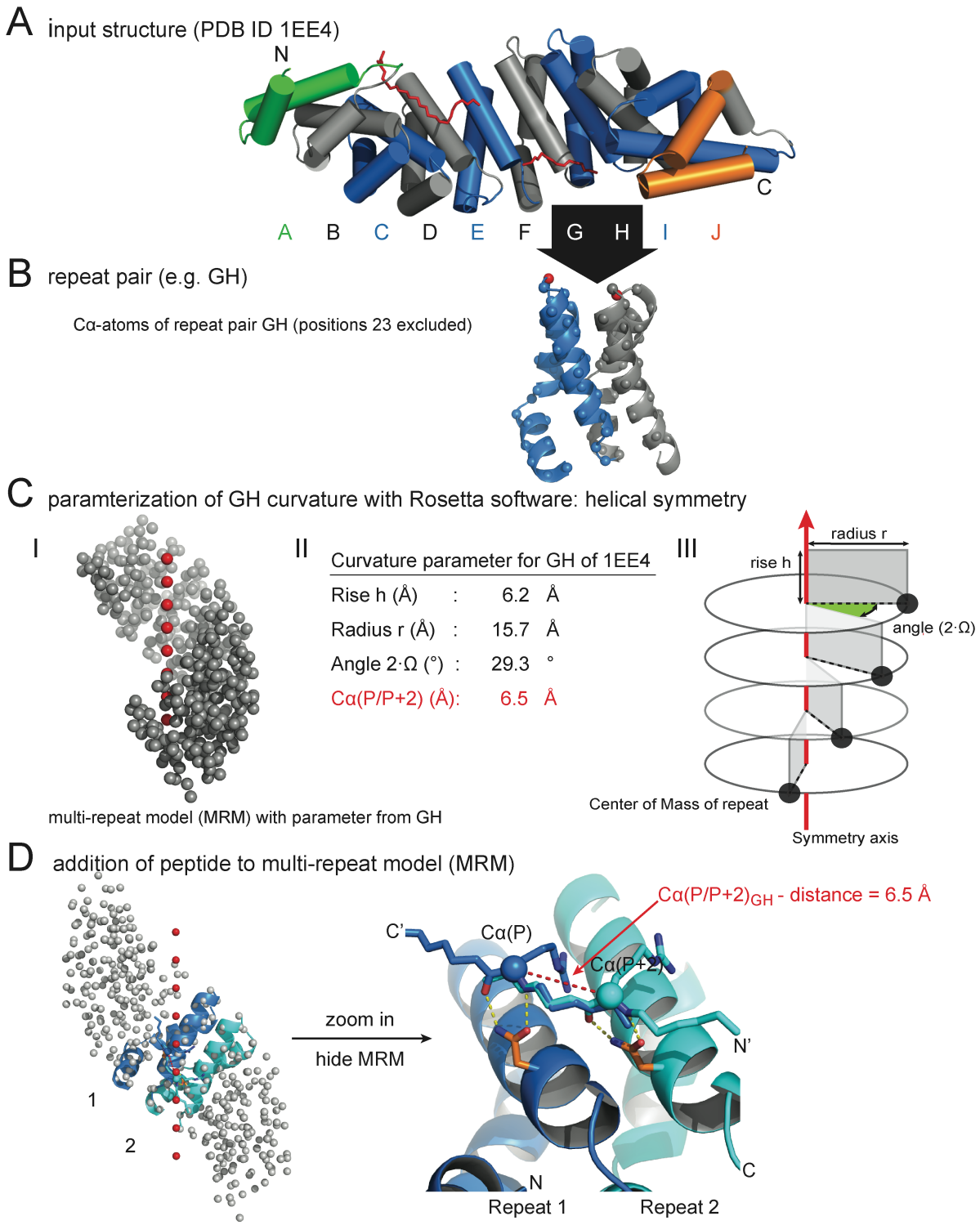
[c] H1: C-terminal region of the human epidermal growth factor receptor (EGFR) (NCBI Reference Sequence: NP_005219; residue: 994-1001)

[d] H2: C-terminal region of the human epidermal growth factor receptor 2 (HER2) (NCBI Reference Sequence: NP_004439; residue: 1245-1252)

[e] H3: C-terminal region of the human epidermal growth factor receptor 3 (HER3) (NCBI Reference Sequence: NP_001973; residue: 1194-1212)

