Supplementary Material

for

Curvature of designed armadillo repeat proteins allows modular peptide binding

Simon Hansen[a,1,†], Patrick Ernst[a,†], Sebastian L. B. König[a], Christian Reichen[a,2], Christina Ewald[a,3] Peer R.E. Mittl[a], Daniel Nettels[a], Ben Schuler[a], Andreas Plückthun[a,*]

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

[†] equal contribution

[1] Present address: Department of Early Discovery Biochemistry, Genentech, 1 DNA Way, South San Francisco, CA 94080, USA

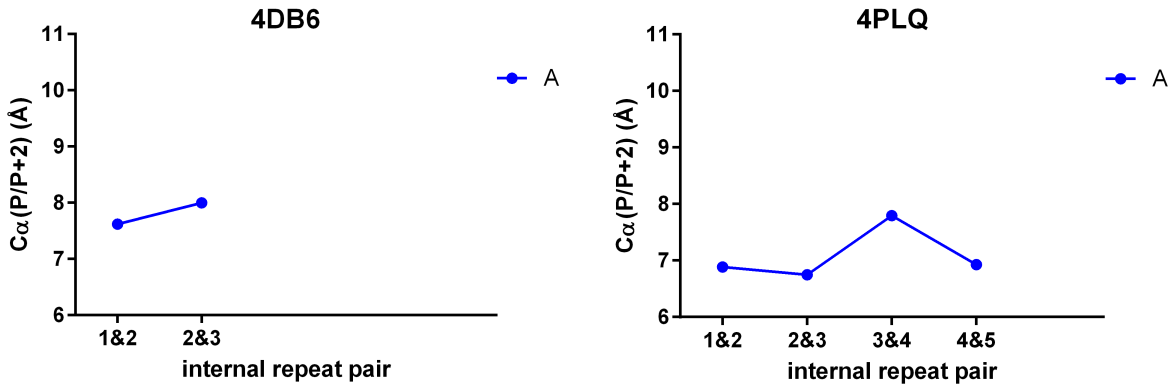
[2] Present address: Molecular Partners AG, Wagistrasse 14, 8952 Schlieren, Switzerland

[3] Present address: Flow Cytometry Facility, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

