Supporting information

for

Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity

Simon Hansen[a], Dirk Tremmel[a], Chaithanya Madhurantakam[a, 2], Christian Reichen[a, 3], Peer R. E. Mittl*[, a] and Andreas Plückthun*[, a]

[a]: Department of Biochemistry, University Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[2]: Present address: Department of Biotechnology, TERI University, 10, Institutional Area, Vasant Kunj, New Delhi 110070, India

[3]: Present address: Molecular Partners AG, Wagistrasse 14, 8952 Schlieren, Switzerland

Correspondance;

*Andreas Plückthun, Tel. +41-44-635 5570, Fax. +41-44-635 5712, e-mail: plueckthun@bioc.uzh.ch

*Peer Mittl, Tel. +41-44-635 6559, e-mail: mittl@bioc.uzh.ch

SI Methods

Cloning of expression plasmids for peptide-sfGFP fusions, peptide-pD fusions and dArmRP

Coding sequences for the desired peptides were produced by two overlapping (by 12 bp) DNA oligomers (Microsynth, Switzerland) that included a BamHI and a HindIII restriction site, respectively. They were annealed and filled up to double-stranded DNA using Phusion® High-Fidelity DNA Polymerase (New England BioLabs). The resulting fragments were *Bam*HI / *Hind*III digested (FastDigest enzymes, Fermentas) and ligated (T4 ligase, Fermentas) into a BamHI / HindIII / FastAP digested (FastDigest enzymes, Fermentas) plasmid pAnN (sfGFP fusions) or pAT223 (pD-fusions[1]). The pAnN vector is a pQE30 (Qiagen) derivative containing an N-terminal MRGSHis$_6$-tag, followed by a BamHI and HindIII restriction site and sfGFP gene. The vector also contains a *lacI^q* and an ampicillin (*amp*) resistance gene. Chemically competent *E. coli* XL1 Blue cells were transformed with the ligation mix and plated on LB / amp (100 µg/ml) / IPTG (500 µM) agar plates. Yellow colored colonies (indicating sfGFP expression) were used to inoculate 5 ml 2YT medium (100 µg/ml amp), from which plasmids were extracted with QIAprep Spin Miniprep Kits (Qiagen). Plasmids were sent for sequencing (Microsynth). Chemically competent *E. coli* BL21 (DE3) or XL1 blue cells were transformed with plasmids with confirmed sequences, for in-vivo biotinylation of pD-fusions double transformation with pAT223 and pBirAcm plasmids (Avidity Inc.) were performed, in this case 35 µg/ml chloramphenicol (cam) was added to plates and media. From single colonies an overnight culture was grown (2YT / amp / [cam if needed]) and a glycerol stock (20% glycerol) was made for future expressions.

The genes for the dArmRPs have already been described previously[2]. For crystallization they were cloned into the expression vector pQE30LIC_3C, a pQE30 derivative with a cleavage site for Human Rhinovirus 3C Protease after the MRGSHis$_6$ tag. For other experiments they were cloned

into the expression vector pQIq containing an uncleavable N-terminal $MRGSHis_6$ tag[3]. An overview of the protein sequences is shown in Fig. S3.

Expression of dArmRP, peptide-pD- and peptide-sfGFP-fusions

An inoculation culture (ca. 25 ml for 1 l of expression culture) was grown overnight (2YT, 100 µg/ml amp) from glycerol stocks. Baffled shake flasks were inoculated to an OD_{600} of ca. 0.1, grown to OD_{600} of ca. 0.8 and induced by adding 500 µM IPTG. sfGFP-fusions were grown at 30°C and 28°C after incubation overnight in Terrific Broth (TB) medium (containing 100 µg/ml amp). For dArmRP and peptide-pD-fusion expression, cells were grown at 37°C for 5 h after induction in 2YT medium (100 µg/ml amp). For in-vivo biotinylation cam (35 µg/ml) was included in the medium and biotin solution was added to a final concentration of 50 µM prior to induction. Cells were harvested by centrifugation (5 min, 4000 *g*) and resuspended in TBS_W (50 mM Tris·HCl pH 7.5, 400 mM NaCl, 20 mM imidazole, 10% glycerol; 60 ml per liter of expression culture). Resuspended cells were frozen in liquid N_2 and stored at -80°C until further use.

