Acta Cryst D STRUCTURAL BIOLOGY

ISSN 2059-7983

Structures of designed armadillo-repeat proteins show propagation of inter-repeat interface effects

Christian Reichen,‡ Chaithanya Madhurantakam,§ Simon Hansen, Markus G. Grütter, Andreas Plückthun* and Peer R. E. Mittl*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. *Correspondence e-mail: plueckthun@bioc.uzh.ch, mittl@bioc.uzh.ch

Received 18 September 2015
Accepted 1 December 2015

Edited by Z. S. Derewenda, University of Virginia, USA

‡ Present address: Molecular Partners AG, Wagistrasse 14, 8952 Zürich-Schlieren, Switzerland.
§ Present address: Department of Biotechnology, TERI University, 10 Institutional Area, Vasant Kunj, New Delhi 110 070, India.

Keywords: armadillo repeat; protein engineering; calcium binding; peptide binding; solenoid protein.

PDB references: His-$Y_{III}M_5A_{II}$, 4v3o; His-$Y_{III}M_4A_{II}$, 4v3q; $Y_{III}M_4A_{II}$, 4v3r

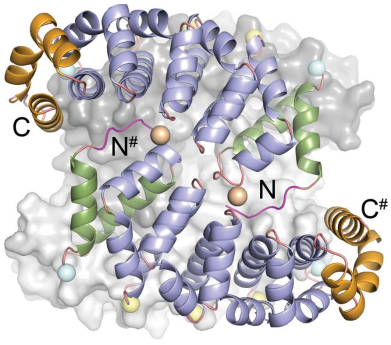
Supporting information: this article has supporting information at journals.iucr.org/d

The armadillo repeat serves as a scaffold for the development of modular peptide-recognition modules. In order to develop such a system, three crystal structures of designed armadillo-repeat proteins with third-generation N-caps (Y_{III}-type), four or five internal repeats (M-type) and second-generation C-caps (A_{II}-type) were determined at 1.8 Å (His-$Y_{III}M_4A_{II}$), 2.0 Å (His-$Y_{III}M_5A_{II}$) and 1.95 Å ($Y_{III}M_5A_{II}$) resolution and compared with those of variants with third-generation C-caps. All constructs are full consensus designs in which the internal repeats have exactly the same sequence, and hence identical conformations of the internal repeats are expected. The N-cap and internal repeats M_1 to M_3 are indeed extremely similar, but the comparison reveals structural differences in internal repeats M_4 and M_5 and the C-cap. These differences are caused by long-range effects of the C-cap, contacting molecules in the crystal, and the intrinsic design of the repeat. Unfortunately, the rigid-body movement of the C-terminal part impairs the regular arrangement of internal repeats that forms the putative peptide-binding site. The second-generation C-cap improves the packing of buried residues and thereby the stability of the protein. These considerations are useful for future improvements of an armadillo-repeat-based peptide-recognition system.

1. Introduction

For the design of artificial peptide-binding modules, scaffolds with modular architectures are highly suitable. In particular, the armadillo repeat reveals structural properties that facilitate the design of peptide-binding modules on a rational basis (Andrade *et al.*, 2001; Kajander *et al.*, 2006; Boersma & Plückthun, 2011; Reichen, Hansen *et al.*, 2014). In natural armadillo-repeat proteins such as importin-α and β-catenin, each repeat comprises three α-helices that are assembled in a triangular spiral staircase arrangement. All repeats are fused into a single protein with an elongated hydrophobic core (Figs. 1*a* and 1*b*). They recognize their target peptides in extended β-sheet conformations with very regular binding topologies. The main chain of the peptide is bound in an antiparallel direction by conserved asparagine residues on the concave side of the armadillo-repeat protein (Huber *et al.*, 1997; Conti *et al.*, 1998; Kobe, 1999; Fontes *et al.*, 2003). Differences exist in side-chain preferences because the importin-α and β-catenin subfamilies recognize peptides with positively and negatively charged side chains, respectively (Conti & Kuriyan, 2000; Ishiyama *et al.*, 2010; Poy *et al.*, 2001).

It is the goal of this protein-engineering project to develop a stable full-consensus armadillo-repeat scaffold. Internal repeats with identical sequences are characteristic of full-consensus designs. Later, the internal repeats will be functionalized to recognize different amino-acid side chains.



OPEN ACCESS

The modularity of the design, which is imposed by the repetitive architecture, should enable us to generate artificial peptide-binding proteins with properties that are precisely tailored according to the length and sequence of the target peptide (Parmeggiani *et al.*, 2008; Reichen, Hansen *et al.*, 2014). Binding proteins with sequence-specific recognition properties for unstructured peptides should be of great interest in research and development because peptide–protein interactions represent 15–40% of all cellular interactions (Petsalaki *et al.*, 2009). Here, many protein–protein interaction scaffolds are unsuitable because they recognize targets based on surface-complementarity properties and thus require a folded counterpart. Conversely, many recognition modules used in intracellular signalling recognize only very short sequences and thus have very low affinity (Pawson & Nash, 2003). Indeed, specific peptide–protein interaction strategies are required to cope with the intrinsic flexibility of unstructured peptides (London *et al.*, 2010).

The first designed armadillo-repeat proteins (dArmRPs) were constructed using a consensus design approach based on 133 and 110 sequences from the importin-α and β-catenin subfamilies, respectively, in combination with structure-aided modifications of the hydrophobic core (Parmeggiani *et al.*, 2008). They possess the overall composition $Y_zM_nA_z$, where Y, M and A represent the N-terminal, internal and C-terminal repeats, respectively. The subscripts denote the generation (version) count (z) and the number of internal repeats (n) in roman and arabic numbers, respectively. Since structure-based techniques are vital for this design approach, several structures of proteins from the $Y_{II}M_nA_{II}$ and $Y_{III}M_nA_{III}$ series have been determined. Initial crystal structures of dArmRPs with second-generation N- and C-caps revealed domain-swapped N-caps, suggesting that the Y_{II}-type N-cap was unstable in solution. To improve the thermodynamic stability of the caps, nine and six mutations were inserted in the N- and C-caps, respectively. These modifications had complementary effects on the thermodynamic stability of the proteins. Introduction of the third-generation N-cap (Y_{III}-type) increased the melting temperature by 4.5°C, but the modifications in the C-cap (A_{III}-type) decreased it by 5.5°C (Madhurantakam *et al.*, 2012). The thermodynamic stabilities of dArmRPs that have so far been designed in this project have been summarized in Reichen, Hansen *et al.* (2014).