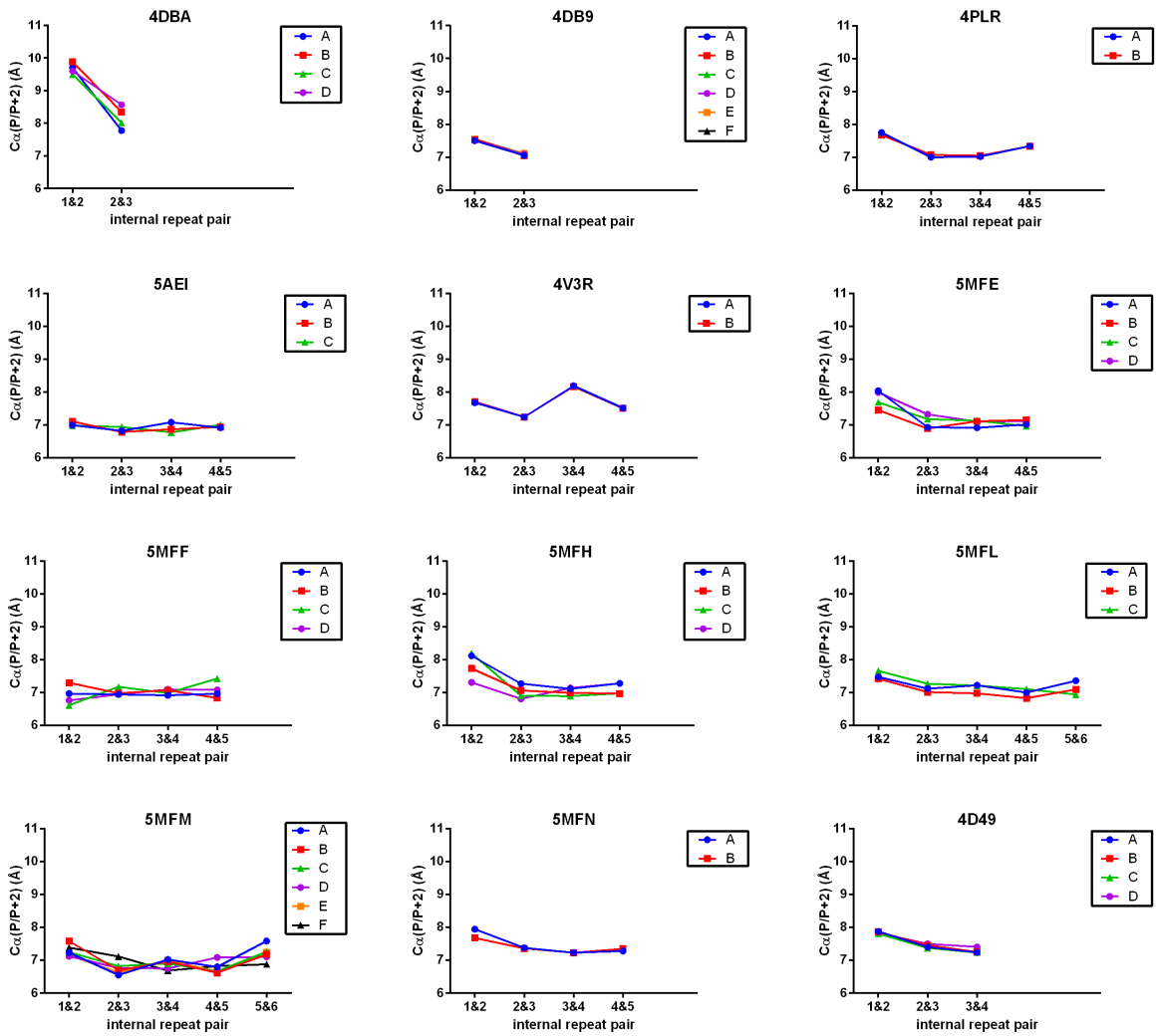
Correspondance;