Purification of dArmRP, peptide-pD- and peptide-sfGFP fusions

Frozen resuspended *E. coli* cells (corresponding to 1 l original culture) were thawed on ice and a spatula tip of chicken lysozyme (Sigma) was added. Cells were lysed by passage through a French press and by subsequent sonication (Branson). Cell debris was pelleted by centrifugation (25,000 *g*, 30 min). Supernatant was loaded (gravity flow) onto Ni-NTA agarose columns (Qiagen) that were equilibrated with TBS_W. Columns were washed with 30 column volumes (CV) of TBS_W. For peptide-sfGFP fusions a modified TBS_W buffer containing 50 mM imidazole was used. Proteins were eluted with TBS_E (50 mM Tris HCl pH 7.5, 500 mM NaCl, 350 mM imidazole; 2.5 CV). Eluted proteins were dialyzed against PBS (overnight, 4°C). After dialysis 10% glycerol was added to sfGFP constructs and the concentration was determined by absorbance at 280 nm,

proteins were stored at -20 °C. pD-fusions were often affected by partial proteolysis of the fused peptide, to remove proteolyzed proteins, cation exchange chromatography was used. Full-length proteins were checked by mass spectroscopy.

Purification of dArmRP for crystallization

Lysis of the *E. coli* cells and a first IMAC purification was carried out as described above including elution with TBS_E. For MRGSHis$_6$-tag removal by 3C Protease, the eluted proteins were mixed with 2% (w/w) of 3C Protease. The mixture was dialyzed (overnight, 4°C) against 50 mM Tris HCl, pH 8, 300 mM NaCl. The cut-off (12-14 kDa) of the dialysis bag was chosen such that the cleaved-off MRGSHis$_6$ fragments could pass through the membrane but the dArmRP and the 3C-Protease would remain inside. A second passage over Ni-NTA columns was performed to remove uncleaved dArmRP as well as 3C Protease (which carries an uncleavable His-tag). The flow-through was collected, including an additional elution (2 CV) with the dialysis buffer. Monomeric protein fractions were isolated by preparative size exclusion chromatography (Superdex 200 HiLoad 26/60 column) with 10 mM Tris HCl pH 7.4, 100 mM NaCl as running buffer. The monomeric fractions were pooled and concentrated (spin concentrator devices, Millipore) to a final concentration of 20 mg/ml. Free $(KR)_5$ peptide (LifeTein) was added to the proteins in a 1.5 fold molar excess.

Extrapolation of dissociation constants of tight interactions (addition to main text)

The extrapolations based on the number of armadillo repeats resulted in the following equations for the respective peptides: $(KR)_3$: $Y = -0.9222 \cdot X - 1.534$ (R^2=0.9901); $(KR)_4$: $Y = -1.342 \cdot X - 1.036$ (R^2=0.9931) and $(KR)_5$: $Y = -1.692 \cdot X - 0.6118$ (R^2=0.9670). The equation for $(KR)_4$ was used to estimate the K_d for the interaction between $Y_{III}M_7A_{II}$ with $(KR)_4$; the equation for $(KR)_5$ was used to estimate the K_ds of $Y_{III}M_7A_{II}$ with $(KR)_5$ and $Y_{III}M_6A_{II}$ with $(KR)_5$.

The extrapolation based on the number of lysine to alanine mutations resulted in the following equations: $Y_{III}M_7A_{II}$ / $(KR)_4$: $Y = 0.8595 \cdot X - 10.13$ (R^2=0.9806); $Y_{III}M_7A_{II}$ / $(KR)_5$: $Y = 0.4852 \cdot X - 10.29$ (R^2=0.9794); and $Y_{III}M_6A_{II}$ / $(KR)_5$: $Y = 0.6552 \cdot X - 10.07$ (R^2=0.9837). The Y-intercepts of these equations were used to estimate the K_ds of the respective dArmRP – peptide combination without any mutation.