Although the initial crystal structures of His-$Y_{III}M_3A_{III}$ and His-$Y_{III}M_3A_{II}$ revealed monomeric proteins (Reichen, Madhurantakam *et al.*, 2014), later studies on $Y_{III}M_5A_{III}$ (third-generation N-cap and C-cap) without an N-terminal His tag revealed domain-swapped N-caps and C-caps in the presence of calcium ions. However, domain swapping of $Y_{III}M_5A_{III}$ was not observed either in the absence of calcium ions or in the presence of the His tag because the His tag prevented the unfolding of the N-cap by binding to the neighbouring His-$Y_{III}M_5A_{III}$ molecule (Reichen, Madhurantakam *et al.*, 2014). To investigate the impact of the cap design on the structural parameters of dArmRPs, particularly in the absence of the His tag, we investigated the crystal structures of the more stable dArmRPs with third-generation N-caps and second-generation C-caps.

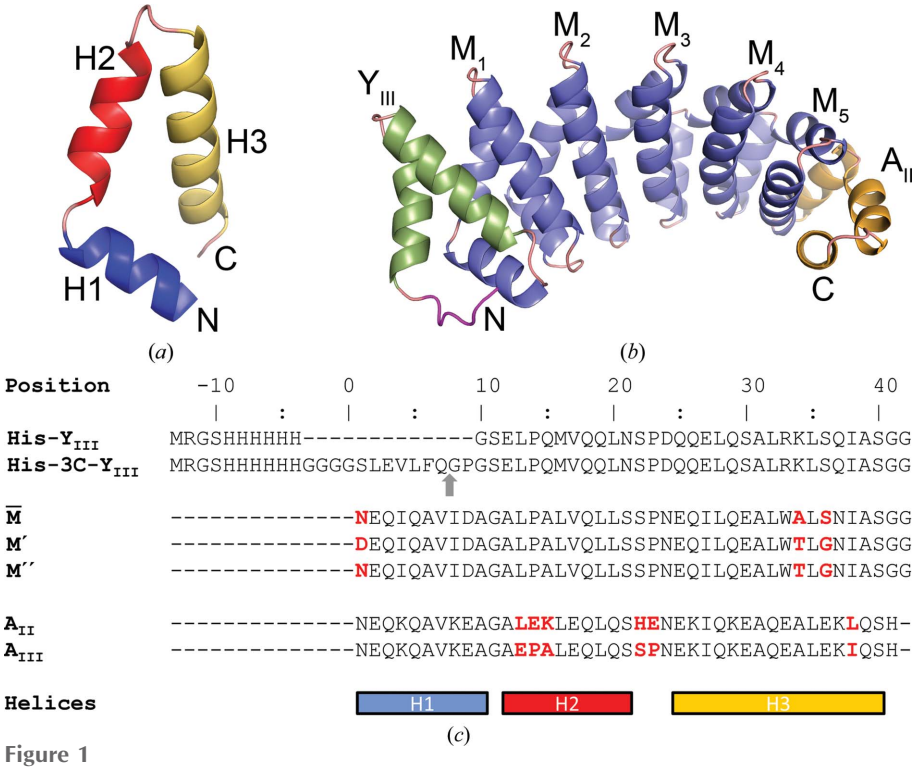


Figure 1
(*a*) The triangular spiral staircase arrangement of helices indicative of the armadillo repeat. (*b*) Ribbon diagram of His-$Y_{III}M_5A_{II}$. The His_6 tag, Y_{III}-type, M-type and A_{II}-type repeats are shown in magenta, green, blue and orange, respectively. (*c*) Sequence alignment of N-caps with and without a 3C protease cleavage site (the scissile bond is indicated by a grey arrow), internal repeats and C-caps. Residues distinguishing different repeat versions are highlighted in red.

2. Materials and methods

2.1. Cloning, protein expression and purification

dArmRPs with cleavable and non-cleavable N-terminal His_6 tags have been expressed and purified as described by Reichen, Madhurantakam *et al.* (2014) with the following modifications: vectors pPank and p148_3C were used for the expression of proteins with and without a cleavable His_6 tag, respectively. The initial designs had noncleavable His_6 tags, but in order to facilitate the elimination of the purification tag, a 3C protease cleavage site was inserted between the His_6 tag and the N-terminus of the N-cap. The amino-acid sequences of the internal and capping repeats are depicted in Fig. 1(*c*).

Table 1
Data and refinement statistics.

Values in parentheses are for the highest resolution shell.

Structure	His-$Y_{III}M_4A_{II}$	His-$Y_{III}M_5A_{II}$	$Y_{III}M_5A_{II}$
PDB code	4v3q	4v3o	4v3r
Data statistics			
Crystallization condition	25% PEG 2000 MME, 0.2 *M* calcium acetate, 0.1 *M* sodium acetate pH 5.5	15% PEG 4000, 0.2 *M* calcium acetate, 0.1 *M* sodium acetate pH 5.5	30% PEG 4000, 0.2 *M* magnesium chloride, 0.1 *M* Tris–HCl pH 8.5
Space group	$P3_2$	$P4_1$	$I4$
No. of molecules in asymmetric unit	4	4	2
Unit-cell parameters			
$a = b$ (Å)	96.50	102.59	129.91
c (Å)	96.34	111.11	70.20
$\alpha = \beta$ (°)	90	90	90
γ (°)	120	90	90
Resolution (Å)	1.80 (1.91–1.80)	2.00 (2.11–2.00)	1.95 (2.06–1.95)
R_{merge} (%)	9.1 (88.6)	10.0 (75.0)	8.8 (47.6)
No. of observations	744192 (120009)	601165 (75390)	107908 (15424)
No. of unique reflections	93024 (15191)	76669 (10657)	41831 (6089)
$\langle I/\sigma(I)\rangle$	12.6 (2.3)	12.2 (2.6)	7.5 (2.0)
Completeness (%)	100 (100)	94.3 (94.3)	98.1 (98.2)
Refinement statistics			
Resolution (Å)	96.34–1.80	111.11–2.00	91.86–1.95
R_{cryst} (%)	18.9	16.8	17.3
R_{free} (%)	23.6	22.4	22.9
B factors			
Wilson *B* (Å^2)	27.0	28.7	21.4
Mean *B* value (Å^2)	35.5	35.2	23.2
R.m.s.d. from ideal values			
Bond lengths (Å)	0.018	0.017	0.017
Bond angles (°)	1.83	1.71	1.72
Total No. of atoms			
Protein	7487	8618	4243
Water	654	767	419
Metal ions	16	12	3
Ligands	2	1	0
Ramachandran plot			
Favoured (%)	98.81	99.02	100.00
Allowed (%)	1.19	0.98	0.00
Outliers (%)	0.00	0.00	0.00

