

Structures of designed armadillo repeat proteins binding to peptides fused to globular domains

Simon Hansen, Jonathan D. Kiefer, Chaithanya Madhurantakam,
 Peer R. E. Mittl, and Andreas Plückthun *

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Received 16 May 2017; Accepted 3 July 2017
 DOI: 10.1002/pro.3229
 Published online 10 July 2017 proteinscience.org

Abstract: Designed armadillo repeat proteins (dArmRP) are α -helical solenoid repeat proteins with an extended peptide binding groove that were engineered to develop a generic modular technology for peptide recognition. In this context, the term “peptide” not only denotes a short unstructured chain of amino acids, but also an unstructured region of a protein, as they occur in termini, loops, or linkers between folded domains. Here we report two crystal structures of dArmRPs, in complex with peptides fused either to the N-terminus of Green Fluorescent Protein or to the C-terminus of a phage lambda protein D. These structures demonstrate that dArmRPs bind unfolded peptides in the intended conformation also when they constitute unstructured parts of folded proteins, which greatly expands possible applications of the dArmRP technology. Nonetheless, the structures do not fully reflect the binding behavior in solution, that is, some binding sites remain unoccupied in the crystal and even unexpected peptide residues appear to be bound. We show how these differences can be explained by restrictions of the crystal lattice or the composition of the crystallization solution. This illustrates that crystal structures have to be interpreted with caution when protein–peptide interactions are characterized, and should always be correlated with measurements in solution.

Keywords: protein–peptide interactions; armadillo repeat; solenoid proteins; repeat proteins; protein engineering; protein crystallization

Introduction

Natural armadillo repeat proteins (nArmRP) are α -solenoid proteins which bind to unstructured regions

of their targets in processes of signal transduction or nuclear transport. Each armadillo repeat is composed of 42 residues¹ that fold into three helices (H1, H2,

Importance Statement We are developing a modular technology for generic peptide recognition, based on designed armadillo repeat proteins (dArmRP). Here we show that dArmRPs bind unstructured peptide stretches of folded proteins in the intended manner. This is a crucial milestone for future applications of the technology. Furthermore, we point out how crystal packing can induce significant deviations of the observed protein–peptide interaction from the behavior in solution, cautioning against an uncritical interpretation of such complex structures.

Grant sponsor: Swiss National Science Foundation; Grant numbers: CRSI33_122686, CRSI_141832/1; Grant sponsor: Forschungskredit of the University of Zurich; Grant number: FK-13-028.

Present address: Simon Hansen, Department of Early Discovery Biochemistry, Genentech, South San Francisco, California 94080

Present address: Jonathan D. Kiefer, Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH Zurich), 8093 Zürich, Switzerland

Present address: Chaithanya Madhurantakam, Department of Biotechnology, Structural and Molecular Biology Laboratory (SMBL), TERI University, 10, Institutional Area, Vasant Kunj, New Delhi 110070, India

*Correspondence to: Andreas Plückthun, Department of Biochemistry, University Zürich, Zürich 8057, Switzerland. E-mail: plueckthun@bioc.uzh.ch

and H3) in a triangular arrangement. The stacked repeats, capped at both termini with special repeats to protect the hydrophobic core from solvent exposure, build up a continuous superhelical domain with an extended peptide-binding groove defined by the adjacent H3 helices.^{2–5} These proteins possess an intrinsically conserved modular binding mode towards peptides. Upon complex formation, the peptide and the nArmRP form a complex where the N- to C-terminal directions of peptide and protein are antiparallel. Peptides are bound in an extended conformation, and the peptide main chain of every second residue is fixed in a β -strand conformation by two hydrogen bonds with conserved asparagine residues (N³⁷, superscripted numbers refer to positions within one repeat) on the H3 helices. The specificity for the peptide sequence is mediated by contacts to the peptide side chains, and modularity is ensured, since each repeat binds two residues of the peptide.^{3,6–10}

Designed armadillo repeat proteins (dArmRP) were engineered based on natural armadillo repeat proteins (nArmRP) by consensus design and computational optimization of the hydrophobic core,¹¹ molecular dynamics simulations to further improve thermodynamic stability,¹² and rational engineering of the N-terminal capping repeat based on crystal structures.¹³ The final construct, called Y_{III}M_xA_{II} (the roman numbers indicate the design cycle of the caps, x stands for the number of internal repeats), is monomeric, has high expression yields in *E. coli*, improved biophysical stabilities compared to nArmRP, and adopts the expected solenoid fold.¹³ An overview of all engineering steps is given in Ref. 14.

Our goal is to develop a generic detection technology based on dArmRP for any unstructured peptide sequence, especially in the context of unstructured regions of folded proteins. Such a technology would significantly speed up the process of creating affinity reagents against linear epitopes as signature sequences of given proteins. dArmRPs are especially suited as a scaffold for such a technology since they are formed by several repeats that stack on one another, allowing facile modification of the length of the binding groove by introduction or omission of repeats. Furthermore, they adopt more uniform super-helical curvatures than the more irregular nArmRP, which thus allows binding of longer peptide stretches.^{13–16}

Previously, we described the first structure and detailed characterization of the high-affinity interaction between Y_{III}M_xA_{II} and peptides made of alternating lysine and arginine residues ((KR)_n). This study showed that our design exhibited the expected characteristics of a modular peptide binder. The binding affinity was modulated by changing the length of the interaction partners and a first Y_{III}M₅A_{II}:(KR)₅ complex structure (between dArmRP and a synthetic peptide, PDB ID: 5AEI) revealed almost the same interactions in each dipeptide:protein-repeat unit. The regularity was

disturbed only by a crystal contact between symmetry-related peptides.¹⁶

Here we present structural evidence that dArmRP cannot only interact with synthetic peptides but also with unstructured regions of folded protein domains in a very similar manner. Two high-resolution structures are presented, one of Y_{III}M₅A_{II} interacting with a N-terminal fusion of (KR)₄ to superfolder Green Fluorescent Protein ((KR)₄_sfGFP)¹⁷ and of Y_{III}M₆A_{II} bound to (KR)₅ that was C-terminally fused to a phage lambda protein D domain (pD_(KR)₅).^{18,19} We analyze the structures with a focus on the dArmRP–peptide interaction and show how the protein fusions and crystallization conditions influence the observed binding mode and how it differs from binding in solution.

Results

Binding stoichiometry determination by SEC-MALS

To determine the stoichiometry of dArmRP and peptide fusions in solution we performed size exclusion chromatography experiments with a multi-angle light scattering detector (SEC-MALS). dArmRPs and 1:1 mixtures of dArmRPs and either pD_(KR)₅ or (KR)₄_sfGFP were injected at 1 mg mL⁻¹. All samples eluted as single peaks and the calculated molecular masses corresponded very closely to the theoretical molecular weights of monomeric proteins or 1:1 complexes. Therefore, we concluded that the 1:1 complexes are the predominant species in solution [Fig. 1(A), Table I].