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

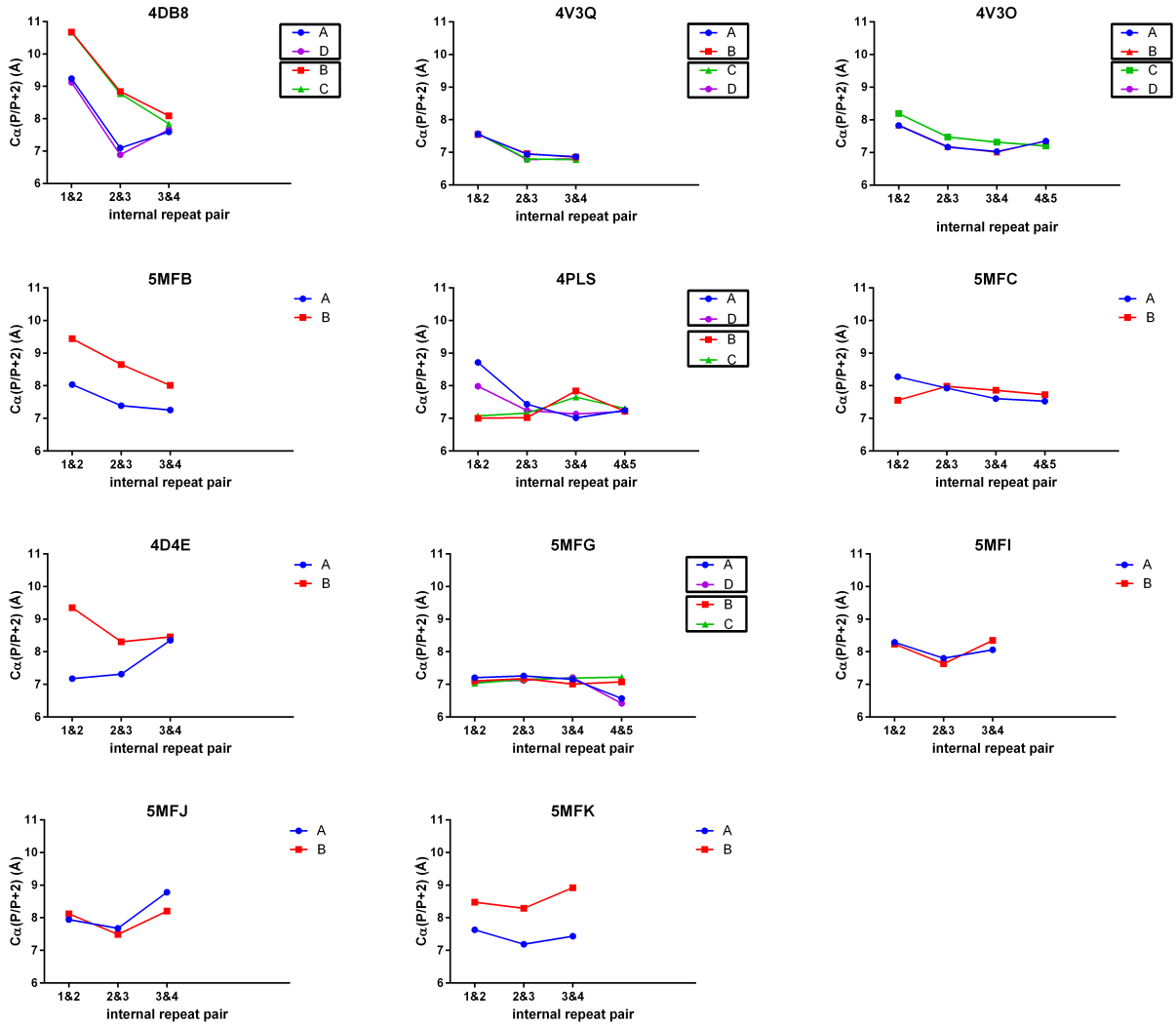
Supplementary Figures and captions



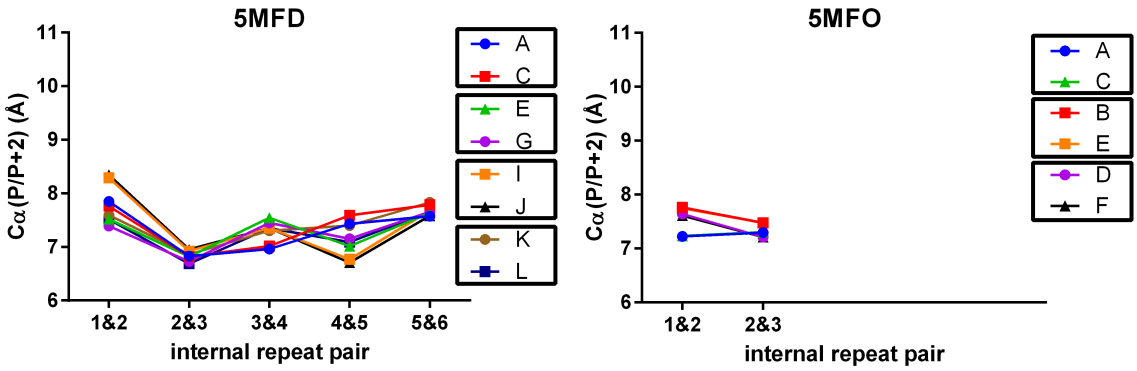
Supplementary Figure S1. Geometry populations of dArmRP – one protein chain per asymmetric unit (AU). C^{α}(P/P+2) of all internal repeat pairs of each chain (chain A) are plotted. The plot title indicates the PDB-ID of the structure analyzed. Here only one chain per AU is present, hence no averaging of C^{α}(P/P+2) is possible.