The proteins comprise third-generation N-caps, second-generation C-caps and four or five internal repeats. All three constructs are full-consensus designs, with internal repeats derived from the $\overline{M}$-type internal repeat described in Alfarano *et al.* (2012). His-$Y_{III}M_4A_{II}$ and $Y_{III}M_5A_{II}$ contain M′-type internal repeats, whereas His-$Y_{III}M_5A_{II}$ contains the M″-type. In the M″-type the aspartic acid at position 1, which was introduced to mimic a potential arginine-binding pocket, was mutated back to the consensus asparagine residue (for all sequences, see Fig. 1*c*). To improve readability, we refer to M-type internal repeats throughout the text.

2.2. Crystallization and structure determination

A Phoenix crystallization robot (Art Robbins Instruments) was used to set up sitting-drop vapour-diffusion experiments in 96-well Corning plates (Corning, New York, USA). Initial crystallization conditions were identified by sparse-matrix screens from Hampton Research (California) and Molecular Dimensions (Suffolk, England), and were later refined by grid screens in which the pH and the precipitant concentrations were varied simultaneously. To confirm the expected peptide-binding site, $(KR)_5$ peptide was added to $Y_{III}M_5A_{II}$ in a 1.5:1 molar ratio prior to crystallization. $(KR)_5$ peptide was used for this experiment because the designed molecular surface of $Y_{III}M_5A_{II}$ resembled the most conserved importin-α peptide-binding site, which recognizes with its core repeats (major and minor binding sites) positive dipeptide motifs composed of lysine and arginine residues. The rationale for this experiment is discussed in Reichen, Hansen *et al.* (2014). Protein solutions were mixed with reservoir solutions in 1:1, 1:2 or 2:1 ratios (200–300 nl final volume) and the mixtures were equilibrated against 50 µl reservoir solution at 4°C. Reservoir conditions are summarized in Table 1. After washing, the crystals in reservoir solutions supplemented with glycerol were flash-cooled in liquid nitrogen.

Data were collected on beamlines X06SA and X06DA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) using a Pilatus detector (Dectris, Baden, Switzerland) and a wavelength of 1.0 Å. Diffraction data were processed using *MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 2006). Structures were solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007) together with the following search models. For His-$Y_{III}M_4A_{II}$ we used the structure of $Y_{III}M_3A_{II}$ (PDB entry 4db6; Madhurantakam *et al.*, 2012). The refined His-$Y_{III}M_4A_{II}$ structure was then used to solve the His-$Y_{III}M_5A_{II}$ and finally the $Y_{III}M_5A_{II}$ structures. The structures were refined using *PHENIX* (Adams *et al.*, 2010) and *REFMAC*5 (Murshudov *et al.*, 2011). For manual model building we used the program *Coot* (Emsley & Cowtan, 2004). The decrease in R_{free} suggested the use of different refinement strategies for His-$Y_{III}M_4A_{II}$ and His-$Y_{III}M_5A_{II}$. His-$Y_{III}M_4A_{II}$ was refined without NCS restraints, whereas tight NCS restraints between chains *A*/*B* and *C*/*D* were applied for the refinement of His-$Y_{III}M_5A_{II}$. Figures were prepared using *PyMOL* (DeLano, 2002). Metal ions were placed manually into strong difference electron-density peaks, taking into account the coordination geometry and the composition of the crystallization buffer. Calcium ions were validated by inspecting the anomalous difference map calculated with phases from the final structure. Water molecules were placed into well defined difference

electron-density peaks at hydrogen-bond distance from the protein. No $(KR)_5$ peptide was identified in the final electron-density map of $Y_{III}M_5A_{II}$. Side-chain conformations were assigned according to the rotamer library of Dunbrack & Cohen (1997) as implemented in *Coot*.

3. Results and discussion

3.1. Structures of His-$Y_{III}M_4A_{II}$ and His-$Y_{III}M_5A_{II}$

The crystal structures of His-$Y_{III}M_4A_{II}$ and His-$Y_{III}M_5A_{II}$ were refined at 1.8 and 2.0 Å resolution, respectively. In both cases the asymmetric units contain tetramers with 222 point symmetry and very similar topologies. The quaternary structures of His-$Y_{III}M_4A_{II}$ and His-$Y_{III}M_5A_{II}$ are governed by calcium ions that connect neighbouring chains in a zipper-like manner and the His_6 tag that binds to the supposed peptide-binding site, albeit in different orientations (see below).

The His-$Y_{III}M_4A_{II}$ tetramer contains 16 calcium ions. Five calcium ions connect two His-$Y_{III}M_4A_{II}$ chains in an anti-parallel orientation (Fig. 2*a*). Considering the large size of this interface (average interface area of 1163 Å^2) there are relatively few direct hydrogen bonds, and most interactions are made *via* calcium ions in the loops between helices H2 and H3. The coordination number of each calcium ion in His-$Y_{III}M_4A_{II}$ is seven, which agrees very well with the statistical analysis of calcium-coordination geometry in protein and small-molecule complexes. Typically, the coordination number of calcium varies between six and eight, with an average length for coordination bonds of between 2.35 and 2.45 Å (Katz *et al.*, 1996). In His-$Y_{III}M_4A_{II}$ the coordination geometry of calcium differs among ions that are bound to internal or capping repeats.