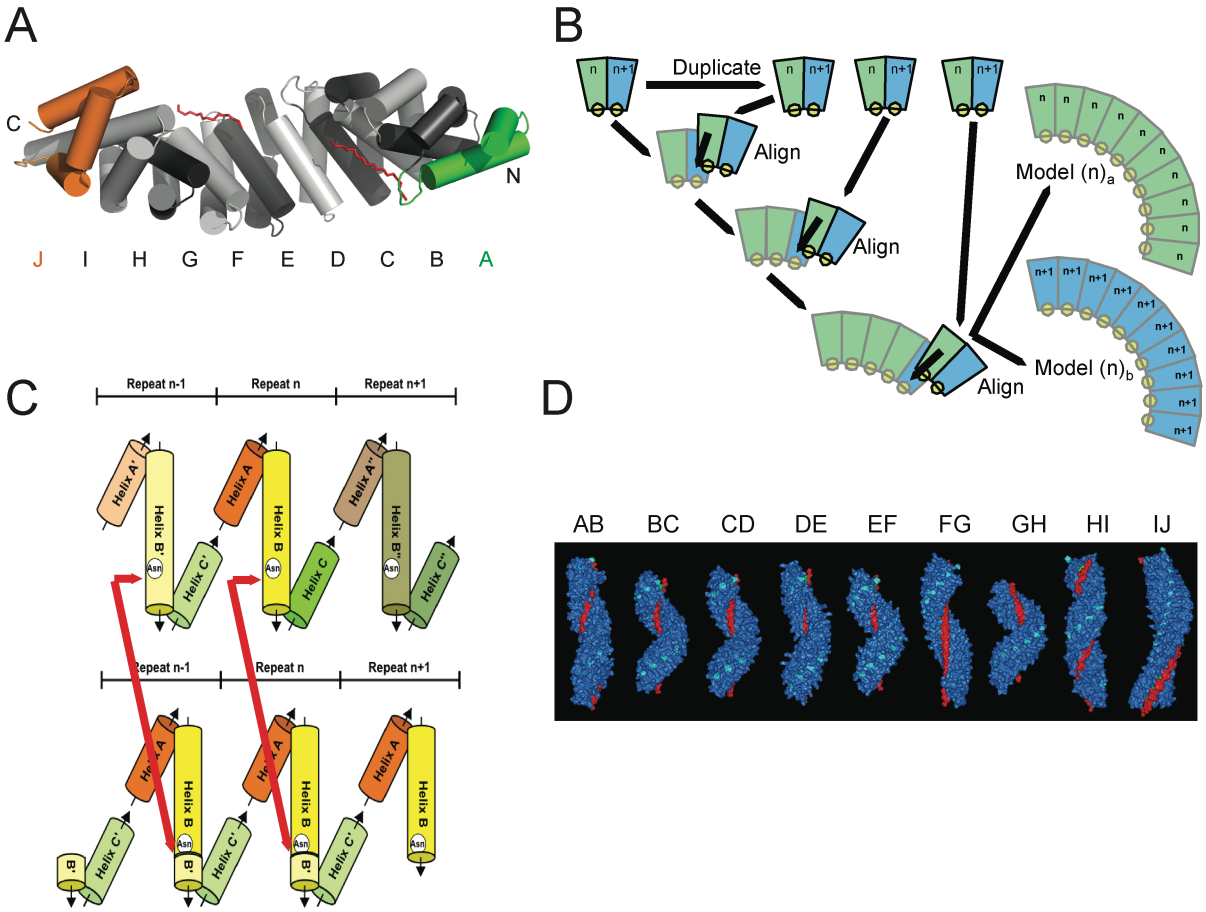
[f] H4: C-terminal region of the human epidermal growth factor receptor 4 (HER4) (NCBI Reference Sequence: NP_005226; residue: 1218-1225)

[g] H0: unspecific control peptide



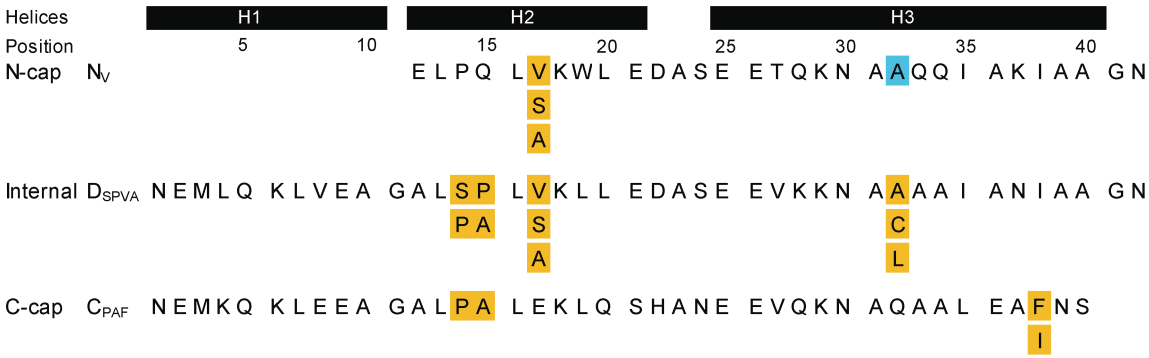
Supplementary Figure S1. Parameterization of the curvature of ArmRPs. A: Input structure shown in cylinder representation, e.g., 1EE4 [2]. The N- and C-terminal repeat is indicated in green and orange, respectively. Internal repeats are colored alternatingly in grey and blue. Repeats are labeled from A to J. B: For the description of the geometry between repeat pair GH, an input file was generated, containing the coordinates of the Cα-atoms of 41 residues of internal repeat G and H, respectively

(residue 23 was excluded, because it is absent in some nArmRP. The nomenclature is according to Figure 2). C: Parameterization of the GH repeat pair curvature based on helical symmetry was performed by Rosetta's symmetric framework as described by DiMaio *et al.* [3] The result of the parameterization are: (I) several output files (e.g. the multi-repeat model (MRM)) and (II) curvature parameters: *rise* h (Å), *radius* r (Å) and *angle'* (rad). The angle given in this manuscript is *angle* Ω (°) = 2 x *angle'* (°). The MRM (with nine identical repeats, named 1-9) and symmetry axis are shown in grey and red spheres, respectively. (III) Schematic MRM according to the parameters obtained between two internal repeats. Each repeat is represented as black sphere at its center of mass (CoM). Helical symmetry operators (rise h, radius r and angle 2Ω) describe the transformation between neighboring repeats, building a right-handed MRM. Accordingly, parameters *rise*, *radius* and *angle* describe the position of the CoM on a cylinder. D: In order to obtain the Cα(P/P+2)-distance for the geometry of the repeat pair, a peptide bound in the conserved and modular orientation is modelled into the MRM. Left: Thereby, two copies of repeat D of the dArmRP binding to the (KRK)-peptide fragment [4] (shown in ribbon representation and colored blue and light blue) were superimposed on the repeat 1 and 2 from the MRM (based on Cα–atoms). Right: The Cα(P/P+2)-distance is measured between the Cα-atom of an amino acid (P) of the bound peptide in repeat 1 to the Cα-atom of an amino acid (P+2) of the bound peptide in repeat 2.