SI Tables

Table S1

Significance levels for ΔΔG values of peptides with different Ala mutations

	$(KR)_4$	K1A	R2A	K3A	R4A	K5A	R6A	K7A	R8A	K3A, K5A	R4A R6A	K3A, R4A	K5A, R6A
$(KR)_4$		****	****	****	****	****	****	****	****	****	****	****	****
K1A	****		****	ns	****	ns	****	ns	ns	****	****	****	****
R2A	****	****		****	ns	****	ns	****	****	****	****	****	****
K3A	****	ns	****		****	ns	****	ns	*	****	****	****	****
R4A	****	****	ns	****		****	ns	****	****	****	****	****	****
K5A	****	ns	****	ns	****		****	ns	ns	****	****	****	****
R6A	****	****	ns	****	ns	****		****	****	****	****	****	****
K7A	****	ns	****	ns	****	ns	****		*	****	****	****	****
R8A	****	ns	****	*	****	ns	****	*		****	****	****	****
K3A, K5A	****	****	****	****	****	****	****	****	****		****	****	****
R4A R6A	****	****	****	****	****	****	****	****	****	****		****	****
K3A, R4A	****	****	****	****	****	****	****	****	****	****	****		ns
K5A, R6A	****	****	****	****	****	****	****	****	****	****	****	ns	

ns: not significant (p-value > 0.05)

*: p-value < 0.05

**: p-value <0.01

***: p-value < 0.001

****: p-value < 0.0001

All peptides are based on $(KR)_4$, mutations are indicated; measurements were performed with $Y_{III}M_5A_{II}$

A one-way ANOVA with multiple comparisons was performed

Table S2

Measured K_d values with Lys to Ala mutations used in the extrapolation of tight binding

Peptide sequence	Number of Lys mutations	K_d ± Std. Dev. (nM)	
		with $Y_{III}M_6A_{II}$	with $Y_{III}M_7A_{II}$
(KR)$_4$-based:			
KRKRARKR	1	n.d.	0.48±0.18
KRARARKR	2	n.d.	4.9±0.53
ARARARKR	3	n.d.	36.9±4.6
ARARARAR	4	n.d.	157±17.6
(KR)$_5$-based:			
KRKRARKRKR	1	0.28±0.02	t.t.
KRKRKRARKR	1	0.40±0.11	t.t.
KRKRARARKR	2	2.7±0.41	0.48±0.13
KRARARARKR	3	8.6±1.3	1.49±0.16
KRARARARAR	4	36±1.7	5.2±0.73
ARARARARAR	5	142±11.7	13.2±2.0

n.d.: not determined

t.t.: too tight to be measured

All values are means of at least three independent assays.

Table S3

ΔΔG values for the increase of repeat number (x) in $Y_{III}M_xA_{II}$ (each line for a constant $(KR)_n$-peptide) or the increase of dipeptide units (n) in $(KR)_n$ (each line for a constant $Y_{III}M_xA_{II}$).

	slope*	ΔΔG† (kJ/mol)
Extension of $Y_{III}M_xA_{II}$		
$(KR)_3$	-0.922	-5.3
$(KR)_4$	-1.342	-7.7
$(KR)_5$	-1.692	-9.7
Extension of $(KR)_n$		
$Y_{III}M_3A_{II}$	-0.645	-3.7
$Y_{III}M_4A_{II}$	-1.059	-6.0
$Y_{III}M_5A_{II}$	-1.505	-8.6
$Y_{III}M_6A_{II}$	-2.069	-11.8