Ca^{2+} ions that bind to internal repeats are contacted by Pro^{23} O and Glu^{25} OE1 from two symmetry-related chains (superscripts indicate the position in the repeat as indicated in Fig. 1*c*) and three water molecules (Fig. 2*b*). Here, Glu^{25} contributes one coordination bond (Glu^{25} OE1–Ca distance 2.5 Å). In contrast, calcium ions that bind between an internal repeat and the N-cap are contacted by two water molecules, two O atoms from Glu^{25} (Glu^{25} OE1–Ca distance of 2.5 Å and Glu^{25} OE2–Ca distance of 3.0 Å), Gln^{25} OE1 and Pro^{23} O (Fig. 2*c*). Thus, the replacement of glutamic acid at position 25 by glutamine in the N-cap displaces one water molecule and allows Glu^{25} to serve as a bidentate ligand. This observation agrees well with previous data on the statistics of calcium binding, in which it was shown that bidentate binding of carboxylate groups to calcium is particularly prevalent if the coordination number is greater than six (Katz *et al.*, 1996). In contrast to many natural calcium-binding sites, where all coordination bonds are approximately equal in length, the His-$Y_{III}M_4A_{II}$ calcium-binding sites are distorted. In His-$Y_{III}M_4A_{II}$ the axial calcium–ligand distances are shorter than the equatorial distances (axial distances 2.1–2.2 Å; equatorial distances 2.4–3.0 Å) and the Glu^{25} OE2–Ca bonding distances differ significantly from the average coordination bond length. The second coordination bond of Glu^{25} is longer, because the carboxylate group is rotated away from the Ca^{2+} ion. In contrast to natural calcium-binding sites that have evolved over time, the His-$Y_{III}M_4A_{II}$ calcium-binding sites are distorted because they are artificial and are therefore less perfect. Besides these zipper-like Ca^{2+} ions bound to the N-termini of H3 helices, four well defined calcium ions additionally bind close to the twofold axes. These Ca^{2+} ions also show pentagonal-bipyramidal coordination spheres involving the Ser^{40} carbonyl O atom, the Glu^{2} side chain

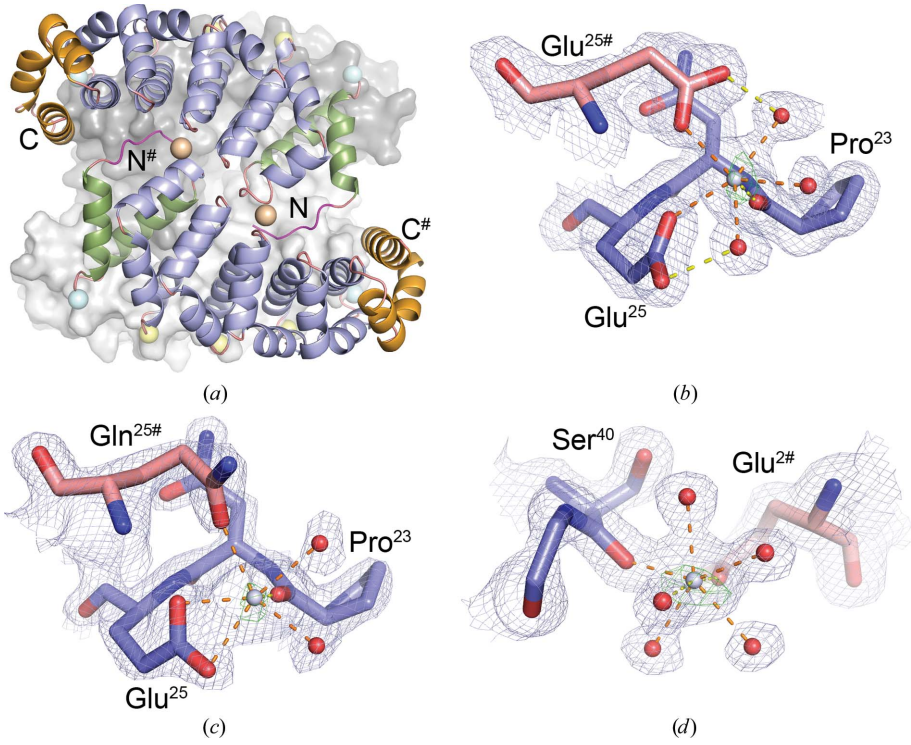


Figure 2
(*a*) The subunits of the His-$Y_{III}M_4A_{II}$ tetramer are connected *via* calcium ions. Two chains are sketched as ribbons and coloured as described in Fig. 1(*b*). Two chains are shown as grey surfaces. Calcium ions are indicated as spheres. Calcium ions binding only to internal repeats are in yellow, those involving the N-cap in light blue and those at the twofold axis in salmon. (*b*) Calcium-binding site between internal repeats viewed along the axial direction (from the direction of $Pro^{23\#}$ O, which was omitted for clarity). Residues from different chains are shown as sticks with blue and salmon C atoms. Calcium ions and water molecules are depicted as grey and red spheres with reduced atomic radii, respectively. Polar interactions in the pentagonal plane involving the calcium ion are shown as dashed lines in orange. Additional interactions are in yellow. The $2F_o - F_c$ and anomalous difference electron-density maps are contoured at 1.3σ (light blue) and 4σ (green), respectively. (*c*) Calcium-binding site involving the N-cap. (*d*) Calcium-binding site at the twofold axis. Colour coding is as described for (*b*).

and five water molecules (Fig. 2*d*). Furthermore, there are two weakly occupied calcium-binding sites involved in crystal contacts.