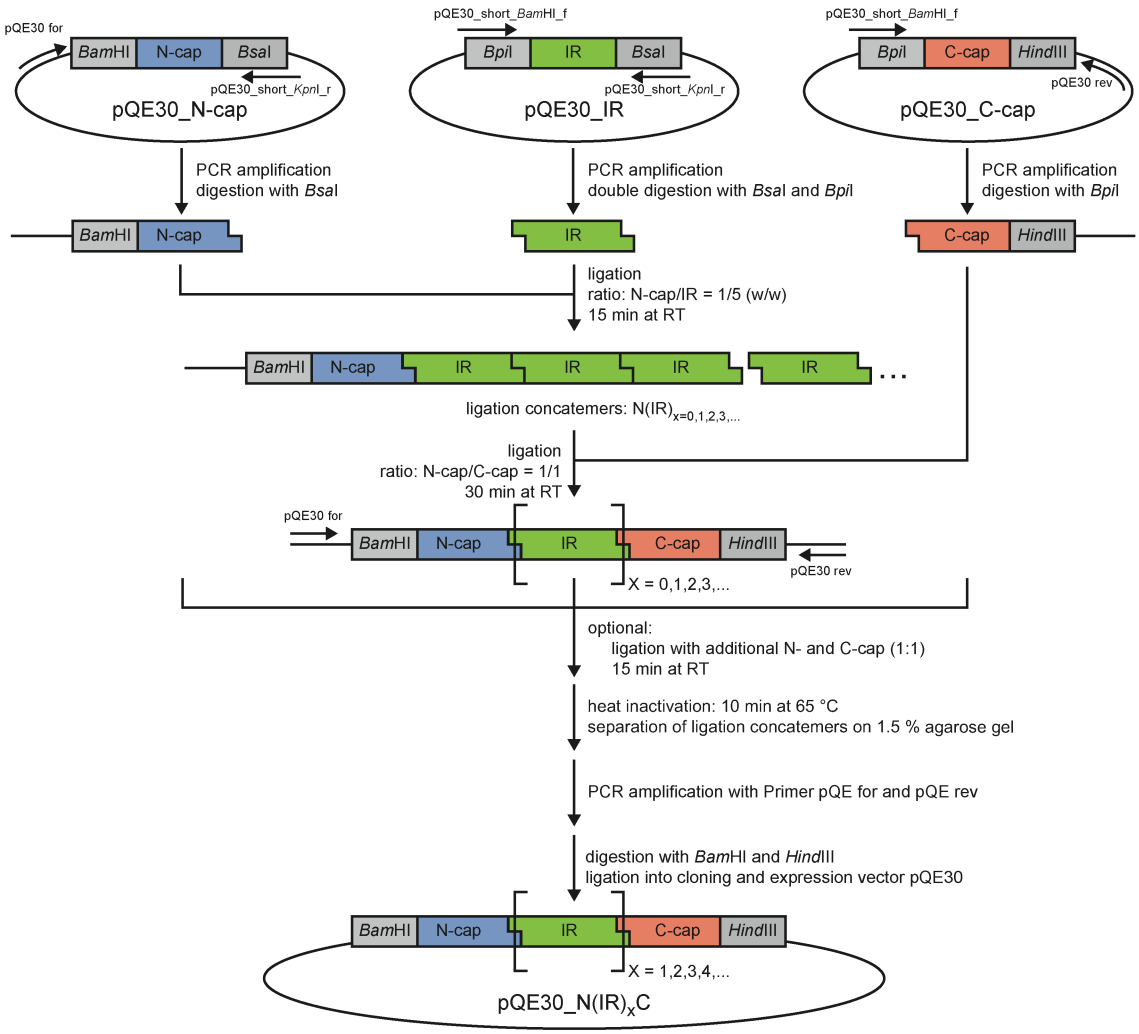


Supplementary Figure S2. Construction a multi-repeat model with an optimal curvature. A: Structure of *S. cerevisiae* importin-α in complex with monopartite NLS (PDB ID 1BK6[2]), showing the right-handed superhelical structure typical for ArmRPs. The cylinders represent the α-helices. The N-terminal repeat is indicated in green, and the C-terminal repeat is shown in orange, respectively. Internal repeats are colored alternatingly in dark and light grey. The bound peptide backbone is shown in stick mode (red). Repeats are labeled from A to J. B: Schematic drawing of the construction procedure. A double repeat was generated from two repeat fragments and the multi-repeat was built by superimposing the first repeat of the second copy on the second repeat of the first copy, then deleting either the first (class b) or the second (class a) repeat of each copy and joining up the remaining residues. C: Construction of single repeat. In this model the single repeat is centered on the helix 3 (called here B) which is the main determinant for the binding site formation and the geometry. Because the interface has to be conserved, the single repeat to be used in the multi-repeat construction was generated from two different parts contributing to the formation of the interface. The junction point is after the conserved asparagine at position 37 (Asn in the picture). The part before is coming from the repeat n, the part after from the repeat n-1. The red arrows indicate the cuts and the