*: calculated trough linear regression shown in Fig. 2c and Fig. S2

†: $\Delta\Delta G = RT\cdot(1/\log(e))\cdot slope$; with T = 298K

Table S4

Data collection and refinement statistics of $Y_{III}M_5A_{II}$ - $(KR)_5$

	$Y_{III}M_5A_{II}$ - $(KR)_5$ (PDB ID: 5aei)
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	88.83, 215.28, 105.99
α, β, γ (°)	90, 90, 90
Resolution (Å)	107.64 - 1.83 (1.877 - 1.83)*
[1]$R_{merge\ (\%)}$	18.95 (153.0)
$I/\sigma I$	14.24 (2.02)
Completeness (%)	99.97 (99.83)
Redundancy	11.1 (11.1)
Refinement	
Resolution (Å)	47.99 - 1.83
No. reflections	89570
[2]$R_{work/}$ R_{free}	0.149 / 0.1791
No. atoms	7694
Protein	6748
Ligand/ion	39
Water	907
B-factors	
Protein	19.8
Ligand/ion	26.0
Water	31.3
R.m.s deviations	
Bond lengths (Å)	0.025
Bond angles (º)	2.06

*Highest resolution shell is shown in parenthesis.

[1]$R_{merge} = \Sigma_{hkl}\Sigma_i||I_i(hkl) - (I(hkl))||/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection hkl and (I(hkl)) is the weighted average intensity for all observations *i* of reflection hkl.

[2]R_{work} and $R_{free} = (\sum||F_o| - |F_c||)/(\sum|F_o|)$, where $|F_o|$ is the observed structure-factor amplitude and $|F_C|$ is the calculated structure-factor amplitude.

The structure was solved from a single crystal.

Table S5

Influence of pH value and NaCl concentration on affinity between $Y_{III}M_5A_{II}$ and $(KR)_5$.

Buffer condition	K_d ± Std. Dev. (nM)
pH value variation	
pH 4.5 (Citrate), 150 mM NaCl	632±71.3
pH 5.0 (MES), 150 mM NaCl	192±28.3
pH 5.6 (MES), 150 mM NaCl	41.3±12.8
pH 6.0 (MES),150 mM NaCl	9.6±5.1
pH 6.5 (MES), 150 mM NaCl	3.2±2.7
pH 7.0 (Phosphate), 150 mM NaCl	0.98±0.4
pH 7.4 (Phosphate), 150 mM NaCl	1.1±0.8
pH 8.0 (Bicine or Tris), 150 mM NaCl	0.86±0.2
pH 8.5 (Bicine or Tris), 150 mM NaCl	0.62±0.4
pH 9.0 (Aminoethanol), 150 mM NaCl	3.1±1.5
pH 9.3 (Aminoethanol), 150 mM NaCl	2.71±1.2
pH 10.0 (CAPS), 150 mM NaCl	9.2±1.9
NaCl concentration variation	
150 mM NaCl, pH 7.4 (Phosphate)	1.1±0.8
250 mM NaCl, pH 7.4 (Phosphate)	7.9±0.6
350 mM NaCl, pH 7.4 (Phosphate)	33.0±4.3
450 mM NaCl, pH 7.4 (Phosphate)	127±34.4

All K_ds are for $Y_{III}M_5A_{II}$ interacting with $(KR)_5$-sfGFP. All values are means of at least three independent assays.

