doi:10.1016/j.jmb.2010.09.023 J. Mol. Biol. (2010) 404, 381–391

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Structural Determinants for Improved Stability of Designed Ankyrin Repeat Proteins with a Redesigned C-Capping Module

Michaela A. Kramer, Svava K. Wetzel, Andreas Plückthun, Peer R. E. Mittl* and Markus G. Grütter

Department of Biochemistry, University of Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland

Received 28 May 2010; received in revised form 6 September 2010; accepted 8 September 2010
Available online 17 September 2010

Edited by I. Wilson

Keywords:
protein engineering;
ankyrin repeat;
protein recognition;
crystal structure;
thermodynamic stability

Designed ankyrin repeat proteins (DARPins) that specifically bind to almost any target can be obtained by ribosome display or phage display from combinatorial libraries. Although DARPins are already very stable molecules, molecular dynamics simulations, equilibrium denaturation experiments, structural studies, and recent NMR experiments suggested that the unfolding of the original C-terminal capping repeat (C-cap), taken from a natural ankyrin repeat protein, limits the stability of the initial DARPin design. Several point mutations had been introduced to optimize the C-cap and were shown to indeed further increase the stability of DARPins. We now determined crystal structures of DARPins with one or three full-consensus internal repeats (NI_1C or NI_3C) between an N-terminal capping repeat and mutants of the C-cap. An NI_1C mutant, in which the C-cap was only extended by three additional helix-forming residues, showed no structural change but reduced *B*-factors in the C-cap. An NI_3C C-cap mutant carrying five additional mutations in the interface to the preceding repeat, previously designed by using the consensus sequence as a guide, showed a rigid-body movement of the C-cap towards the internal repeat. This movement results in an increased buried surface area and a superior surface complementarity and explains the improved stability in equilibrium unfolding, compared to the original C-cap. A C-cap mutant with three additional mutations introducing suitably spaced charged residues did not show formation of salt bridges, explaining why its stability was not increased further. These structural studies underline the importance of repeat coupling for stability and help in the further design of this protein family.

Introduction

Repeat proteins are among the most abundant classes of naturally occurring protein–protein interaction modules. They are found in all kingdoms of life and constitute rather simple overall architectures, consisting of structurally identical repeats that are stacked together to form an elongated non-globular protein domain. Repeat proteins serve as unique templates for protein engineering experiments and protein folding studies, because their

*Corresponding author. E-mail address: mittl@bioc.uzh.ch.

Abbreviations used: DARPin, designed ankyrin repeat protein; GdnHCl, guanidine hydrochloride; VdW, van der Waals; NCS, non-crystallographic symmetry; AU, asymmetric unit; C-cap, C-terminal capping repeat; N-cap, N-terminal capping repeat; AR, ankyrin repeat; MD, molecular dynamics; PDB, Protein Data Bank.

stability is dominated by intra-repeat interactions and interactions between neighboring repeats.[1,2]

Several classes of repeat proteins exist. Besides armadillo,[3] tetratricopeptide,[4] and leucine-rich repeat proteins,[5] ankyrin repeats (AR) have probably been most widely used as protein engineering templates. The AR consists of two antiparallel α-helices and a short β-turn.

Two approaches have been applied to design AR consensus proteins by protein engineering. While Mosavi *et al.*[6] reported full consensus proteins, Binz *et al.*[7] directly designed a library and only later reported a full consensus design.[8] The two full consensus designs differ in the amino acid sequence of the repeats (discussed in Ref. 8) and, most importantly, in the presence of capping repeats.

The original design of Mosavi *et al.*[6] without capping repeats yielded proteins that were only soluble at acidic pH, and the subsequent introduction of positive charges in the C-terminal repeat then allowed the protein to be soluble at neutral pH,[9] but it still had to be produced by refolding from inclusion bodies made in *Escherichia coli*. Unfortunately, this gain in solubility was accompanied by a significant loss in stability at pH 4, while at pH 7, the stability could not be compared because the initially designed protein was not soluble.

In contrast, the design of Binz *et al.*[7] had incorporated both N- and C-terminal capping repeats (N-cap and C-cap, respectively) with a polar and charged surface, and these caps were later shown to be crucial to allow the protein to fold to the native state in bacterial expression systems.[10] This efficient *in vivo* folding was a prerequisite for the generation of specific binders and subsequent biotechnological applications. From libraries of designed ankyrin repeat proteins (DARPins), in which in the internal repeats 6 of the 33 residues were randomized, binders to many targets were selected by either ribosome display[11] or phage display,[12] and crystal structures of isolated DARPins and DARPins in complex with various target proteins have been determined (see, e.g., Refs. 13–15). We denote these proteins obtained from the libraries as N*x*C, where N and C denote the N- and C-capping repeats and *x* denotes the number of randomized internal repeats, and we denote the full consensus proteins as NI_xC, where I denotes a full consensus internal repeat and the subscript *x* denotes their number.