Supplementary Figure S2. Geometry populations of dArmRP – one population per asymmetric unit (AU). C^{α}(P/P+2) of all internal repeat pairs of each chain (shown in different colors and connected by lines, see legend) are plotted. The plot title indicates the PDB-ID of the structure analyzed. Here only one population per AU is present, hence geometry parameters of the same internal repeat pairs were averaged among all chains in the AU. Justifications for classifications in one or more populations are explained in Materials and Methods in the main text.



Supplementary Figure S3. Geometry populations of dArmRP – two populations per asymmetric unit (AU). C^{α}(P/P+2) of all internal repeat pairs of each chain (shown in different colors and connected by lines, see legend) are plotted. The plot title indicates the PDB-ID of the structure analyzed. Here two populations per AU are present. Hence, in case of only two chains per AU, the geometry parameters of each chain were treated as independent measurement in the analysis dataset. If four chains are present in the AU, the geometry parameters were averaged among chains of the same population for each repeat pair. Chains within the same population are indicated by black boxes around the chain IDs in the legend. Justifications for classifications in one or more populations are explained in Materials and Methods in the main text.



Supplementary Figure S4. Geometry populations of dArmRP – more than two populations per asymmetric unit (AU). C^{α}(P/P+2) of all internal repeat pairs of each chain (shown in different colors and connected by lines, see legend) are plotted. The plot title indicates the PDB-ID of the structure analyzed. Here more than two populations per AU are present. Hence the geometry parameters were averaged among chains of the same population for each internal repeat pair and used as independent measurements in the analysis dataset. Chains within the same population are indicated by black boxes around the chain IDs in the legend. Justifications for classifications in one or more populations are explained in Materials and Methods in the main text.

Supplementary Tables

Supplementary Table ST1.

Crystallization conditions, data collection and refinement statistics.

Protein	$Y_{III}(Dq)_4CqI$	$Y_{III}M_5A_{II}/(RR)_4$
PDB-ID	5MFB	5MFE
Crystalization condition	25 % w/v PEG 2000 MME 0.3 M Na Acetate 0.1 M Na Cacodylate pH 6.5	5% w/v PEG 8000 27.3% v/v MPD 0.2 M $CaCl_2$ 0.1 M Na Cacodylate pH 7.0
Data collection		
Resolution range (Å)	42.81 - 2.30 (2.42 – 2.30)	48.35 – 1.95 (2.00 – 1.95)
Space group	C 2	P 2_1 2_1 2_1
Unit cell parameters		
a, b, c (Å)	139.23, 46.56, 90.81	82.00, 103.67, 134.18
α, β, γ (°)	90, 128.55, 90	90, 90, 90
Unique reflections	20111	83955
Multiplicity	3.8 (4.0)	8.6 (8.8)
Completeness	97.3 (97.1)	99.9 (99.9)
R_{merge}	0.081 (1.85)	0.241 (3.68)
R_{meas}	0.095 (2.14)	0.256 (3.91)
$\langle I \rangle / \sigma I$	10.3 (0.76)	7.94 (0.60)
CC(1/2)	0.999 (0.298)	0.996 (0.155)
Wilson B-factor ($Å^2$)	57.99	33.38
Refinement		
R_{work}	0.184	0.181
R_{free}	0.241	0.230
RMSD of bond lengths (Å)	0.010	0.010
RMSD of bond angles (°)	1.29	1.13
Average B-factor ($Å^2$)	75.98	40.90
Ramachandran plot (%)		
favored	97.89	99.82
allowed	2.11	0.181
outliers	0.00	0.00
Non-hydrogen atoms		
protein	3486	8433
ligands	0	74
waters	69	458

Statistics for highest resolution shell in parentheses

Supplementary Table ST1 (continued).