SI Figures

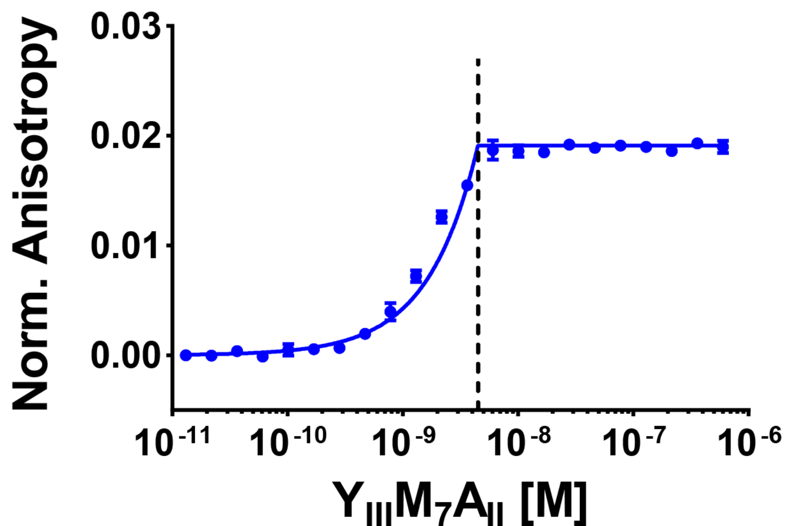

Figure S1: Exemplary saturation binding experiment for the determination of stoichiometry between $Y_{III}M_7A_{II}$ and $(KR)_5$. Here $(KR)_5$_sfGFP (4.5 nM) was titrated with $Y_{III}M_7A_{II}$ and the curve was fit with a 1:1 binding model (only for illustration). The point where $(KR)_5$_sfGFP (constant concentration) and $Y_{III}M_7A_{II}$ have the same concentration (equivalence point) is indicated by a broken vertical line. Since in all cases this corresponds to the concentration were the upper plateau is reached, we concluded that the system shows 1:1 stoichiometry.

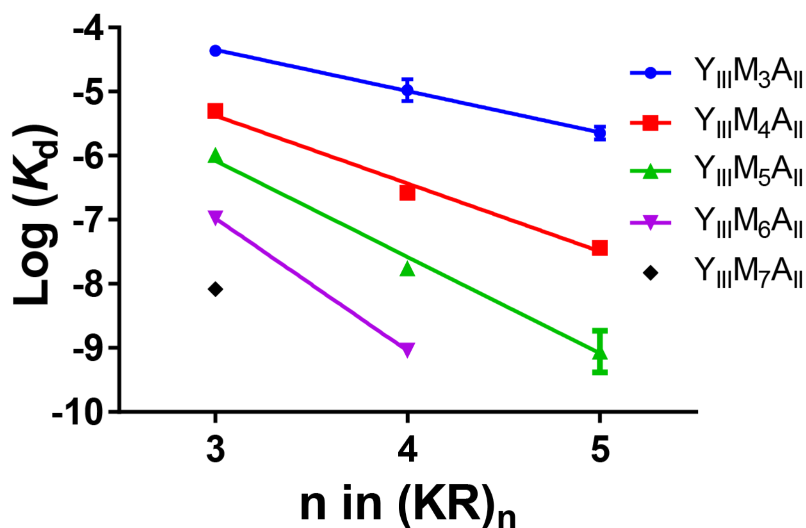

Figure S2: Dependence of K_ds on the length of $(KR)_n$ peptides for different dArmRP. The decimal logarithm of K_ds of a given dArmRP ($Y_{III}M_3A_{II}$ – $Y_{III}M_7A_{II}$) is plotted against the number of (KR)-repeats of the respective peptide. Linear regression is indicated. The following equations and R^2 values were obtained: for $Y_{III}M_3A_{II}$: Y = -0.6450·X -2.412 (R^2 = 0.9480); for $Y_{III}M_4A_{II}$: Y = -1.059·X -2.197 (R^2 = 0.9874); for $Y_{III}M_5A_{II}$: Y = -1.505·X – 1.562 (R^2 = 0.9533); for $Y_{III}M_6A_{II}$: Y = -2.069·X -0.7681 (R^2 = 0.9960); from the slopes ΔΔG values for the addition of one (KR)-repeat can be calculated, they are summarized in SI Table 3.