Many dArmRP structures contain dimers of dArmRPs in the crystal that are crosslinked by Ca²⁺ ions binding to carboxylate groups on different molecules, and thereby interact with each repeat in a zipper-like fashion.²⁰ This motif is also found in the Y_{III}M₆A_{II}:pD_(KR)₅ structure (see below). We conducted SEC-MALS measurements to elucidate whether these dimers are also observed in solution. Y_{III}M₅A_{II} (1 mg mL⁻¹) was injected once with Tris buffered saline (TBS), and once with TBS that was supplemented with 200 mM CaCl₂ as running buffer. These measurements resulted in identical molecular weights that suggest monomeric proteins, but a significant increase of elution volumes was observed, suggesting either a strongly compacted structure or a Ca²⁺-induced interaction with the column material. A similar shift was observed with bovine serum albumin as a control protein in the same running buffers. Hence, we conclude that Ca²⁺ ions do not influence the oligomerization state of dArmRP in solution. However, the buffer composition most likely influences column interactions, which results in increased elution volumes [Fig. 1(B), Table I].

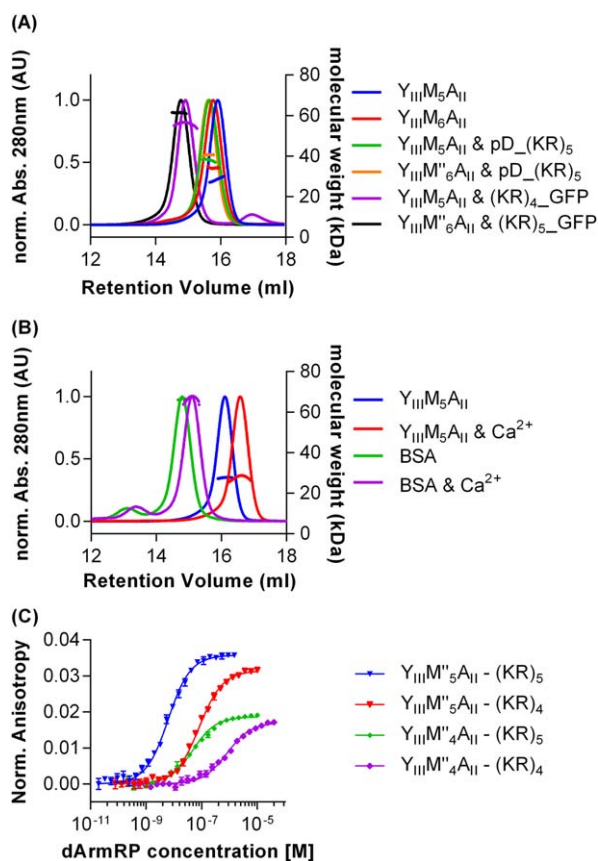


Figure 1. Size exclusion chromatography multi-angle light scattering (SEC-MALS) and fluorescence anisotropy (FA) experiments. (A) Mixtures of dArmRP and pD or sfGFP peptide fusions and dArmRPs alone were analyzed on a SEC-MALS instrument, chromatograms are shown as lines, MALS data as dots, extracted molecular weights are given in Table I. (B) dArmRP and bovine serum albumin (BSA) were analyzed by SEC-MALS in different buffers (TBS or TBS with CaCl_2). Presentation as in (A), extracted molecular weights are given in Table I. (C) Examples of fluorescence anisotropy (FA) measurements (symbols) for different combinations of dArmRP and peptide-sfGFP fusions with fits (lines). Extracted K_D values are given in Table II.

Affinity of M'' -type internal repeats to $(KR)_n$ peptides

Up to now, many dissociation constants (K_D) have been determined for dArmRPs containing M-type repeats.¹⁶ Here we are investigating proteins with M'' -type repeats, which differ from M-type repeats by carrying two point mutations: mutation $S^{36}G$ (superscripted numbers indicate positions within an internal repeat as shown in Fig. 2) provides more space to N^{37} to optimally bind peptide backbones, and mutation $A^{34}T$ was introduced to exert a stronger twist of the superhelix.¹⁵ The K_D values between $Y_{III}M''_x A_{II}$ and $(KR)_n$ -sfGFP fusions were determined by fluorescence anisotropy [FA, Fig. 1(C)], and the mean K_D s of at least three independent assays are reported in Table II. As already observed for $Y_{III}M_x A_{II}$, affinities for $Y_{III}M''_x A_{II}$ also increase for longer $(KR)_n$ peptides or for proteins with more internal repeats. A comparison of identical dArmRP:peptide combinations (same length of $(KR)_n$ peptide and same number of internal protein repeats) reveals that affinities of M'' -type internal repeats are lower by a factor of two to five, compared to the M-type repeat [Fig. 1(C), Table II].

Structure of $Y_{III}M''_6 A_{II}$:pD $_{(KR)_5}$

The structure of the complex $Y_{III}M''_6 A_{II}$:pD $_{(KR)_5}$ was determined at 2.3-Å resolution, which is surprisingly good, considering a solvent content of 70.2%. Crystallization condition, data processing and refinement statistics are given in Table III. The asymmetric unit (AU) consists of eight dArmRP and four pD-domains fused to $(KR)_5$. The eight dArmRP chains form four dimers (chains A and C, E and I, G and J, K and L) linked via Ca^{2+} ions that are complexed in the loops between helices H2 and H3 of two dArmRP, as described previously.^{15,20} As detailed above, these dimers are not formed in solution. In the crystal, dArmRP-dimers are bridged by pD-domains. This arrangement is extended to neighboring AUs to form

Table I. Molecular Weights (MW) Extracted From SEC-MALS Experiments

Samples	Retention volume (mL)	Measured MW (kDa)	Calculated MW (kDa)	Ratio ^a
$Y_{III}M_5 A_{II}$ ^b	15.91	28.6	30.0	0.95
$Y_{III}M''_6 A_{II}$ ^b	15.78	34.1	34.3	0.99
$Y_{III}M_5 A_{II}$ & $(KR)_4$ -sfGFP ^b	14.92	56.8	58.9	0.96
$Y_{III}M_5 A_{II}$ & pD $_{(KR)_5}$ ^b	15.65	38.4	41.5	0.93
$Y_{III}M''_6 A_{II}$ & $(KR)_5$ -sfGFP ^b	14.78	62.0	63.5	0.98
$Y_{III}M''_6 A_{II}$ & pD $_{(KR)_5}$ ^b	15.62	41.8	45.8	0.91
$Y_{III}M_5 A_{II}$ ^c	16.10	27.7	30.0	0.92
$Y_{III}M_5 A_{II}$ & Ca^{2+} ^d	16.58	27.8	30.0	0.93
Bovine serum albumin ^{c,e}	14.79	66.5	66.5	1.00
Bovine serum albumin & Ca^{2+} ^{d,e}	15.10	66.7	66.5	1.00

^a Ratio = (measured MW)/(calculated MW).