Crystallization conditions, data collection and refinement statistics.

Protein	$Y_{III}M_5A_{II}$/ $(RR)_5$	$Y_{III}M_5A_{II}$/ $(RR)_4$
PDB-ID	5MFF	5MFG
Crystalization condition	60% sat. $(NH_4)_2SO_4$ 0.1 M HEPES pH 8.0	18 % w/v PEG 3350 0.15 M KSCN 0.2 M $CaCl_2$ 0.1 M Na Acetate pH 5.5
Data collection		
Resolution range (Å)	48.31 -1.90 (1.95 – 1.90)	48.65 – 1.90 (1.95 – 1.90)
Space group	P $2_12_12_1$	P 6_2
Unit cell parameters		
a, b, c (Å)	81.76, 90.89, 179.66	168.54, 168.54, 80.58
α, β, γ (°)	90, 90, 90	90, 90, 120
Unique reflections	106118	102534
Multiplicity	11.0 (11.2)	20.6 (20.5)
Completeness	100 (99.9)	99.9 (98.3)
R_{merge}	0.108 (4.68)	0.138 (6.14)
R_{meas}	0.113 (4.91)	0.141 (6.34)
$\langle I \rangle / \sigma I$	14.51 (0.70)	19.47 (0.64)
CC(1/2)	0.999 (0.273)	1.000 (0.19)
Wilson B-factor ($Å^2$)	44.06	35.80
Refinement		
R_{work}	0.184	0.182
R_{free}	0.213	0.213
RMSD of bond lengths (Å)	0.010	0.003
RMSD of bond angles (°)	1.11	0.496
Average B-factor ($Å^2$)	52.2	44.76
Ramachandran plot (%)		
favored	99.73	98.96
allowed	0.27	0.94
outliers	0	0.09
Non-hydrogen atoms		
protein	8502	7998
ligands	56	13
waters	396	486

Statistics for highest resolution shell in parentheses

Supplementary Table ST1 (continued).

Crystallization conditions, data collection and refinement statistics.

Protein	$Y_{III}M_5A_{II}$/ $(RR)_5$	$Y_{III}(Dq.V2)_4CqI$ / $(KR)_4$
PDB-ID	5MFH	5MFI
Crystalization condition	5% w/v PEG 8000 27.3% v/v MPD 0.2 M $CaCl_2$ 0.1 M Na Cacodylate pH 6.7	25 % w/v PEG 2000 MME 0.3 M Na Acetate 0.1 M Na Acetate pH 5.5
Data collection		
Resolution range (Å)	48.39 – 2.00 (2.05 – 2.00)	46.04 – 1.45 (1.49 – 1.45)
Space group	$P\ 2_12_12_1$	$P\ 2_12_12_1$
Unit cell parameters		
a, b, c (Å)	81.48, 103.79, 134.01	50.80, 89.25, 107.50
α, β, γ (°)	90, 90, 90	90, 90, 90
Unique reflections	77426	87337
Multiplicity	13.1 (11.0)	13.1 (12.5)
Completeness	100 (99.9)	99.9 (99.9)
R_{merge}	0.543 (9.23)	0.117 (5.44)
R_{meas}	0.565 (9.68)	0.118 (5.85)
$<I>/\sigma I$	5.52 (0.29)	15.1 (0.52)
CC(1/2)	0.990 (0.108)	1.00 (0.230)
Wilson B-factor ($Å^2$)	35.55	24.21
Refinement		
R_{work}	0.199	0.202
R_{free}	0.232	0.228
RMSD of bond lengths (Å)	0.008	0.009
RMSD of bond angles (°)	1.05	1.04
Average B-factor ($Å^2$)	44.57	34.92
Ramachandran plot (%)		
favored	99.74	97.94
allowed	0.18	2.06
outliers	0.09	0.00
Non-hydrogen atoms		
protein	8581	3726
ligands	21	0
waters	532	580

Statistics for highest resolution shell in parentheses

Supplementary Table ST1 (continued).