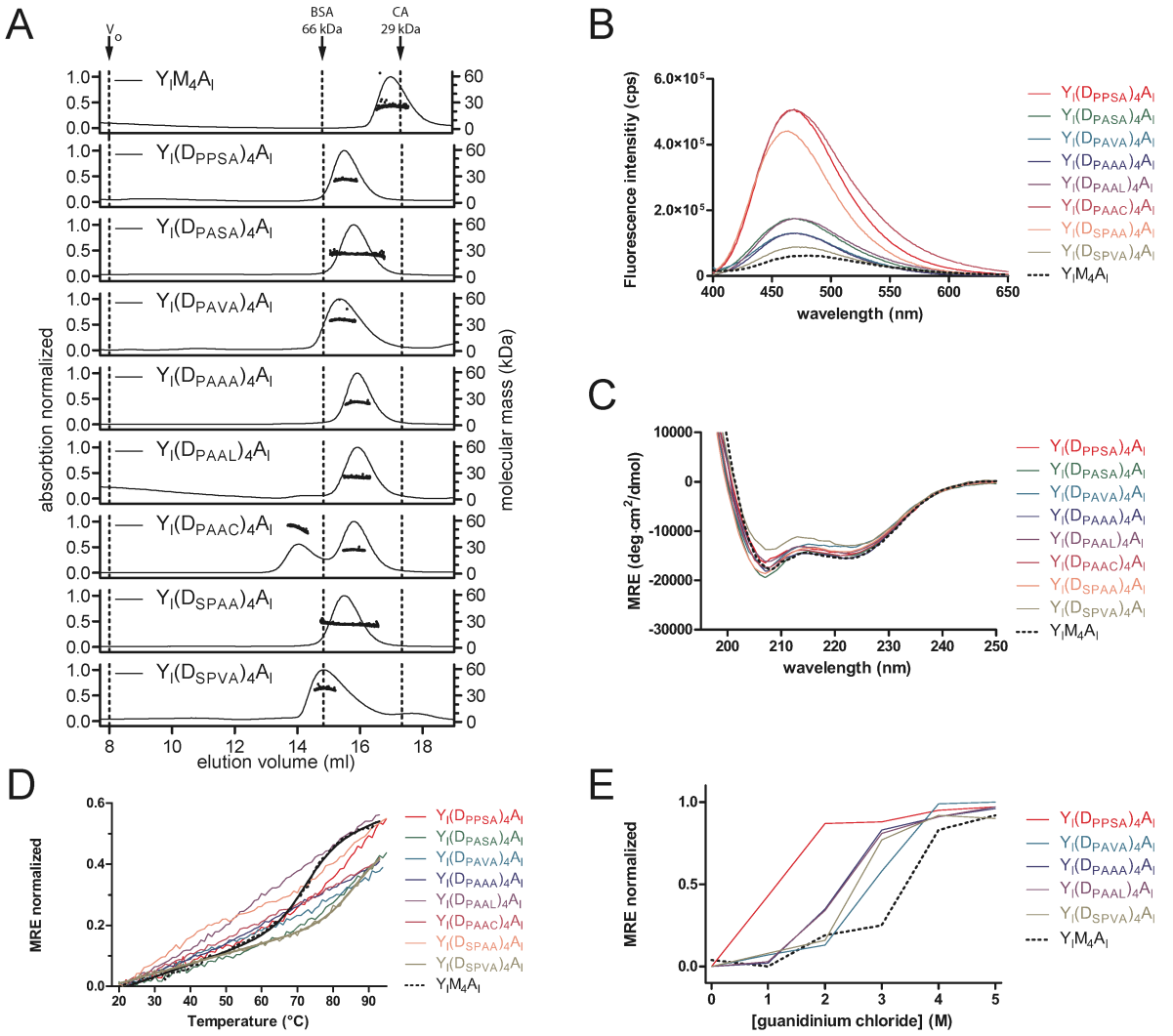
corresponding junction points. D: Multi-repeat models of class a. Each model was built using the interface between two consecutive repeats, indicated by the letters, for the construction of the single repeat, and then used for the construction of the dimer including the geometrical information. A model corresponding to each couple of repeats from importin yeast, importin mouse and catenin was realized. The multi-repeat models based on yeast importin (PDB 1EE4) are shown here. A poly-alanine peptide (PDB ID 1Q1S[5], interacting with repeat pair C), added by superposition of helix 3 to each repeat is depicted in red. The model GH (of class a) was used as starting point for the sequence design, because of its most favorable geometry for peptide binding without distortion of the target.



Supplementary Figure S3. Sequence alignment of computationally designed N-terminal capping repeats (e.g. N_V), internal repeats (e.g. D_{SPVA}) and C-terminal repeats (e.g. C_{PAF}). Subscripts describe the variable residues that were experimentally tested in each module are highlighted in orange. Variable position 32 in the N-cap (blue) was kept constant due to the negative impact of alternative residues Cys and Leu (tested in the internal repeats (see Supplementary Figure S5)).