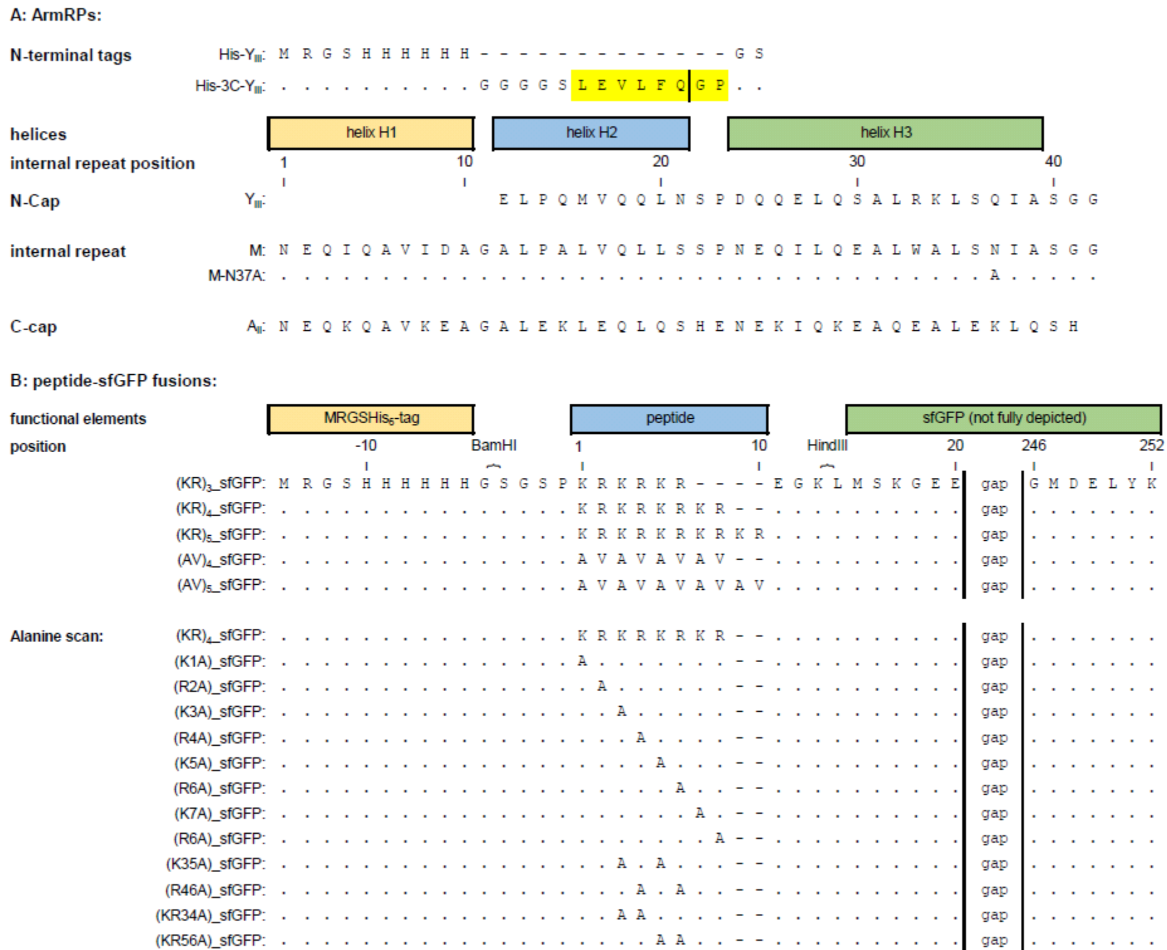

Figure S3: Sequences of dArmRPs and peptide-sfGFP fusions. (A) Sequences of dArmRPs. Amino acid sequences of the functional elements are given in one-letter code. Identical residues are indicated by a period (.), gaps by a hyphen (-). Numbering represents the position within a armadillo repeat. The recognition sequence for 3C-protease is highlighted in yellow with the cutting site as a vertical line. (B) Alignment of peptide-sfGFP fusions. Amino acid sequences are given in one-letter code. Identical residues are indicated by a period (.), gaps by a hyphen (-). Numbering is chosen so that the first peptide residue is number 1.

SI References

(1) Kawe, M.; Forrer, P.; Amstutz, P.; Plückthun, A. *J. Biol. Chem.* **2006**, *281*, 40252.

(2) Madhurantakam, C.; Varadamsetty, G.; Grütter, M. G.; Plückthun, A.; Mittl, P. R. *Protein Sci.* **2012**, *21*, 1015.

(3) Simon, M.; Zangemeister-Wittke, U.; Plückthun, A. *Bioconjugate Chem.* **2012**, *23*, 279.