^b Measured in PBS.

^c Measured in TBS.

^d Measured in TBS with 200 mM CaCl_2 .

^e Data for the major elution peak.

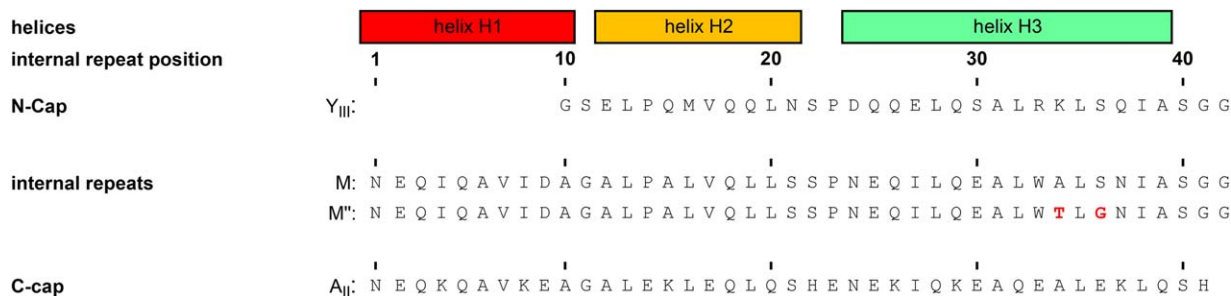


Figure 2. Sequence alignment of dArmRP. Differences in sequences between M and M' internal repeats are highlighted in red. The positions of the three helices that form an armadillo repeat are schematically shown as boxes on top.

long rods of dArmRP and pD-domains [Fig. 3(A)]. The distribution of pD-peptide fusions among dArmRP dimers is complex. In the crystal, there are two dArmRP dimers where only one dArmRP interacts with a pD-peptide fusion (chain E of dimer EI binds pD-peptide fusion F and chain G of dimer GJ binds pD-peptide fusion H), while the AC dimer binds two pD-peptide fusions simultaneously (chains AC and BD) and the KL dimer does not bind any pD-peptide fusion at all [Fig. 3(A)]. All dArmRP/pD-peptide fusion interactions are observed among molecules in the AU and there are no isolated pD-peptide fusions that could bind to symmetry-related dArmRPs. Whether a (KR)₅ peptide is bound or not by the dArmRP depends on the crystal contacts between the dArmRP and the folded portion of the pD-fusion: dArmRPs that interact with a peptide are involved in only few crystal contacts of the N- or C-terminal capping repeat, whereas non-interacting dArmRPs mediate crystal contacts also via the binding surface of internal repeats five and six to the folded pD-domain. To compare the overall dArmRP structures C α atoms were superimposed and RMSD values were calculated. Superimposing all dArmRP chains on each other results in a mean RMSD value \pm SD of 0.43 ± 0.17 Å (with a maximal RMSD of 0.71 Å of chain G vs. chain J), demonstrating that all dArmRP have very similar structures.

The four Y_{III}Mⁿ₆A_{II}:pD-(KR)₅ complexes within the AU can be divided into two groups (AB, EF and CD, GH, named after the chain IDs) that differ by the relative orientation of the pD-domain to the dArmRP [Fig. 3(B)]. Superposition of all complexes on Y_{III}M₅A_{II}:(KR)₅, the complex structure with the

unfused peptide, shows that the peptide is less regularly bound in Y_{III}Mⁿ₆A_{II}:pD-(KR)₅. Furthermore, the four Y_{III}Mⁿ₆A_{II}:pD-(KR)₅ complexes from the AU are not identical as they exhibit different positions of the pD-domain relative to the dArmRP, which puts strain on the peptide and thereby influences peptide binding. The orientation in complexes CD and GH is better suited for the peptide to bind to the binding groove. In complexes AB and EF, the C-terminus of the pD-domain to which (KR)₅ is fused is positioned further away from the binding groove than in complexes CD and GH; as a consequence, not the entire peptide is bound in AB and EF [Fig. 3(C)]. Therefore, only two and three (complex EF and AB, respectively) out of ten possible conserved bidentate hydrogen bonds between N³⁷ and the peptide backbone are formed (distance cut-off: 3.6 Å). In complexes CD and GH, the positioning of the (KR)₅ peptide is closer to the binding groove, but also here only five out of the ten possible hydrogen bonds between N³⁷ and peptide backbone are formed (distance cut-off: 3.6 Å). This result shows that in all complexes strain is applied to the peptide by the pD-fusion, which has to fit into the crystal lattice. Analysis of the binding interface between (KR)₅ and Y_{III}Mⁿ₆A_{II} in complex CD, which is most similar to the unfused peptide complex Y_{III}M₅A_{II}:(KR)₅, shows that the peptide and the dArmRP form an antiparallel complex (N- to C-terminal directions run opposite to each other for peptide and dArmRP) and the expected residues (E³⁰, W³³, N³⁷, and S⁴⁰) of Y_{III}Mⁿ₆A_{II} are involved in binding [Fig. 3(D), top]. For complex EF [Fig. 3(D), bottom], which deviates the most from Y_{III}M₅A_{II}:(KR)₅, antiparallel binding is also established, but as mentioned above, only few N³⁷s are involved in binding, and especially some lysine side chains point away from the binding surface. In all complexes, the first (KR) dipeptide, closest to pD, interacts mainly with residues of dArmRP internal repeats four and five. This means that dArmRP internal repeat six does not contribute to binding at all and consequently the C-terminal (KR) dipeptide of (KR)₅ is overhanging towards the N-cap of the dArmRP, and thus not resolved in the electron density.

Table II. K_D s of dArmRPs Interacting With Different Peptides

	$K_D \pm$ S.D. (nM)			
	(KR) ₄	(KR) ₅	(KK) ₄	(RR) ₄
Y _{III} M ⁿ ₄ A _{II}	670 \pm 150	63 \pm 14	n.d.	n.d.
Y _{III} M ⁿ ₅ A _{II}	38 \pm 4.6	5.4 \pm 0.2	n.d.	n.d.
Y _{III} M ₄ A _{II}	265 \pm 23 ^a	36 \pm 1.2 ^a	14,200 \pm 3000	28 \pm 3.3
Y _{III} M ₅ A _{II}	18 \pm 3.5 ^a	1.1 \pm 0.8 ^a	1400 \pm 150	2.0 \pm 0.3

n.d.: not determined.