The His-$Y_{III}M_4A_{II}$ tetramer is further stabilized by interactions between the N-terminal His_6 tag and the supposed peptide-binding site. This contact is formed by His6, which interacts with Glu^{30} and Trp^{33} (Glu156 and Trp159) from the third internal repeat, and His8, which interacts with Trp^{33} (Trp201) from the fourth internal repeat and Glu^{33} (Glu243) from the C-cap (Fig. 3*a*). Besides the salt bridges between histidine and glutamic acid side chains, the aromatic stacking interaction between His6 and Trp^{33} might contribute significant binding energy because the spatial orientation of side chains seen here is frequently found in protein structures (cluster 4 of His–Trp interactions in the atlas of protein side-chain interactions; Singh & Thornton, 1992). Since all four chains of His-$Y_{III}M_4A_{II}$ are very similar (r.m.s.d. of 0.28 Å for residues 14–246) these interactions are equivalent in all four subunits of the crystallographic tetramer.

In contrast to this, the crystallographic tetramer of His-$Y_{III}M_5A_{II}$ is less symmetric. Here, chains *A*/*B* and *C*/*D* are pairwise identical (r.m.s.d. of 0.05 Å), whereas an r.m.s.d. of 0.85 Å for the comparison between pairs (*e.g.* chain *A* with *D*) suggests substantial differences. Furthermore, His-$Y_{III}M_5A_{II}$ chains *A*/*B* are more similar to His-$Y_{III}M_4A_{II}$ (r.m.s.d. of 0.72 Å for the superposition of residues 14–210 on the equivalent residues from His-$Y_{III}M_4A_{II}$) than chains *C*/*D* (r.m.s.d. of 1.17 Å). These differences are caused by different contacts within the tetramer. In chains *C*/*D* of His-$Y_{III}M_5A_{II}$ the side chain of Glu198 interacts with His8 from chain *D*/*C* (Fig. 3*b*), whereas in chains *A*/*B* the side chain of Glu198 intercalates between internal repeats 3 and 4 and forms a hydrogen bond to the side chain of Gln68 from chains *B*/*A* (similar to the interaction shown in Fig. 3*a* for His-$Y_{III}M_4A_{II}$). As a consequence of this asymmetry, two calcium ions close to the twofold axis, which are present in all four chains of His-$Y_{III}M_4A_{II}$ (Fig. 2*d*), are only present in His-$Y_{III}M_5A_{II}$ chains *A*/*B* and are absent from chains C/*D*.

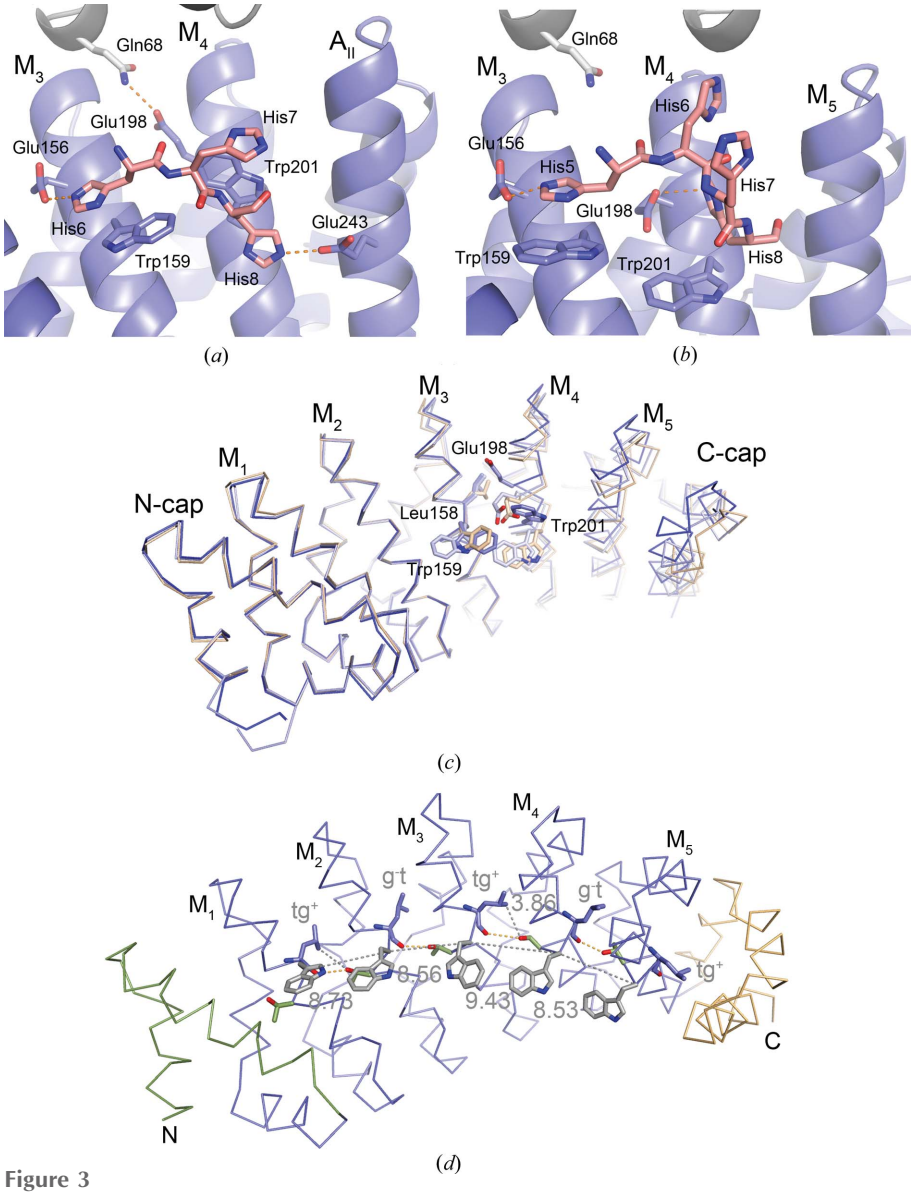


Figure 3
Interface between internal repeats M_3 and M_4 in chain *C* of His-$Y_{III}M_4A_{II}$ (*a*) and His-$Y_{III}M_5A_{II}$ (*b*). The dArmRPs are shown in blue and grey and the His_6 tag with salmon C atoms. (*c*) Superposition based on the N-cap and internal repeats M_1–M_3 of His-$Y_{III}M_5A_{II}$ chain *A* (dark blue), His-$Y_{III}M_5A_{II}$ chain *C* (light blue) and $Y_{III}M_5A_{II}$ (orange). Residues at the M_3–M_4 interface are labelled. (*d*) C^{α} trace of $Y_{III}M_5A_{II}$ coloured in green (N-cap), blue (internal repeats) and orange (C-cap). The Leu^{32}, Trp^{33} and Thr^{34} side chains are shown as sticks in blue, grey and green, respectively. Hydrogen bonds and general distances are shown as orange and grey dotted lines, respectively. Distances and conformations of Leu^{32} side chains are indicated (tg^+, *trans*/*gauche*$^+$; g^-t, *gauche*$^-$/*trans*).