Crystallization conditions, data collection and refinement statistics.

Protein	YIII(Dq.V2)4CqI / $(KR)_5$	$Y_{III}(Dq.V1)_4C_{PAF}/(KR)4$
PDB-ID	5MFJ	5MFK
Crystalization condition	25 % w/v PEG 2000 MME 0.3 M Na Acetate 0.1 M Tris pH 8.5	25 % w/v PEG 2000 MME 0.2 M KSCN 0.1 M NaAcetate pH 5.5
Data collection		
Resolution range (Å)	46.13 - 1.53 (1.57 – 1.53)	40.41 - 2.30 (2.42 – 2.30)
Space group	$P\ 2_12_12_1$	$P\ 2_1$
Unit cell parameters		
a, b, c (Å)	51.11, 89.25, 107.19	67.47, 51.92, 69.68
α, β, γ (°)	90, 90, 90	90, 112.58, 90
Unique reflections	74727	20024
Multiplicity	12.9 (6.2)	3.3 (3.2)
Completeness	99.9 (100)	97.5 (95.4)
R_{merge}	0.093 (5.37)	0.042 (0.366)
R_{meas}	0.097 (5.61)	0.051 (0.44)
$<I>/\sigma I$	16.98 (0.56)	21.8 (5.36)
CC(1/2)	1.00 (0.179)	0.999 (0.918)
Wilson B-factor ($Å^2$)	28.29	42.95
Refinement		
R_{work}	0.206	0.184
R_{free}	0.237	0.242
RMSD of bond lengths (Å)	0.009	0.009
RMSD of bond angles (°)	1.00	1.13
Average B-factor ($Å^2$)	39.51	50.82
Ramachandran plot (%)		
favored	99.59	98.96
allowed	0.41	0.83
outliers	0.00	0.21
Non-hydrogen atoms		
protein	4085	3685
ligands	0	0
waters	476	158

Statistics for highest resolution shell in parentheses

Supplementary Table ST1 (continued).

Crystallization conditions, data collection and refinement statistics.

Protein	$(KR)_5$_GS_{10}_$Y_{III}M_6A_{II}$	$Y_{III}M_6A_{II}$_GS_{11}_$(KR)_5$
PDB-ID	5MFL	5MFM
Crystalization condition	37.5% (MPD/PEG 1000/PEG 3350) 0.03 M $MgCl_2$ 0.03 M $CaCl_2$ Bicine/TRIS pH 8.5	37.5% (MPD/PEG 1000/PEG 3350) 0.03 M $MgCl_2$ 0.03 M $CaCl_2$ Bicine/TRIS pH 8.5
Data collection		
Resolution range (Å)	49.56 - 2.50 (2.58 – 2.50)	44.69 – 2.3 (2.36 – 2.30)
Space group	P $3_2$2 1	P 2_1
Unit cell parameters		
a, b, c (Å)	93.10, 93.10, 188.48	81.76, 89.39, 123.85
α, β, γ (°)	90, 90, 120	90, 97.07, 90
Unique reflections	33299	77719
Multiplicity	6.2 (6.3)	4.2 (4.3)
Completeness	99.4 (99.8)	98.6 (99.1)
R_{merge}	0.165 (2.34)	0.114 (1.63)
R_{meas}	0.179 (2.55)	0.131 (1.86)
$\langle I \rangle / \sigma I$	8.21 (0.61)	8.67 (1.07)
CC(1/2)	0.998 (0.191)	0.999 (0.301)
Wilson B-factor ($Å^2$)	66.91	50.87
Refinement		
R_{work}	0.206	0.217
R_{free}	0.255	0.251
RMSD of bond lengths (Å)	0.008	0.008
RMSD of bond angles (°)	1.11	1.03
Average B-factor ($Å^2$)	76.69	78.71
Ramachandran plot (%)		
favored	99.70	98.79
allowed	0.30	1.21
outliers	0.00	0.00
Non-hydrogen atoms		
protein	7441	15536
ligands	42	57
waters	156	358

Statistics for highest resolution shell in parentheses

Supplementary Table ST1 (continued).