^a Data from Hansen *et al.*, 2016.¹⁶

Table III. Crystallization Conditions, Data Collection, and Refinement Statistics

Complex PDB-ID	$Y_{III}M^p_6A_{II}:pD_{-}(KR)_5$ 5MFD	$Y_{III}M_5A_{II}:(KR)_4$ _sfGFP 5MFC
Crystallization condition	28.0% v/v PEG 400 0.2 M CaCl ₂ 0.1 M Na HEPES pH 7.5	8% w/v PEG 4000 0.1 M Na Acetate pH 4.6
Data collection		
Resolution range (Å)	49.12–2.30	40.56–2.40
Space group	P 6 ₃	C 2 2 2 ₁
Unit cell parameters		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	194.66, 194.66, 241.74	81.11, 124.24, 245.33
α , β , γ (°)	90, 90, 120	90, 90, 90
Unique reflections	229078	47804
Multiplicity	10.4 (10.2)	12.4 (12.8)
Completeness	99.9 (99.7)	97.9 (98.2)
<i>R</i> _{merge}	0.175 (5.11)	0.128 (4.59)
<i>I</i> / σ (<i>I</i>)	11.02 (0.61)	15.17 (0.58)
CC(1/2)	0.999 (0.345)	0.999 (0.174)
Wilson <i>B</i> -factor (Å ²)	60.92	72.39
Refinement		
<i>R</i> _{work}	0.190	0.218
<i>R</i> _{free}	0.214	0.24
RMSD of bond lengths (Å)	0.009	0.01
RMSD of bond angles (°)	1.11	1.25
Average <i>B</i> -factor (Å ²)	76.08	99.2
Ramachandran plot (%)		
Favored	98.96	98.45
Allowed	0.91	1.55
Outliers	0.13	0
Nonhydrogen atoms		
Protein	22206	8037
Ligands	49	4
Waters	485	78

Statistics for highest resolution shell in parentheses.

Structure of $Y_{III}M_5A_{II}:(KR)_4$ _sfGFP

The complex structure between $Y_{III}M_5A_{II}$ and $(KR)_4$ _sfGFP was determined at 2.4 Å resolution. Crystallization condition, data collection and refinement statistics are reported in Table III. The AU contains two 1:1 complexes [Fig. 4(A)]. To avoid negative residue numbers in the peptide that was fused N-terminally to sfGFP but still stay close to the common numbering convention of GFP (the chromophore is formed from residues 64–66) we added 1000 to all residues of $(KR)_4$ _sfGFP. The $(KR)_4$ peptide is bound by $Y_{III}M_5A_{II}$ via its designated binding groove in the expected antiparallel orientation [Fig. 4(B)]. Compared to the complex $Y_{III}M_5A_{II}:(KR)_5$, the binding in $Y_{III}M_5A_{II}:(KR)_4$ _sfGFP is less regular [Fig. 4(C)]. Most importantly, it is not exclusively the $(KR)_4$ peptide which is bound by $Y_{III}M_5A_{II}$, but the linker between $(KR)_4$ and sfGFP also participates in binding (residues Glu997–Gly998–Lys999–Leu1000). In complex AB, Leu1000 (the last residue of the linker) is bound on both its NH and CO group with the typical bidentate hydrogen bonds to N³⁷ (in this case Asn79), and the Leu side chain is involved in hydrophobic interactions with W³³. Lys999 forms a hydrogen bond with S⁴⁰. Hence, for these two residues the interaction is similar to the interaction of (KR) dipeptides in

$Y_{III}M_5A_{II}:(KR)_5$ when Leu1000 takes the place of the arginine. Binding of the preceding peptide residues deviates more from $Y_{III}M_5A_{II}:(KR)_5$, even though they constitute KR pairs, since the two linker residues, Gly998 and Glu997, also have to be accommodated. This is achieved by a sharp kink of the peptide backbone at this position. Gly998 interacts with N³⁷ (Asn121), while Glu997 occupies the arginine binding pocket formed by Trp117 and Trp159. This is at first surprising, considering the opposite charges of Glu997 and arginine, but crystals were obtained at a pH value of 4.6. Hence, Glu997 can be regarded as protonated and Glu156 (labeled as E³⁰ in the figure), which usually forms a salt bridge with an arginine, flips its side chain away from the pocket [Fig. 4(D)]. Because Glu997 occupies an arginine binding pocket, the neighboring Arg996 occupies the lysine binding pocket. This behavior is seen for all residues towards the N-terminus, meaning that all lysines bind in the arginine pockets and vice versa [Fig. 4(D)].

The peptide adopts the expected backbone binding mode forming bidentate main chain interactions with N³⁷. Only the H-bonds between Asn247 and LysK991 are stretched (3.8 Å for Asn247-OD...Lys991-N and 3.6 Å for Asn247-ND...Lys991-O). Because of the register shift, the N³⁷ hydrogen bonds fix the backbone of