3.2. Structure of $Y_{III}M_5A_{II}$ without His tag

The structure of $Y_{III}M_5A_{II}$ without His tag was determined in the absence of calcium ions and refined at 1.95 Å resolution. This structure is most similar to chains *C*/*D* of His-$Y_{III}M_5A_{II}$ (r.m.s.d.s of 1.14 and 0.60 Å for C^{α} atoms of residues 14–288 of chains *A*/*B* and *C*/*D*, respectively). These differences are a consequence of a rigid-body movement of the C-terminal repeats (internal repeats M_4 and M_5 and the C-cap). A superposition of $Y_{III}M_5A_{II}$ on His-$Y_{III}M_5A_{II}$ based on the N-cap and internal repeats M_1–M_3 (residues 14–168) reveals that this part is very similar in all chains. However, in this superposition the C-terminal repeats of

$Y_{III}M_5A_{II}$ match nicely with the C-terminal repeats of His-$Y_{III}M_5A_{II}$ chains *C*/*D*, but they are shifted towards M_3 in chains *A*/*B* (1.4 Å shift of Trp201 CA towards Leu158 CA). This movement can be described as an 8° rotation around an axis that runs parallel to the stacking direction of the C-terminal part and is probably a consequence of different side-chain conformations of Leu158, Trp159, Glu198 and Trp201 at the interface between M_3 and M_4 (Fig. 3*c*). The structures of His-$Y_{III}M_4A_{II}$ and $Y_{III}M_5A_{II}$ represent extreme cases that are most different. In His-$Y_{III}M_5A_{II}$ these differences are combined into a single structure. His-$Y_{III}M_5A_{II}$ chains *A*/*B* and *C*/*D* represent the conformations seen in His-$Y_{III}M_4A_{II}$ (all chains) and $Y_{III}M_5A_{II}$ (all chains), respectively. Similar structural plasticity has been observed previously for the comparison of β-catenin crystallized in two different crystal forms. For β-catenin the C-terminal repeats were rotated 11.5° around an axis that runs approximately parallel to the axis of the superhelix (Huber *et al.*, 1997).

Thus, dArmRPs with second-generation C-caps and third-generation N-caps possess substantial flexibility, particularly for the side chains of Glu^{30}, Leu^{32} and Trp^{33} (equivalent to Glu156, Leu158 and Trp159 in repeat M_3 and Glu198, Leu200 and Trp201 in repeat M_4). Experimental structural data for importin-α in complex with nuclear localization sequence (NLS) peptides (Conti *et al.*, 1998) and modelling studies on dArmRPs–peptide complexes (Reichen, Hansen *et al.*, 2014) indicate that the superhelix parameters and the conformations of Glu^{30} and Trp^{33}, which also participate in binding the His_6 tag as outlined above, are important structural features for proper binding of the target peptide. In a first approximation, the curvature of the peptide-binding site can be described by the distances of C^{α} atoms at position 33. In the major NLS peptide-binding site of importin-α (PDB entry 1bk6; Conti *et al.*, 1998) the distance between C^{α} atoms of adjacent Trp^{33} residues (*e.g.* Trp153, Trp195 and Trp237 in repeats 1–3) varies between 8.6 and 8.8 Å. In $Y_{III}M_5A_{II}$ the average distance between these atoms is 8.82 ± 0.39 Å. However, in $Y_{III}M_5A_{II}$ the spread between Trp^{33} C^{α}-atom distances is extremely large, with the largest distance observed between repeats M_3 and M_4 (the distances between Trp159 CA and Trp201 CA are 9.42 Å in chain *A* and 9.43 Å in chain *B*). This distance is probably too large for binding the target peptide in the desired conformation and this mismatch is located almost at the centre of the putative peptide-binding site. It is possible that this mismatch is responsible for the fact that the $(KR)_5$ peptide was not observed in the electron-density map, although it was present during crystallization. Interestingly, the rigid-body movement of the C-terminal part as seen in His-$Y_{III}M_4A_{II}$ (all chains) and His-$Y_{III}M_5A_{II}$ (chains *C*/*D*) brings this value to the other extreme. Here, the distance of Trp^{33} C^{α} atoms between repeats M_3 and M_4 is 8.14 ± 0.06 Å, which might be too short for proper binding.

Although $Y_{III}M_5A_{II}$ is considered to be a full consensus design regarding the sequence of internal repeats, the internal repeats are not identical in terms of structure. These differences can be exerted either by different lattice contacts (Figs. 3*a* and 3*b*) or by improper design, which prevents the internal repeats from obtaining a unique conformation throughout the protein. Different distances between adjacent repeats are probably the result of both effects. In particular, the side-chain conformations of buried residues in the hydrophobic core, such as Ile^{27}, Leu^{32}, Thr^{34}, Gly^{36} and Ile^{38}, mediate the contacts between adjacent repeats. In the structure of $Y_{III}M_5A_{II}$ the side-chain conformations of Thr^{34}, Ile^{38} and of course Gly^{36} are invariant. The side chain of Thr^{34} cross-links internal repeats by forming hydrogen bonds to the main-chain carbonyl groups of Leu^{32} and Glu^{30} from adjacent repeats. The side chain of Ile^{27} adopts mainly *gauche*$^-$/*trans* conformations, whereas the side chain of Leu^{32} alternates between *trans*/*gauche*$^+$ and *gauche*$^-$/*trans* (Fig. 3*d*).