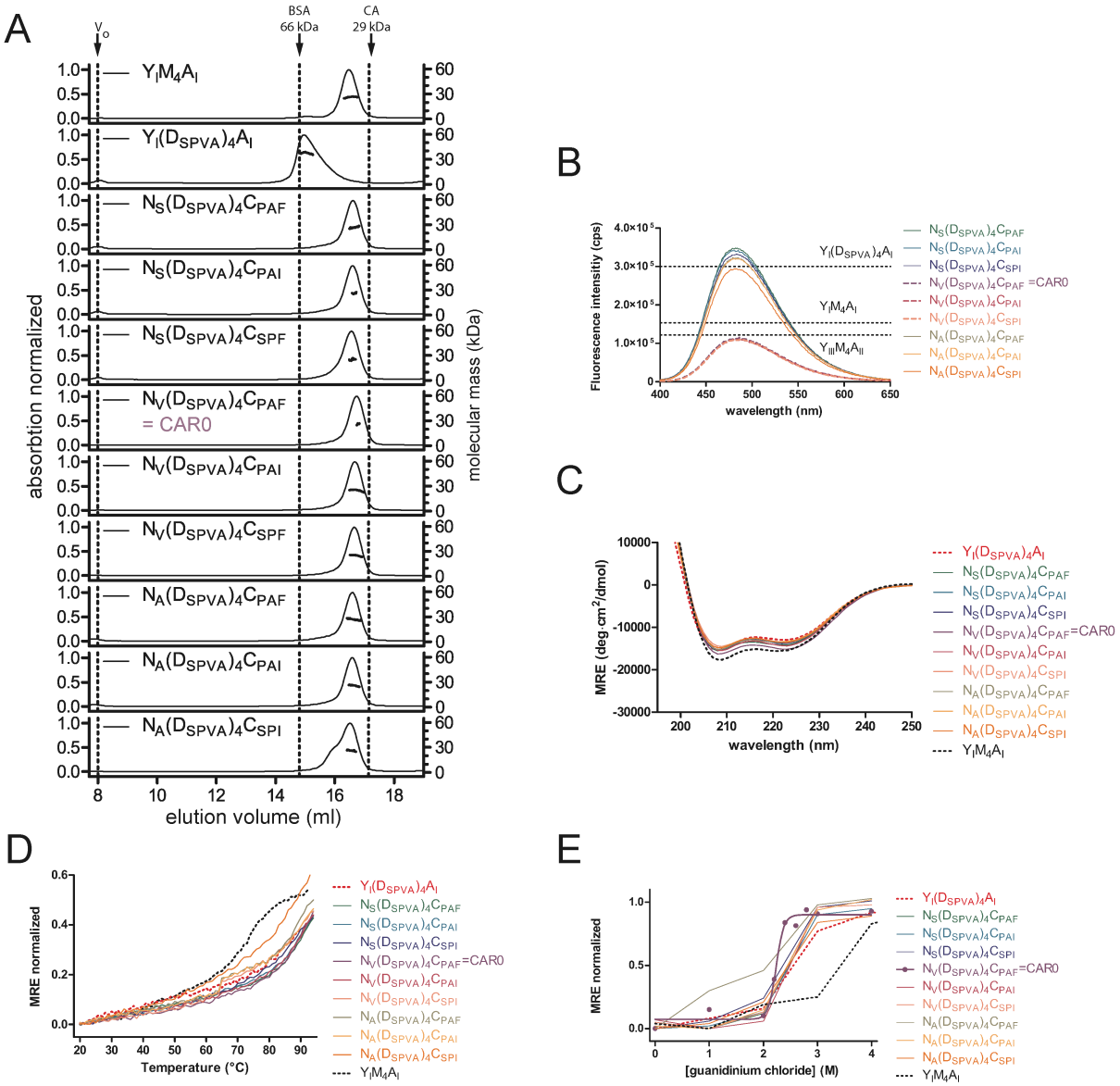


Supplementary Figure S4. Scheme of the multi-fragment assembly strategy for designed armadillo repeat protein constructs at the DNA level, to assemble full-length repeat proteins within one day. The single modules (N-cap (e.g., N_V), internal repeat (e.g., Dq) and C-cap (e.g., C_{PAF})) were amplified from pQE30 vectors, using primers to produce short overhang sites, downstream of the N-cap, on both sides of the internal repeat and upstream of the C-cap, respectively. In a first step, *Bsa*I digested N-cap fragments were ligated with a five-fold excess of *Bsa*I and *Bpi*I double-digested internal repeats to produce ligation concatemers. The ligation reaction is stopped by adding *Bpi*I-digested C-capping fragments. Optionally, the mixture was supplemented again with single-digested N- and C-terminal fragments to saturate the formation of full-length constructs. The ligation mixture was heat-inactivated and separated on a 1.5 % agarose gel. Corresponding fragments, containing 1-10 internal repeats, were extracted and separately amplified by PCR, using outer primers (pQEfor and pQErev), before inserting them via *Bam*HI and *Kpn*I restriction enzymes into a cloning and expression vector.



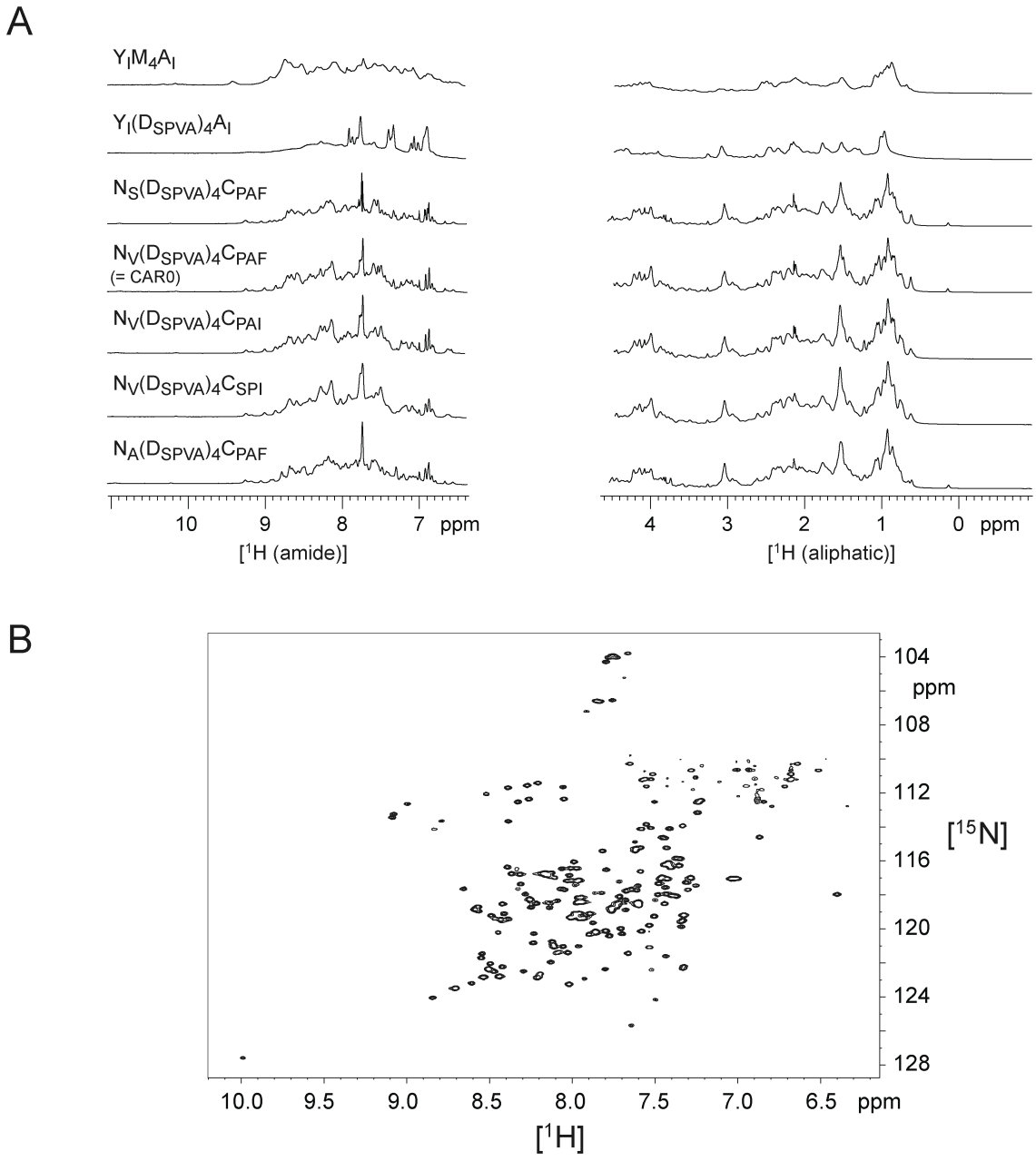
Supplementary Figure S5. Biophysical characterization of designed ArmRPs with different computationally designed internal repeats (D-type). A: SEC (normalized absorption at 230 nm) and MALS (dots) of $Y_I(D)_4A_I$ proteins, containing consensus-based capping repeats (Y_I and A_I)[6] and four identical computationally designed internal repeats (e.g. D_{SPVA}) (cf. sequences in Supplementary Figure S2). Elution volumes corresponding to the void volume (V_0), bovine serum albumin (MW: 66 kDa) and carbonic anhydrase (MW: 29 kDa) are indicated by dashed lines and used as molecular weight standards. All proteins, except of $Y_I(D_{PAAC})_4A_I$, having a cysteine at position 32 and forming a monomer and a dimer, are monomeric, shown by MALS, even though some protein indicate an rather extended conformation with lower elution volume. In comparison, the consensus based $Y_IM_4A_I$ protein eluted at higher volume, fitting closer to the expected molecular weight with globular MW standards. B: ANS fluorescence spectra, measured on a PTI QM-2000-7 fluorimeter (Photon Technology International, USA). C: CD spectra of all proteins are shown, expressed as the mean residue ellipticity (MRE). D: Temperature-induced unfolding of designed proteins. Calculated values of MRE, monitored