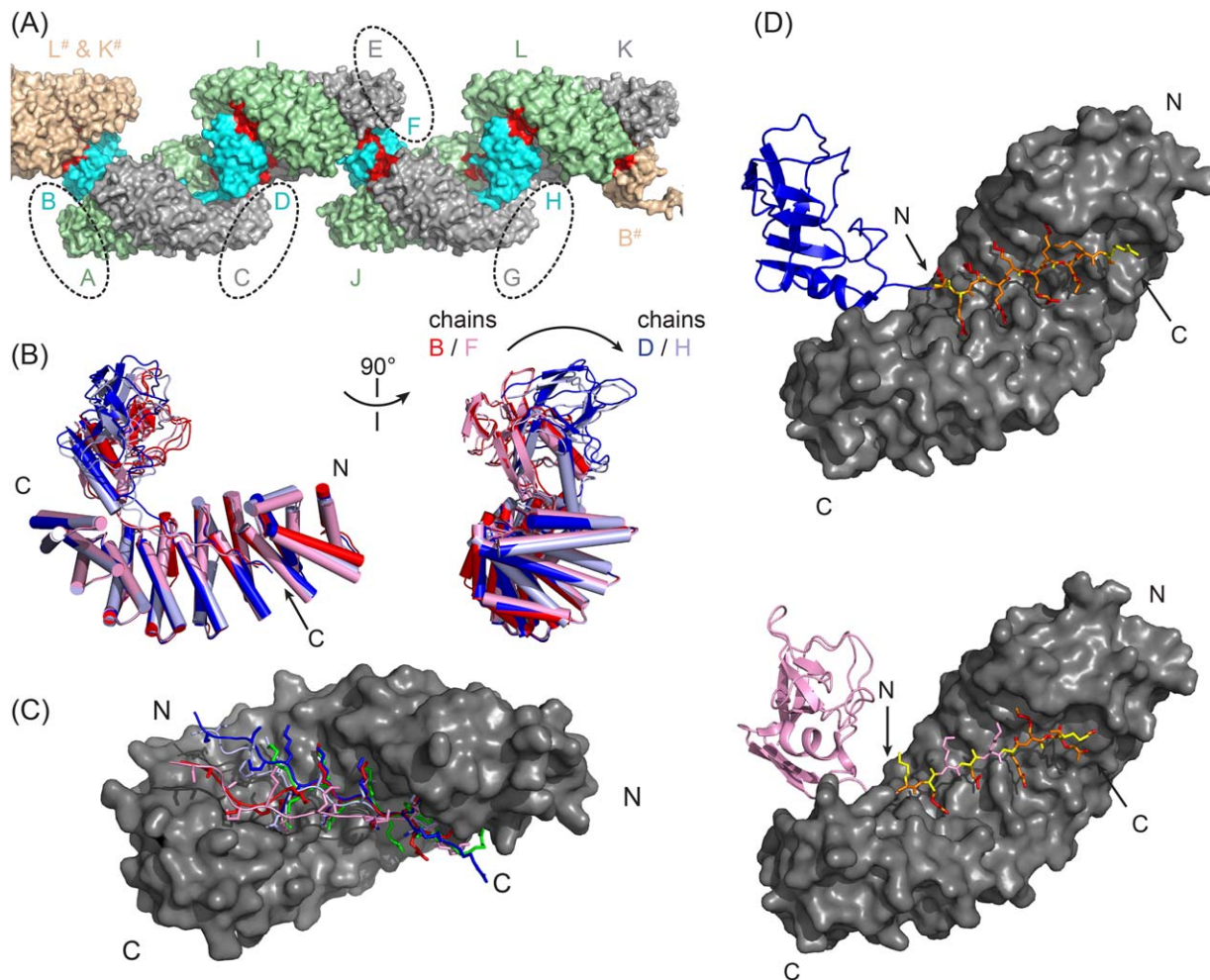


Figure 3. Analysis of the $Y_{III}M^{\#}_6A_{II}:$ pD₅(KR)₅ structure. (A) Overview of the asymmetric unit (AU), dArmRP are shown in green and grey, pD-domains in cyan, molecules from symmetry-related AUs in beige. Crystal contacts ($< 4 \text{ \AA}$) between dArmRP and pD-domains (excluding contacts of the (KR)₅-peptide) in red. Chain IDs are given in the respective colors, # indicates symmetry-related chains. Chain labels of dArmRP:peptide complexes are connected by broken ovals. (B) Superposition of the four complexes between dArmRP and pD₅(KR)₅ viewed from two angles. N- and C-termini of both proteins are indicated on the left. The four complexes are shown in different colors (AB: red, CD: blue, EF: pink, GH: light blue). (C) Superposition of the four $Y_{III}M_5A_{II}:(KR)_5$ complexes. The dArmRP is shown as a grey surface, the unfused peptide of $Y_{III}M_5A_{II}:(KR)_5$ in green, all other peptides in the same color as in (B), the pD-domain is omitted for clarity. (D) Complexes CD (top) and EF (bottom): dArmRP as grey surface, the pD-domain as cartoon and the peptide colored according to distance from dArmRP (red: $< 3.6 \text{ \AA}$, orange: $< 5 \text{ \AA}$, yellow: remaining atoms of residues with a contact $< 5 \text{ \AA}$, pink: $> 5 \text{ \AA}$).

lysines (binding to its NH and CO) and not the backbone of arginines, as seen previously in the $Y_{III}M_5A_{II}:(KR)_5$ structure. Side chains of Lys995 and Lys993 interact with W³³ but not with E³⁰. Side chains of Arg996, Arg994, and Arg992 interact with S⁴⁰ and, in some cases, also with N¹. Because a substantial part of the binding groove is occupied by the linker, the N-terminus of (KR)₄ is pointing toward the C-terminus of the dArmRP [Fig. 4(D)].

Complex CD is similar to complex AB with essentially the same interactions. However, the whole ArmRP is shifted relative to the peptide by one repeat toward the peptide C-terminus. This means that in complex AB internal repeats one to five interact with the peptide, whereas in the case of complex CD, internal repeats two to five bind the peptide, one repeat

less than in complex CD [Fig. 4(B)]. Hence, a larger portion of the peptide is overhanging at the C-terminus of the dArmRP in complex CD, compared to complex AB, and the first residue K989 is not resolved in the electron density map. The observation of a shift by one repeat relative to the peptide is not unexpected, since the internal repeats all have exactly the same sequence and are capable of interacting in the same way. However, two different registers have not been observed within the same X-ray structure before.

Affinities of (RR)₄ and (KK)₄ peptides

In the structure of $Y_{III}M_5A_{II}:(KR)_4$ -sfGFP the lysine/arginine recognition of the peptides is flipped; arginines bind to lysine pockets whereas lysines bind to arginine pockets. To estimate the effect of this flipping,