DARPins are very stable molecules.[7,13] The stability of the full consensus NI_xC DARPins is even higher than the stability of DARPins with randomized internal repeats. At least for DARPins with up to six internal repeats, the stability increases with the number of internal repeats.[8] DARPins with more than two internal repeats cannot be denatured by boiling alone.[8,10] Depending on the ionic strength, the full-consensus NI_3C possesses melting temperatures, in the presence of 4 M guanidine hydrochloride (GdnHCl), between 74 and 80 °C.[8,16]

Table 1. Sequence and structural parameters of full-consensus DARPins

Repeat	Sequence[a]	T_m[b] (°C)	Buried surface[c] (Å²)	SC[c]	Hydrophobic contacts	Polar contacts[d]	Average *B*-factor[e] (Å²)
N-cap	--D**LGKKLLEAARA**GQ**DDEVRILMAN**GADVNAK : 20 : 30 : 40						2QYI: 23.0 Mutant 4: 21.9 Mutant 5: 31.7 ± 0.4 Mutant 6: 31.5 ± 2.9
Internal	DKDGYT**PLHLAARE**GH**LEIVEVLLKA**GADVNAK : 50 : 60 : 70 :						
C-cap							
original	DKFGKT**AFDISIDN**GN**EDLAEILQ**--------- : 150 : 160 :	60	625	0.705	20	2	34.1
Mut4	DKFGKT**AFDISIDN**GN**EDLAEILQ<u>KAA</u>**------ 80 : 90 : 100	64	654	0.711	14	4 (1)	22.6
Mut5	DKFGKT**<u>P</u>FD<u>L</u>AIDN**GN**ED<u>I</u>AE<u>V</u>LQ<u>KAA</u>**------ : 150 : 160 :	77	720 ± 6	0.801	26	4	33.1 ± 1.6
Mut6	DKFGKT**<u>P</u>FD<u>L</u>AI<u>RE</u>GH<u>ED</u>I<u>A</u>E<u>V</u>LQ<u>KAA</u>**------ : 150 : 160 :	77	723 ± 6	0.802	30	4	39.4 ± 3.5

[a]Residues participating in α-helices are indicated in bold characters. Mutated residues are underlined. Residue numbering refers to NI_1C for mutant 4 and to NI_3C for mutants 5 and 6 and NI_3C with the original C-cap (PDB ID: 2QYI).

[b]Melting temperature of NI_1C with the original C-cap and mutants, taken from Ref. 10.

[c]Buried surface area and surface complementarity (SC) between last internal repeat and C-cap. Errors represent standard deviations between values of NCS-related molecules.

[d]Number of H-bonds. The number of salt bridges is given in parentheses.

[e]Given are the average *B*-factors for residues 13 to 142 (NI_3 parts) for 2QYI, mutant 5, and mutant 6 and residues 13 to 76 (NI_1 part) for mutant 4.

Molecular dynamics (MD) calculations of two DARPins, E3_5 and E3_19,[17] and an investigation of full-consensus DARPins combining MD calculations with GdnHCl-induced unfolding experiments[10] had shown that the original C-cap possessed the lowest stability of all repeats. This original C-cap had been taken from the guanine–adenine binding protein β1.[7] Comparison of several DARPin crystal structures also revealed that the C-cap possesses the largest structural flexibilities.[13,16,18]

From the analysis of the MD calculations, Interlandi *et al.* suggested a total of 11 mutations to improve the stability of the C-cap and the packing of the C-cap to the last internal repeat (Table 1).[10] It had been found that in the simulation at 300 K, the C^{α} atoms of the last internal repeat fluctuate less than those of the C-cap. The new design had been inspired by making the C-cap more similar to the consensus sequence of the internal repeats. This includes an Ala-to-Pro substitution, which reintroduces the conserved Thr-Pro-Leu-His motif, as well as four mutations located at the interface to the last internal repeat, which should improve side-chain packing. Additionally, the C-terminus was extended by three residues (sequence Lys101-Ala102-Ala103) to elongate the C-terminal helix (all these changes are present in NI_1C_Mut5). This helix extension was also investigated by itself (NI_1C_Mut4). Finally, three charged residues were introduced in addition to those of NI_1C_Mut5, to allow for inter-repeat salt bridges and hydrogen bonds (to generate NI_1C_Mut6), and it had been observed that, indeed, this protein showed the lowest fluctuations in MD calculations.[10]

To investigate the structural consequences of those mutations at the atomic level, we determined the crystal structures of NI_3C_Mut5 and NI_3C_Mut6, as well as NI_1C_Mut4, and compared them to the crystal structure of the ancestor molecule NI_3C with the C-cap from the guanine–adenine binding protein β1.