at 222 nm, were normalized by setting the initial values (folded) as 0 and the putative complete unfolded protein (MRE=0) as 1. $Y_I(D_{SPVA})_4A_I$ and $Y_IM_4A_I$ (dots) with fit (smooth lines); all other proteins are shown as lines connecting the data points. (E) Normalized GdnHCl-induced unfolding of a selection of dArmRPs.



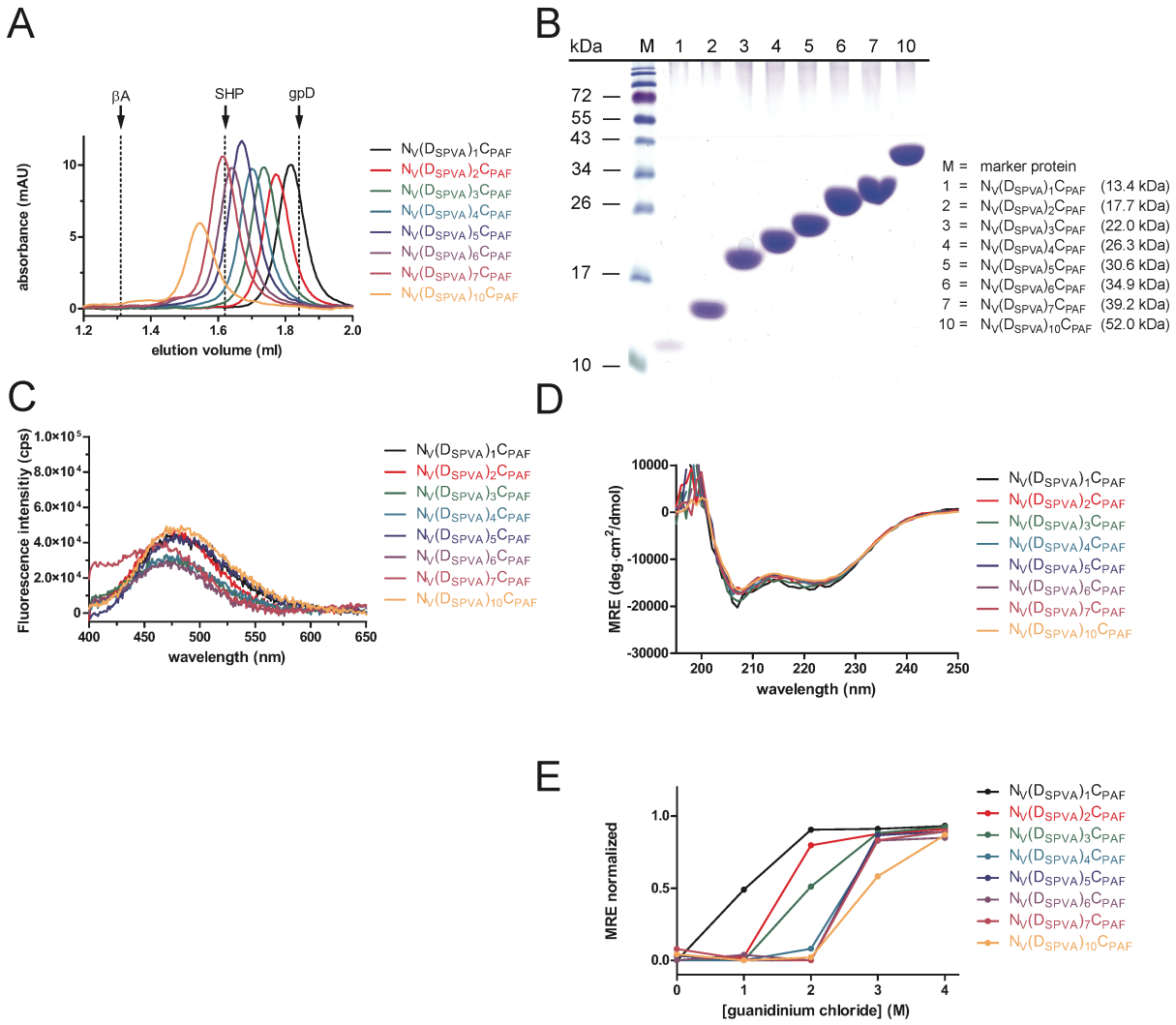
Supplementary Figure S6. Biophysical characterization of dArmRPs carrying internal repeat D_{SPVA} and different computational-designed capping repeats (N-caps: N_S, N_V and N_A; C-caps: C_{PAF}, C_{PAI} and C_{SPI}). A: SEC (normalized absorption at 230 nm) and MALS (dots) of $N(D_{SPVA})_4C$ proteins. Elution volumes corresponding to the void volume (V_0), bovine serum albumin (MW: 66 kDa) and carbonic anhydrase (MW: 29 kDa) are indicated by dashed lines and used as molecular weight standards. All proteins, except of $N_A(D_{SPVA})_4C_{SPI}$ forming in addition a dimeric protein peak, were converted from the tailing peak found in $Y_I(D_{SPVA})_4A_I$, corresponding to a mixture of monomers and dimers, to a single monomeric peak by the corresponding mutations. This peak coincides with the folded consensus-based ArmRP $Y_IM_4A_I$. B: ANS fluorescence spectra. Lowest signals were observed for proteins containing the N_V-capping repeat (dashed curve). Horizontal dashed lines indicated the highest signal