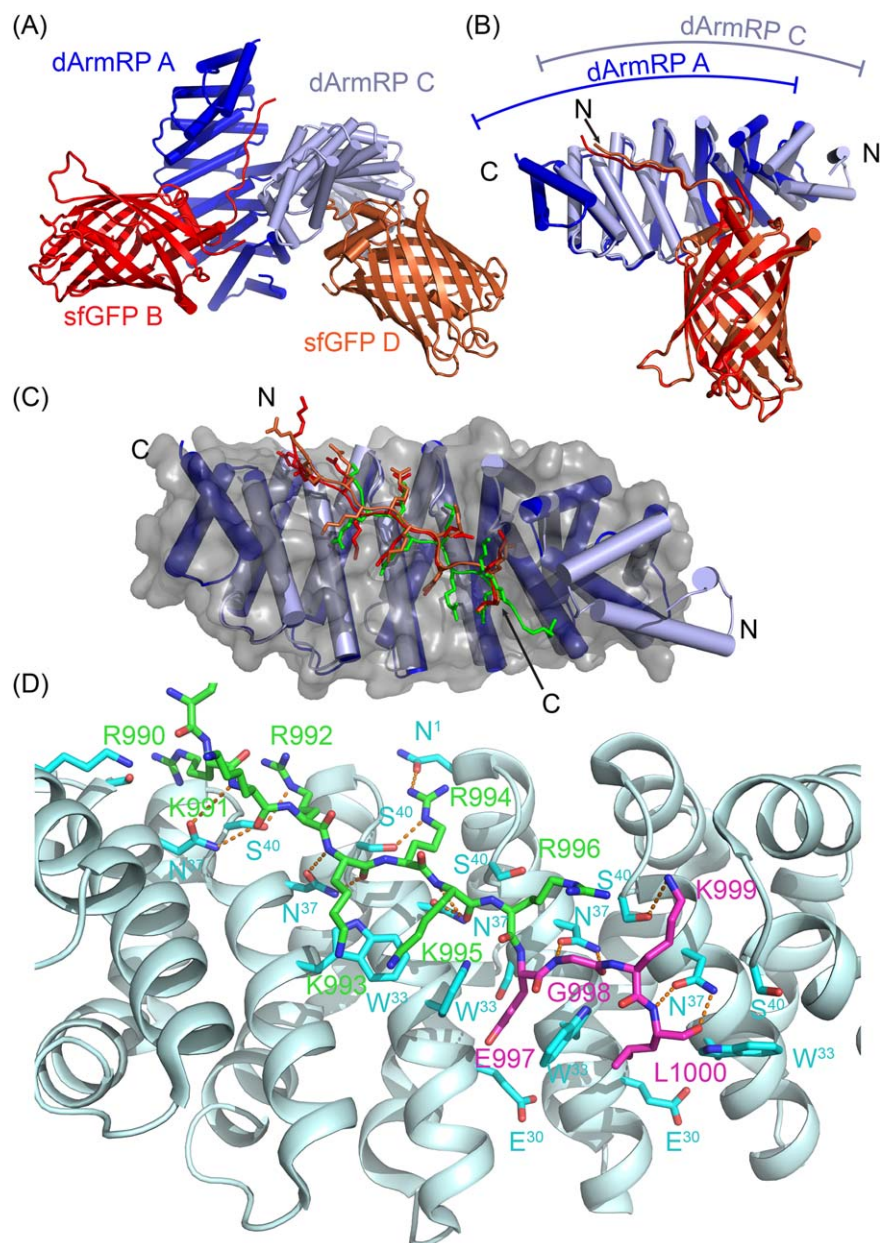


Figure 4. Analysis of the $Y_{III}M_5A_{II}:(KR)_4$ _sfGFP structure. (A) Overview of the asymmetric unit (AU). All protein chains are labeled in the respective color with protein type and chain ID. (B) Superposition of the two complexes reveals a register shift between dArmRP A and dArmRP C. N- and C-termini are indicated. (C) Superposition of experimental complexes with complex of the unfused peptide, $Y_{III}M_5A_{II}:(KR)_5$ (PDB-ID: 5AEI). $Y_{III}M_5A_{II}$ is shown as transparent surface and unfused $(KR)_5$ peptide in green, complex AB in blue and red, and complex CD in light blue and orange. (E) Detailed view of the peptide:dArmRP interaction in complex AB. The peptide is colored green for the $(KR)_4$ portion and magenta for the linker between $(KR)_4$ and sfGFP, sfGFP is omitted. dArmRP is shown as cartoon with side chains of interacting residues highlighted as cyan sticks. Hydrogen bonds are shown in orange. All peptide residues are labeled in the respective color. Highlighted dArmRP residues are labeled according to numbering within one armadillo repeat.

affinities of $(RR)_4$ and $(KK)_4$ peptides with $Y_{III}M_4A_{II}$ and $Y_{III}M_5A_{II}$ were determined. This ensures that all pockets are occupied by the same residues, either lysine or arginine, and the comparison with the affinities of $(KR)_4$ peptides allows to quantify the effect of either residue bound in the opposite pocket. $(RR)_4$ is bound more tightly by both dArmRP than $(KR)_4$ while the affinity of $(KK)_4$ is much weaker than $(KR)_4$ (Table II). Hence, arginines bind better than lysines in both

pockets. A flipped peptide where arginines take the usual place of lysines and vice versa will still possess a reasonable affinity, probably in between the affinities of $(KR)_n$ and $(KK)_n$ peptides.

Discussion

Here we present structural evidence that dArmRPs are able to bind unstructured peptide stretches connected to folded protein domains, and not just synthetic peptides.

The general binding mode of $Y_{III}M_xA_{II}$ and $(KR)_n$ was thus confirmed, meaning that the interaction is indeed mediated by the designated binding groove and the expected topology of the complex is adopted. This topology is defined by the directions of the N- and C-termini, which run in opposite directions for dArmRP and peptide (antiparallel complex). The conservation of this topology is crucial for the prediction of binding epitopes based on the primary peptide sequence and hence for the modular concept of the dArmRP technology.¹¹

Structures obtained by X-ray crystallography are usually taken as gold standard to which all other data have to be compared. Here we have observed, however, differences between the interactions of the fused peptide with the dArmRP even within the same unit cell, emphasizing the need for a more critical evaluation of such interactions. In some cases relevant intermolecular contacts can be disfavored because of the strong forces generated when accommodating packing contacts.²¹ The observed structure(s) are in an energy minimum accommodating these crystal packing interaction, the cognate peptide–protein interactions, and any other interaction within and between molecules.

The interaction between dArmRPs and $(KR)_n$ has been described in-depth by in-solution methods and a structure of the complex with the unfused peptide, $Y_{III}M_5A_{II}:(KR)_5$, has been determined that is only marginally influenced by crystal packing.¹⁶ In contrast, both structures described here exhibit some striking differences from the expected behavior in solution. These differences could be attributed to dynamics of the interaction between dArmRP and peptide. However, some changes in occupancy are likely caused by the formation of the crystal lattice. Notably the $Y_{III}M^r_6A_{II}:pD_(KR)_5$ structure consists of eight dArmRPs and only four $pD_(KR)_5$ chains. On the atomic level the $(KR)_5$ peptides interact with only one dArmRP chain. Hence, in the crystal we find only half of the dArmRPs occupied by a peptide, forming four 1:1 complexes and four unoccupied dArmRPs, while in solution, homogenous 1:1 complexes are observed by SEC-MALS [Fig. 1(A), Table I]. It is stunning that despite the high affinity of this interaction half of the peptide binding sites on dArmRPs remain unoccupied in the crystal structure. Probably, a crystal lattice with a homogenous distribution of 1:1 complexes could not be obtained. Also the four dArmRP molecules that form a direct interaction with $pD_(KR)_5$ show different arrangements of the pD domain and dArmRP relative to each other. This influences the observed peptide–dArmRP interaction; the interactions are more similar to the unfused peptide complex $Y_{III}M_5A_{II}:(KR)_5$ when the C-terminus of $pD_(KR)_5$ is positioned closer to the peptide binding groove of the dArmRP [Fig. 3(D)] and thus fewer spatial restrictions are imposed on the peptide. This shows that the crystal

lattice has a profound influence on the observed interaction in the experimental structures.

$Y_{III}M^r_6A_{II}:pD_(KR)_5$ was the first structure determined of an M-type repeat in complex with a peptide. In the M-type repeat two point mutations were introduced that were thought to increase the affinity to $(KR)_n$ peptides. However, this was not achieved, K_{ds} are even slightly weaker compared to the M-type repeats (Table II), probably due to a less optimal binding geometry.