Crystallization conditions, data collection and refinement statistics.

Protein	$Y_{III}M_5A_{II}$	$Y_{III}M_3A_{III}$
PDB-ID	5MFN	5MFO
Crystalization condition	2.4 M Na Malonate pH 7.0 0.2 M $CaCl_2$	25% w/v PEG 2000 MME 0.2 M Ca Acetate 0.1 M Na Acetate pH 5.5
Data collection		
Resolution range (Å)	46.66 - 2.80 (2.94 – 2.80)	48.29 - 1.30 (1.33 – 1.30)
Space group	P 6_32 2	P 4_3
Unit cell parameters		
a, b, c (Å)	142.55, 142.55, 142.12	83.92, 83.92, 166.15
α, β, γ (°)	90, 90, 120	90, 90, 90
Unique reflections	21564	278605
Multiplicity	39.7 (37.4)	11.2 (11.1)
Completeness	99.9 (99.9)	99.3 (98.1)
R_{merge}	0.257 (7.643)	0.056 (1.70)
R_{meas}	0.260 (7.75)	0.058 (1.78)
$<I>/\sigma I$	18.68 (0.69)	19.48 (1.44)
CC(1/2)	1.00 (0.184)	1.00 (0.475)
Wilson B-factor ($Å^2$)	89.30	17.83
Refinement		
R_{work}	0.200	0.159
R_{free}	0.246	0.178
RMSD of bond lengths (Å)	0.010	0.008
RMSD of bond angles (°)	1.320	0.898
Average B-factor ($Å^2$)	131.9	27.83
Ramachandran plot (%)		
favored	98.38	99.58
allowed	1.08	0.42
outliers	0.54	0
Non-hydrogen atoms		
protein	4133	9686
ligands	21	23
waters	16	1305

Statistics for highest resolution shell in parentheses

Supplementary Table ST2: Populations of dArmRP chains with the same curvature within structures

PDB-ID	category	populations (chain IDs)			
		first population	second population	third population	fourth population
4DBA	one population	A, B, C, D	-	-	-
4DB9	one population	A, B, C, D, E, F	-	-	-
4DB6	one chain	A	-	-	-
4DB8	two populations	A, D	B, C	-	-
4V3Q	two populations	A, B	C, D	-	-
4V3O	two populations	A, B	C, D	-	-
4PLR	one population	A, B	-	-	-
5MFB	two populations	A	B	-	-
4PLS	two populations	A, D	B, C	-	-
4PLQ	one chain	A	-	-	-
5AEI	one population	A, B, C	-	-	-
4V3R	one population	A, B	-	-	-
5MFC	two populations	A	B	-	-
5MFD	> 2 populations	A, C	E, G	I, J	K, L
4D4E	two populations	A	B	-	-
5MFE	one population	A, B, C, D	-	-	-
5MFF	one population	A, B, C, D	-	-	-
5MFG	two populations	A, D	B, C	-	-
5MFH	one population	A, B, C, D	-	-	-
5MFI	two populations	A	B	-	-
5MFJ	two populations	A	B	-	-
5MFK	two populations	A	B	-	-
4D49	one population	A, B, C, D	-	-	-
5MFL	one population	A, B, C	-	-	-
5MFM	one population	A, B, C, D, E, F	-	-	-
5MFN	one population	A, B, C	-	-	-
5MFO	> 2 populations	A, C	B, E	D, F	-

Justifications for classifications in one or more populations are explained in Materials and Methods in the main text.