observed for $Y_I(D_{SPVA})_4A_I$ and consensus-based proteins $Y_IM_4A_I$ and $Y_{III}M_4A_{II}$. C: CD spectra of all proteins are shown, given by the mean residue ellipticity (MRE). D: Normalized temperature-induced unfolding of designed proteins, measured by CD spectroscopy at 222 nm. E: Normalized GdnHCl-induced unfolding of dArmRPs. All proteins were measured in steps of 1 M GdnHCl. Protein $N_V(D_{SPVA})_4C_{PAF}$, showing low ANS signal, high secondary structure content by CD and low levels of unfolding at 2 M GdnHCl, was re-measured with 0.2 M steps GdnHCl and is shown as dots with fit (smooth lines).

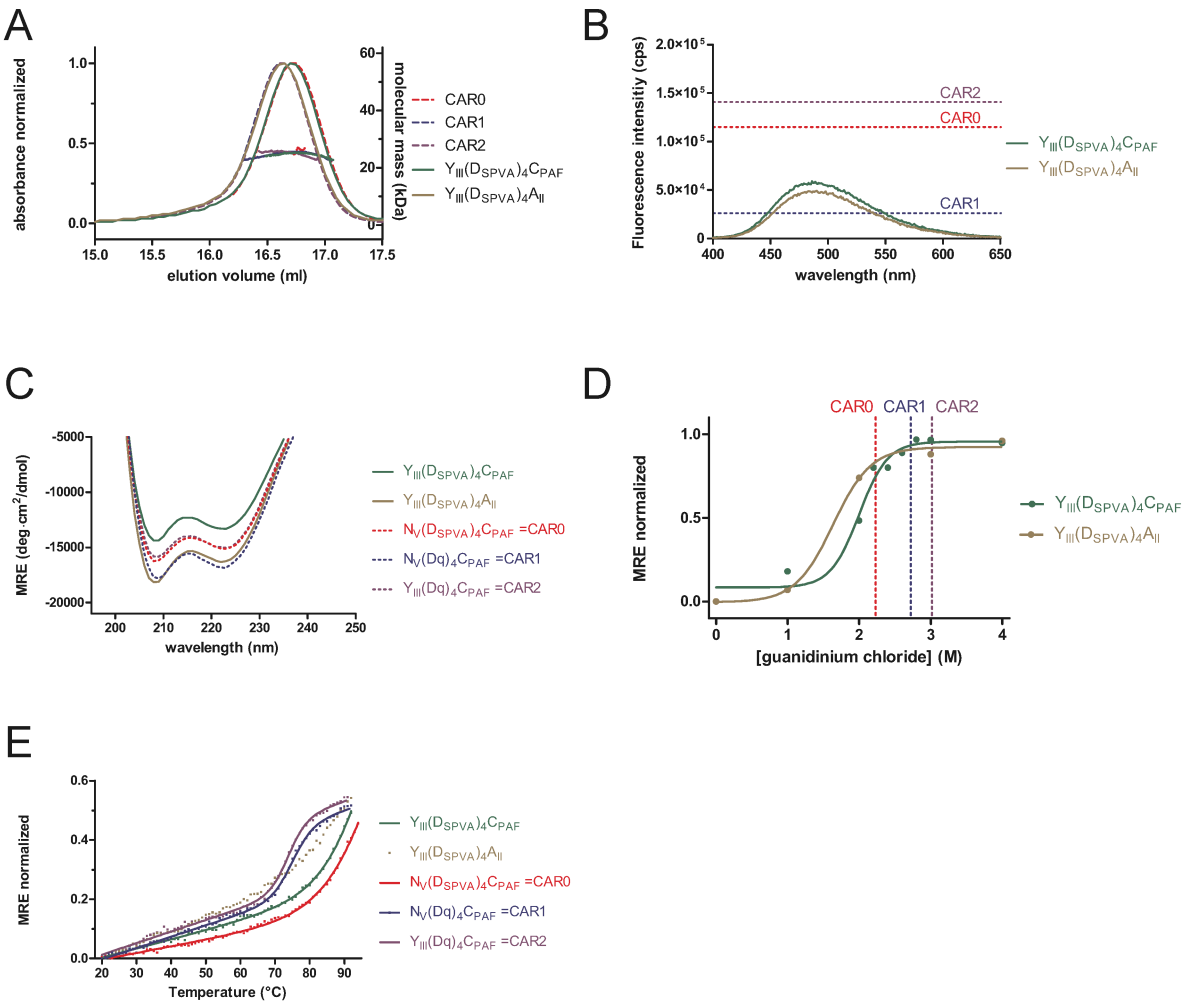


Supplementary Figure S7. NMR spectra of dArmRPs. A: 1D ^{1}H-NMR spectra were recorded to rank proteins according to signal dispersion in the amide- and methyl-region as well as the line width of their proton resonances (amid region was enlarged vertically by a factor of 5 in comparison to the aliphatic region). Poor signal dispersions and broad line widths, indications for less folded and potentially more oligomeric proteins, were found for $Y_I(D_{SPVA})_4A_I$, but were improved in constructs containing computationally designed capping repeats (N-cap: N_S, N_V and N_A; C-cap: C_{PAF}, C_{PAI} and C_{SPI}). Similar to $Y_IM_4A_I$ [1], protein $N_V(D_{SPVA})_4C_{PAF}$ (CAR0) shows good biophysical properties and ^{1}H-NMR spectra. B: [^{15}N,^{1}H]-HSQC spectra of CAR0. From the expected 233 peaks (MRGSH_6GS-tag and prolines excluded), 216 could be counted. Due to overlaps of peaks in repeat proteins, this

number is expected to be less than the maximally possible, and similar to a well folded consensus-based ArmRP [1].



Supplementary Figure S8. Biophysical characterization of dArmRPs with 1-10 D_{SPVA} internal repeats. A: Analytical SEC (normalized absorption at 280 nm) of $N_V(D_{SPVA})_{1\text{-}10}C_{PAF}$ proteins. Elution volumes corresponding to β-amylase (βA: 200.0 kDa), trimeric viral capsid protein [7] (SHP: 35.9 kDa) and protein D [8] (gpD: 11.4 kDa) are indicated by dashed lines and used as molecular weight standards. All proteins are monomeric at a concentration of 10 µM. B: Protein samples used for crystallization are shown in a Coomassie-stained SDS polyacrylamide gel. Proteins run 2-12 kDa below their expected molecular weights. C: All proteins showed low levels of ANS binding in fluorescence spectra, measured on a PTI QM-2000-7 fluorimeter (Photon Technology International, USA). D: CD spectra are plotted according to the mean residue ellipticity (MRE). Based on the close-to-identical MRE values, the α-helical content correlates with the length of the proteins. E: Normalized GdnHCl-induced unfolding based on MRE monitored at 222 nm. Initial values (folded) were normalized to 0 and the putative completely unfolded protein (MRE=0) to 1. Midpoints of unfolding transitions increase according to the number of internal repeats from ~1 M GdnHCl (for N1C) to ~2.9 M (for N10C).



Supplementary Figure S9. Biophysical characterization of dArmRP variants ($Y_{III}(D_{SPVA})_4C_{PAF}$ or $Y_{III}(D_{SPVA})_4A_{II}$) containing the C-terminal consensus-based capping repeat Y_{III}. Protein variants, containing either computationally designed C-terminal capping repeat C_{PAF} (colored greenish) or the consensus-based C-cap A_{II} (colored in brown) are compared to CAR0 ($N_V(D_{SPVA})_4C_{PAF}$), CAR1 ($N_V(Dq)_4C_{PAF}$), or CAR2 ($Y_{III}(Dq)_4C_{PAF}$), colored in red, blue or magenta, respectively. A: SEC (normalized absorption at 230 nm) and MALS (dots) of dArmRPs. All proteins showed a single monomeric peak at elution volumes earlier or close to $N_V(D_{SPVA})_4C_{PAF}$ (CAR0). B: ANS fluorescence spectra. Horizontal dashed lines, indicated the highest signal observed for proteins CAR0-CAR2. C: CD spectra of all proteins are shown, represented by the mean residue ellipticity (MRE). D: Normalized GdnHCl-induced unfolding of dArmRPs. Measured data points (dots) are shown with fits (smooth lines). Transition midpoints of reference proteins are indicated by vertical dashed lines. E: Normalized temperature-induced unfolding of designed proteins (dotts) with fits (smooth lines).

References

[1] P. Alfarano, G. Varadamsetty, C. Ewald, F. Parmeggiani, R. Pellarin, O. Zerbe, et al., Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy, Protein Sci. 21 (2012) 1298-1314.

[2] E. Conti, M. Uy, L. Leighton, G. Blobel, J. Kuriyan, Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha, Cell 94 (1998) 193-204.

[3] F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker, I. André, Modeling symmetric macromolecular structures in Rosetta3, PLoS One 6 (2011) e20450.

[4] S. Hansen, D. Tremmel, C. Madhurantakam, C. Reichen, P.R. Mittl, A. Plückthun, Structure and Energetic Contributions of a Designed Modular Peptide-Binding Protein with Picomolar Affinity, J. Am. Chem. Soc. 138 (2016) 3526-3532.

[5] M.R. Fontes, T. Teh, G. Toth, A. John, I. Pavo, D.A. Jans, et al., Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin-alpha, Biochem. J. 375 (2003) 339-349.

[6] F. Parmeggiani, R. Pellarin, A.P. Larsen, G. Varadamsetty, M.T. Stumpp, O. Zerbe, et al., Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core, J. Mol. Biol. 376 (2008) 1282-1304.

[7] P. Forrer, C.S. Chang, D. Ott, A. Wlodawer, A. Plückthun, Kinetic stability and crystal structure of the viral capsid protein SHP, J. Mol. Biol. 344 (2004) 179-193.

[8] P. Forrer, R. Jaussi, High-level expression of soluble heterologous proteins in the cytoplasm of Escherichia coli by fusion to the bacteriophage Lambda head protein D, Gene 224 (1998) 45-52.