The structure of $Y_{III}M_5A_{II}:(KR)_4_sfGFP$ shows that the dArmRP also interacts with the linker between $(KR)_4$ and sfGFP, which has the sequence EGKL. This would seem to put the specificity towards the $(KR)_n$ peptides in question. We believe, however, that this observed binding is only possible at low pH, where a protonated E997 can be accommodated in an arginine pocket. The conformation where the linker is bound has probably a reduced flexibility compared to a conformation where the linker is not stabilized by the dArmRP, and thus might benefit crystal formation. In solution, no measurable affinity was observed between $(AV)_n$ peptides that were fused with exactly the same linker sequence to sfGFP,¹⁶ pointing to no relevant affinity of the linker region itself to $Y_{III}M_5A_{II}$. Two different binding registers are found within one structure. We already proposed this behavior as a possible cause for the high affinity between $Y_{III}M_xA_{II}$ and $(KR)_n$, because it would increase the configurational entropy of the complexes.¹⁶

It was previously reported that in a dataset of protein–ligand complex structures around one third of the entries exhibit influences on the binding mode of ligands by crystal contacts.²² In the structures described here, a flexible peptide is bound, and residual binding affinity will be preserved even if it is only partially bound or flipped. Therefore, these structures might be especially susceptible to influences on their observed interaction mode. The fusion to a bulky structured domain adds to this problem because energetically it might be more valuable to accommodate this domain in the crystal lattice rather than establishing the full-length interaction of the flexible epitope.²¹ Finally, the high salt concentrations ($Y_{III}M^r_6A_{II}:pD_(KR)_5$) or low pH values ($Y_{III}M_5A_{II}:(KR)_4_sfGFP$) of the crystallization conditions reduce the affinities of $(KR)_n$ peptides to dArmRP.¹⁶

In summary, even though the two structures described here confirm the general binding mode of dArmRP for target peptides in the context of folded protein domains, we show how packing of crystals or crystallization conditions can elicit deviations from the behavior in solution. Hence, we want to stress that interaction studies by protein crystallization should be corroborated by suitable in-solution experiments.

Methods

Cloning

Cloning of the dArmRP genes has been described previously.¹³ For SEC-MALS and crystallization, they were subcloned with BamHI and HindIII restriction enzymes (FastDigest enzymes, Fermentas) into the vector pQE30LIC_3C, which contains a MRGSHis₆-tag cleavable by 3C-protease.¹⁶ For fluorescence anisotropy experiments, dArmRP genes were cloned into the expression vector pQIq containing an uncleavable N-terminal MRGSHis₆ tag.²³ Cloning of (KR)₄-sfGFP and pD₅(KR)₅ has been described previously.¹⁶ For crystallization they were also subcloned into the vector pQE30LIC_3C. Plasmids were extracted from overnight cultures of single *E. coli* XL1 Blue colonies grown on LB agar plates (100 µg mL⁻¹ ampicillin as selection marker) and sequenced. *E. coli* XL1 Blue or *E. coli* BL21 (DE3) cells were retransformed with plasmids with correct sequences, and glycerol stocks (20% glycerol, stored at -80°C) were made.

Protein expression and purification

Protein expression was carried out in 1 l of 2xYT medium (containing 100 µg L⁻¹ ampicillin and 0.5% glucose). Media were inoculated from 25 mL overnight culture, themselves inoculated from glycerol stocks, and grown at 37°C to an OD₆₀₀ of 0.7. Expression was induced by adding 750 µM of isopropyl-β-D-thiogalactopyranoside (IPTG) and left for 5 h at 37°C. Cells were harvested by centrifugation (5000 *g*, 5 min), resuspended in 25 mL TBS_W (50 mM Tris pH 8.0, 400 mM NaCl, 20 mM imidazole) and frozen until further usage. Resuspended cells were thawed on ice and lysed by sonication and passage through a French press system, cell debris was removed by centrifugation (25,000*g*, 20 min). Crude extracts were applied to Ni-NTA superflow resin columns (3 mL, Qiagen). Columns were washed with 30 column volumes of TBS_W and proteins were eluted with TBS_E (TBS_W with 300 mM imidazole). For crystallization and SEC-MALS experiments dArmRP and human rhino virus 3C-protease (2% w/w) were mixed and the reaction mixture was dialyzed against 50 mM Tris pH 7.4 and 300 mM NaCl to remove MRGSHis₆-tags; uncleaved proteins and 3C-protease were removed by reverse IMAC chromatography. For crystallization, dArmRP and pD₅(KR)₅ or (KR)₄-sfGFP were mixed with an 1.5 molar excess of pD₅(KR)₅ or (KR)₄-sfGFP. Complexes were isolated by SEC on an Äkta Explorer chromatography system using a HiLoad 26/60 Superdex 200 pg column and 10 mM Tris, pH 7.4 with 100 mM NaCl as running buffer. Prior to crystallization, complexes were concentrated (Amicon Ultra Centrifugal Filters, Merck Millipore) to 20 mg mL⁻¹.

Size exclusion chromatography multi-angle light scattering

An Agilent LC1100 chromatography system (Agilent Technologies) equipped with an Optilab rEX refractometer (Wyatt Technology) and a miniDAWN three-angle light-scattering detector (Wyatt Technology) was used for the experiments. All samples (50 µL, 1 mg mL⁻¹) were injected into a Superdex 200 10/30 column (24 mL, GE Healthcare) using PBS, Tris-buffered saline (TBS, 50 mM Tris/HCl pH 7.4, 150 mM NaCl) or TBS with 200 mM CaCl₂ as running buffer (Table I). Analysis was conducted with ASTRA software (version 6.0.1.10; Wyatt Technology).

Affinity determination

Black nonbinding 96-well plates (Greiner) and PBS with 0.03% BSA was used for all assays. Peptide-sfGFP was kept at a constant concentration and titrated with increasing concentrations of dArmRP with four replicates for each dArmRP concentration. A Safire II plate reader (Tecan) was used to measure fluorescence anisotropy. Data were averaged and the anisotropy value of the highest dArmRP dilution was subtracted from all other values. Fitting to a simple 1:1 binding model (Eq. 1) was done in Graphpad Prism software

$$Y(K_D, L_t, R_t) = \frac{m \left(-K_D - L_t - R_t + \sqrt{(K_D + L_t + R_t)^2 - 4L_t R_t} \right)}{-2R_t} \quad (\text{Eq. 1})$$

where Y is the actual amplitude, m is the amplitude of maximal anisotropy increase, K_D is the dissociation constant, L_t is the total ligand concentration (dArmRP), and R_t is the total receptor concentration (peptide-sfGFP).