This alternation suggests that a uniform conformation of Leu^{32} is impossible. In the interface between M_3 and M_4 of $Y_{III}M_5A_{II}$, where we observe the largest distance between Trp^{33} C^{α} atoms, Leu158 CD1 (Leu^{32} in M_3) and Thr202 OG1 (Thr^{34} in M_4) are at van der Waals distances (3.86 and 3.97 Å in chains *A* and *B*) because the Leu158 side chain adopts a *trans*/*gauche*$^+$ conformation. Therefore, steric hindrance between Leu158 and Thr202 might be responsible for increasing the distance between Trp^{33} C^{α} atoms and for the failure to obtain a dArmRP–peptide complex structure. To adopt a Trp^{33} C^{α} distance which is similar to the values seen in the major binding site of importin-α, Thr202 OG1 would have to move closer to Leu158, but this approach would require a *gauche*$^-$/*trans* conformation of the Leu158 side chain.

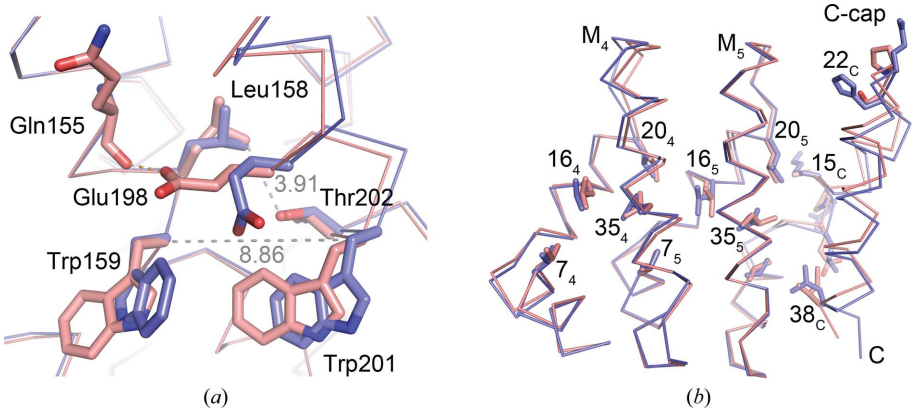


Figure 4
Superposition of $Y_{III}M_5A_{III}$ (third-generation C-cap; PDB entry 4plq; salmon) on $Y_{III}M_5A_{II}$ (second-generation C-cap; blue). (*a*) Residues at the M_3–M_4 interface. General distances and hydrogen bonds are shown as grey and orange dotted lines, respectively. Distance values refer to $Y_{III}M_5A_{III}$. The superposition is based on all C^{α} atoms from M_3. (*b*) Residues at the M_5–C-cap interface. Numbers refer to positions in the repeat (Fig. 1*c*), with subscripts indicating the internal repeat number or the C-cap. Side chains of all residues that differ between Y_{II} and Y_{III} and some residues from the hydrophobic core are shown in stick representation. The superposition is based on all C^{α} atoms from M_5.

Of course, surface-exposed side chains (such as Trp^{33} and Glu^{30}) also adopt different rotamers, but it can be assumed that these differences affect inter-repeat distances to a minor extent because the environments of surface-exposed side chains are usually less densely packed than the environments of buried side chains. However, some side-chain conformations of buried and surface-exposed residues are coupled. For example, the conformation of Trp^{33} is linked to the conformation of Leu^{32} in the preceding repeat. In repeats M_1 and M_3 Leu^{32} adopts *trans*/*gauche*$^+$ conformations and Trp^{33} in repeats M_2 and M_4 is *trans*/+90°, whereas in repeats M_2 and M_4 Leu^{32} is *gauche*$^-$/*trans* and Trp^{33} adopts *trans*/−105° conformations in repeats M_3 and M_5 (Fig. 3*d*). Only Trp243 in chain *B* deviates from this general observation.

3.3. Comparison of dArmRPs with second-generation and third-generation C-caps

The crystal structures of $Y_{III}M_5A_{III}$ with and without a His_6 tag and third-generation C-caps have been published recently (Reichen, Madhurantakam *et al.*, 2014). $Y_{III}M_5A_{III}$ without a His_6 tag but crystallized in the presence of calcium revealed domain-swapped N- and C-caps. Since $Y_{III}M_5A_{II}$ without a His_6 tag and a second-generation C-cap did not crystallize in the presence of calcium, it remains unclear whether the redesign of the C-cap was responsible for calcium-induced domain swapping.

Interestingly, $Y_{III}M_5A_{III}$ also shows an extended distance between Trp^{33} C^{α} atoms of internal repeats M_3 and M_4 (distance between Trp159 CA and Trp201 CA of 8.86 Å), a short distance between Thr202 OG1 and Leu158 CD2 of 3.91 Å and no electron density for the $(KR)_5$ peptide, although it was present during crystallization (Reichen, Madhurantakam *et al.*, 2014). On the other hand, Leu158 shows the *gauche*$^-$/*trans* side-chain conformation, which is *trans*/*gauche*$^+$ in $Y_{III}M_5A_{II}$, probably because Glu198 forms an additional hydrogen bond to Gln155 O (Fig. 4*a*).