Crystallization and structure determination

Sparse-matrix screens (Hampton Research and Molecular Dimensions) in 96-well Corning plates (Corning Incorporated) at 4°C were used in a sitting-drop vapor diffusion set-up to identify initial crystallization conditions. For each condition the reservoir solution was mixed in three ratios with protein solution (1:1, 1:2, and 2:1). Table III summarizes the crystallization conditions as well as data collection and refinement statistics. Crystals of Y_{III}M₅A_{II}:(KR)₄-sfGFP were flash-frozen (liquid N₂) in mother liquor supplemented with 20% glycerol, whereas crystals of Y_{III}M₆A_{II}:pD₅(KR)₅ were frozen directly in mother liquor. Data of Y_{III}M₅A_{II}:(KR)₄-sfGFP was collected on beam line X06DA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) using a Pilatus detector system (Dectris). Data of Y_{III}M₆A_{II}:pD₅(KR)₅ was collected at beamline P14 at Petra III (Deutsches Elektronen Synchrotron, Hamburg) on a Pilatus detector system

(Dectris). Data were processed using programs XDS, XSCALE and XDSCONV.²⁴

PHASER²⁵ was used for molecular replacement to obtain initial phases. Search models for $Y_{III}M_5A_{II}:(KR)_4$ -sfGFP were poly-alanine-models based on PDB IDs 5AEI (dArmRP, chain A) and 1GFL (GFP, chain A). For $Y_{III}M_6A_{II}:pD_{(KR)_5}$ a model based on structure 5AEI but with six internal repeats was used as search model for dArmRPs. This allowed us to obtain initial phases and subsequent manual placement of PDB-ID: 1TCZ¹⁸ into the additional density. Refinement was done using programs REFMAC5,^{26,27} BUSTER,²⁸ and Phenix-Refine,^{29,30} followed by model building in COOT.^{31,32} Five percent of data were used to calculate the R_{free} value.

Acknowledgments

The authors express their gratitude to Céline Stutz-Ducommun and Beat Blattmann from the UZH Protein Crystallization Center for help with crystallization experiments and the staff from beamlines X06DA at the Swiss Light Source (Paul Scherrer Institut, Würenlingen, Switzerland) and P14 at Petra III (Deutsches Elektronen Synchrotron, Hamburg, Germany) for technical support.

References

1. Peifer M, Berg S, Reynolds AB (1994) A repeating amino-acid motif shared by proteins with diverse cellular roles. *Cell* 76:789–791.
2. Huber AH, Nelson WJ, Weis WI (1997) Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* 90:871–882.
3. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J (1998) Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell* 94:193–204.
4. Daniels DL, Spink KE, Weis WI (2001) beta-catenin: molecular plasticity and drug design. *Trends Biochem Sci* 26:672–678.
5. Conti E, Kuriyan J (2000) Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. *Structure* 8:329–338.
6. Fontes MRM, Teh T, Jans D, Brinkworth RI, Kobe B (2003) Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. *J Biol Chem* 278:27981–27987.
7. Kobe B (1999) Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. *Nat Struct Biol* 6:388–397.
8. Marfori M, Mynott A, Ellis JJ, Mehdi AM, Saunders NF, Curmi PM, Forwood JK, Boden M, Kobe B (2011) Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim Biophys Acta* 1813:1562–1577.
9. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H (2009) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J Biol Chem* 284:478–485.
10. Takeda AAS, de Barros AC, Chang CW, Kobe B, Fontes MRM (2011) Structural basis of importin-alpha-mediated nuclear transport for Ku70 and Ku80. *J Mol Biol* 412:226–234.
11. Parmeggiani F, Pellarin R, Larsen AP, Varadamsetty G, Stumpp MT, Zerbe O, Caflisch A, Plückthun A (2008) Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J Mol Biol* 376:1282–1304.
12. Alfarano P, Varadamsetty G, Ewald C, Parmeggiani F, Pellarin R, Zerbe O, Plückthun A, Caflisch A (2012) Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. *Protein Sci* 21:1298–1314.
13. Madhurantakam C, Varadamsetty G, Grütter MG, Plückthun A, Mittl PR (2012) Structure-based optimization of designed Armadillo-repeat proteins. *Protein Sci* 21:1015–1028.
14. Reichen C, Hansen S, Plückthun A (2014) Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. *J Struct Biol* 185:147–162.
15. Reichen C, Madhurantakam C, Plückthun A, Mittl PRE (2014) Crystal structures of designed armadillo repeat proteins: implications of construct design and crystallization conditions on overall structure. *Protein Sci* 23:1572–1583.
16. Hansen S, Tremmel D, Madhurantakam C, Reichen C, Mittl PR, Plückthun A (2016) Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity. *J Am Chem Soc* 138:3526–3532.
17. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24:79–88.
18. Chang CS, Plückthun A, Wlodawer A (2004) Crystal structure of a truncated version of the phage lambda protein gpD. *Proteins* 57:866–868.
19. Yang F, Forrer P, Dauter Z, Conway JF, Cheng NQ, Cerritelli ME, Steven AC, Plückthun A, Wlodawer A (2000) Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat Struct Biol* 7:230–237.
20. Reichen C, Madhurantakam C, Hansen S, Grütter MG, Plückthun A, Mittl PRE (2016) Structures of designed armadillo-repeat proteins show propagation of inter-repeat interface effects. *Acta Cryst D* 72:168–175.
21. Krissinel E (2010) Crystal contacts as nature's docking solutions. *J Comput Chem* 31:133–143.
22. Sondergaard CR, Garrett AE, Carstensen T, Pollastri G, Nielsen JE (2009) Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. *J Med Chem* 52:5673–5684.
23. Simon M, Zangemeister-Wittke U, Plückthun A (2012) Facile double-functionalization of designed ankyrin repeat proteins using click and thiol chemistries. *Bioconjugate Chem* 23:279–286.
24. Kabsch W (2010) XDS. *Acta Cryst D* 66:125–132.
25. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* 40:658–674.
26. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Cryst D* 55:247–255.
27. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA

- (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst D*67:355–367.
28. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vornrhein C, To W (2016) BUSTER version 2.10.2. Cambridge, United Kingdom: Global Phasing Ltd.
29. Afonine PV, Grosse-Kunstleve RW, Chen VB, Headd JJ, Moriarty NW, Richardson JS, Richardson DC, Urzhumtsev A, Zwart PH, Adams PD (2010) phenix.-model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. *J Appl Cryst* 43:669–676.
30. Afonine PV, Mustyakimov M, Grosse-Kunstleve RW, Moriarty NW, Langan P, Adams PD (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Cryst D*66:1153–1163.
31. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst D*60:2126–2132.
32. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Cryst D*66:486–501.