For dArmRPs with three internal repeats it was shown that the redesign of the C-cap (from A_{II} to A_{III}) decreases the melting temperature by 5.5°C (Madhurantakam *et al.*, 2012), and a domain-swapped C-cap was observed for $Y_{III}M_5A_{III}$ (Reichen, Madhurantakam *et al.*, 2014). Both observations suggest that $Y_{III}M_5A_{III}$ is less stable than $Y_{III}M_5A_{II}$. A superposition of $Y_{III}M_5A_{III}$ (PDB entry 4plq) and $Y_{III}M_5A_{II}$ based on the last internal repeat suggests that this destabilization might be owing to subtle rearrangements in the hydrophobic core between internal repeats M_4 and M_5 and the C-cap. Three out of six mutations that were introduced at the C-cap are solvent-exposed and do not seem to have a significant effect on the structure. However, Lys^{15}→Ala, His^{22}→Ser and Leu^{38}→Ile mutations cause a gentle rearrangement of the C-cap (Fig. 4*b*). This rearrangement has implications for the packing of side chains in the hydrophobic core. In the more stable $Y_{III}M_5A_{II}$ structure the side chains of Leu^{16}, Leu^{20} and Val^7 adopt a uniform distribution of side-chain rotamers in all repeats. Val^7 adopts a *trans* conformation. Leu^{16} and Leu^{20} are always *gauche*$^-$/*trans*. In $Y_{III}M_5A_{III}$ this crystal-like arrangement is perturbed by the C-cap. In $Y_{III}M_5A_{III}$ the side chains of Leu^{16}, Leu^{20} and Val^7 adopt the same conformations as in $Y_{III}M_5A_{II}$ only in the N-terminal part, whereas in the C-terminal part their conformations are clearly different. For Leu^{32} the situation is inverted. In $Y_{III}M_5A_{III}$ the rotamer distribution of Leu^{32} is uniform, whereas in $Y_{III}M_5A_{II}$ alternating Leu^{32} conformations are observed (Fig. 3*d*). Uniform distributions of rotamers are frequently observed in polypeptides with very high thermodynamic stabilities, such as amyloid fibrils (Nelson *et al.*, 2005) and β-helix proteins (Schulz & Ficner, 2011). Therefore, it can be assumed that the uniform distribution of side-chain rotamers is related to the stability of dArmRPs and *vice versa*. On the other hand, the deterioration of uniformity, as caused by the third-generation C-cap, is linked to destabilization of the protein.

In conclusion, this detailed investigation of the different versions of dArmRPs has shown that small differences in packing between repeats, notably between internal repeats and the caps, can make the protein susceptible to perturbations caused by crystal contacts and ions used in crystallization, indicating a lack of rigidity. This leads to a surprising long-range effect of changes in the C-cap and helps to explain the astonishing observation that a full-consensus design does not necessarily generate a unique repeat conformation. Although the internal repeats are chemically absolutely identical, their conformations lack uniformity. The current analysis suggests that future improvements of an armadillo-repeat-based peptide-recognition system will have to take three considerations into account. (i) In particular, the deletion of the His tag seems to be crucial for liberating the presumed peptide-binding site. (ii) The second-generation C-cap presented here seems to be superior to the third-generation C-cap, which was initially believed to be more advanced. (iii) The choice of amino acids at the inter-repeat interface, particularly at positions 27, 32 and 34, should be reconsidered because the side chains at these positions show substantial conformational heterogeneity.

Acknowledgements

Beat Blatmann and Celine Stutz at the high-throughput crystallization centre and the staff of beamlines X06SA and X06DA at the Swiss Light Source are acknowledged for their skillful technical support. This work was financially supported by a Swiss National Science foundation grant to AP and PREM (Sinergia S-41105-06-01). SH is the recipient of Forschungskredit from the University of Zurich, grant No. FK-13-028.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Alfarano, P., Varadamsetty, G., Ewald, C., Parmeggiani, F., Pellarin, R., Zerbe, O., Plückthun, A. & Caflisch, A. (2012). *Protein Sci.* **21**, 1298–1314.
Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001). *J. Struct. Biol.* **134**, 117–131.
Boersma, Y. L. & Plückthun, A. (2011). *Curr. Opin. Biotechnol.* **22**, 849–857.
Conti, E. & Kuriyan, J. (2000). *Structure*, **8**, 329–338.

Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. (1998). *Cell*, **94**, 193–204.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Dunbrack, R. L. Jr & Cohen, F. E. (1997). *Protein Sci.* **6**, 1661–1681.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fontes, M. R., Teh, T., Toth, G., John, A., Pavo, I., Jans, D. A. & Kobe, B. (2003). *Biochem. J.* **375**, 339–349.
Huber, A. H., Nelson, W. J. & Weis, W. I. (1997). *Cell*, **90**, 871–882.
Ishiyama, N., Lee, S.-H., Liu, S., Li, G.-Y., Smith, M. J., Reichardt, L. F. & Ikura, M. (2010). *Cell*, **141**, 117–128.
Kajander, T., Cortajarena, A. L. & Regan, L. (2006). *Methods Mol. Biol.* **340**, 151–170.
Katz, A. K., Glusker, J. P., Beebe, S. A. & Bock, C. W. (1996). *J. Am. Chem. Soc.* **118**, 5752–5763.
Kobe, B. (1999). *Nature Struct. Biol.* **6**, 388–397.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
London, N., Movshovitz-Attias, D. & Schueler-Furman, O. (2010). *Structure*, **18**, 188–199.
Madhurantakam, C., Varadamsetty, G., Grütter, M. G., Plückthun, A. & Mittl, P. R. (2012). *Protein Sci.* **21**, 1015–1028.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R. & Eisenberg, D. (2005). *Nature (London)*, **435**, 773–778.
Parmeggiani, F., Pellarin, R., Larsen, A. P., Varadamsetty, G., Stumpp, M. T., Zerbe, O., Caflisch, A. & Plückthun, A. (2008). *J. Mol. Biol.* **376**, 1282–1304.
Pawson, T. & Nash, P. (2003). *Science*, **300**, 445–452.
Petsalaki, E., Stark, A., García-Urdiales, E. & Russell, R. B. (2009). *PLoS Comput. Biol.* **5**, e1000335.
Poy, F., Lepourcelet, M., Shivdasani, R. A. & Eck, M. J. (2001). *Nature Struct. Biol.* **8**, 1053–1057.
Reichen, C., Hansen, S. & Plückthun, A. (2014). *J. Struct. Biol.* **185**, 147–162.
Reichen, C., Madhurantakam, C., Plückthun, A. & Mittl, P. R. (2014). *Protein Sci.* **23**, 1572–1583.
Schulz, E. C. & Ficner, R. (2011). *Curr. Opin. Struct. Biol.* **21**, 232–239.
Singh, J. & Thornton, J. M. (1992). *Atlas of Protein Side-Chain Interactions*. Oxford: IRL Press.