Results

Structure of NI_1C_Mut4

The crystal structure of NI_1C_Mut4 was determined at 2.2 Å resolution. It folds into a compact

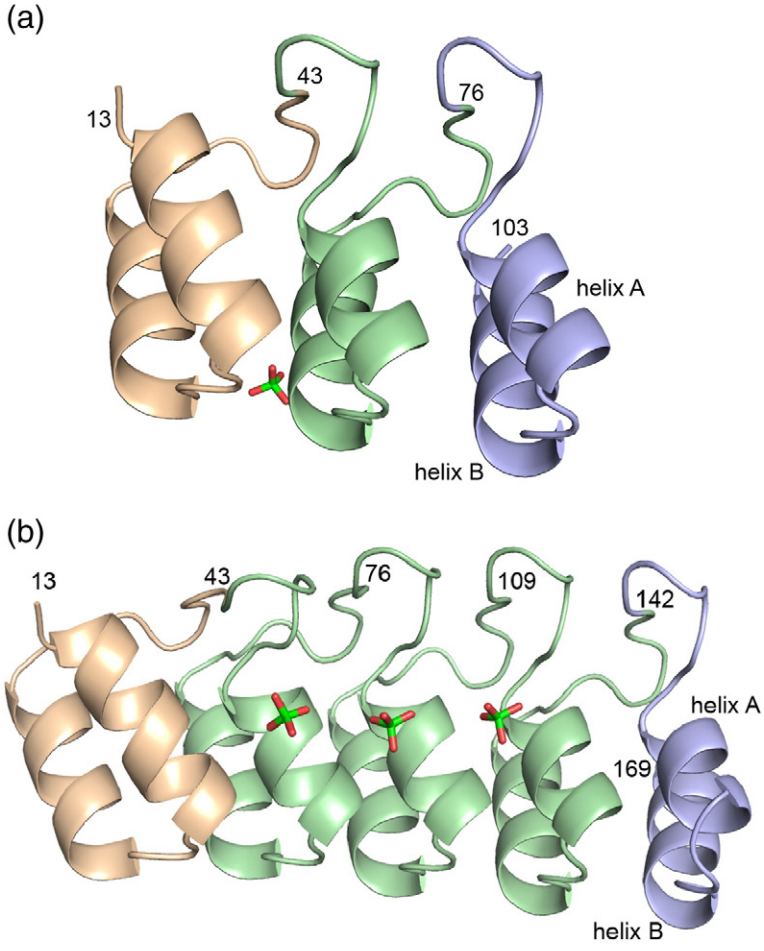

Fig. 1. Ribbon diagrams of NI_1C_Mut4 (a) and NI_3C_Mut6 (b). N-cap, internal repeats, and C-cap are colored salmon, green, and blue, respectively. The N-terminus and the last amino acid of every repeat are labeled. Bound sulfate ions are shown in a stick representation.

globular molecule with overall dimensions of 17 Å × 24 Å × 25 Å (Fig. 1). The structure reveals clear electron density for amino acids 13 to 103, which are subdivided into the N-cap (residues 13 to 43), the single internal repeat (residues 44 to 76), and the C-cap (residues 77 to 103). The first 12 residues corresponding to the N-terminal MRGSHis$_6$-tag are flexible and do not show up in the electron density. Although several crystal structures of DARPins with three internal repeats have been determined, the present structure of NI$_1$C_Mut4 serves as the prototype structure for a DARPin with just one internal repeat. The NI repeat pair (residues 13 to 76) and the IC repeat pair (residues 43 to 100) of NI$_1$C_Mut4 superimpose perfectly onto the corresponding N- and C-terminal repeat pairs of NI$_3$C with the original cap with rmsd of 0.47 Å (66 C$^\alpha$ atoms) and 0.46 Å (54 C$^\alpha$ atoms), respectively (Fig. 2). The N-, I-, and C-terminal repeats of NI$_1$C reveal average *B*-factors of 19.1 Å^2, 24.7 Å^2, and 22.6 Å^2, respectively. Thus, in contrast to previously determined structures of DARPins with three internal repeats, the C-cap of NI$_1$C_Mut4 shows a lower average *B*-factor than the internal repeat (see Discussion and Fig. 3).

The C-cap of NI$_1$C_Mut4 deviates from the original capping repeat of NI$_3$C only by a three-amino-acid extension of the C-terminal helix, the second helix (helix B) of the capping repeat. This extension appears to have decreased the *B*-factors of the C-cap, compared to the original C-cap. As has been anticipated during the design process, residues Lys101-Ala102-Ala103 adopt an α-helical conformation and participate in the hydrogen-bonding network of helix B. The side chain of Lys101 serves as a C-terminal helix cap but forms no direct hydrogen bonds with the C-terminal carboxylate of helix B. Instead, Lys101 forms a surface-exposed salt bridge with the side chain of Glu97. Furthermore, the C$^\alpha$ atom of Ala103 is located at van der Waals

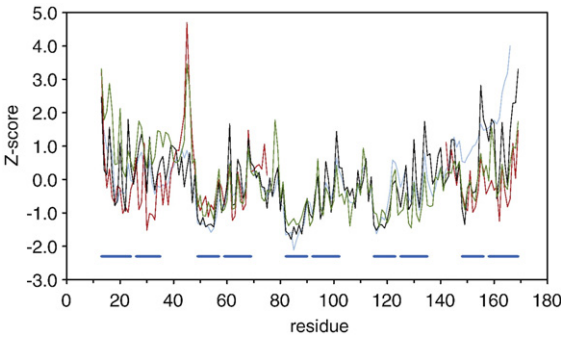

Fig. 3. Average *B*-factor plotted over the residue number for NI$_3$C with the original C-cap (blue), NI$_3$C_Mut5 (black), NI$_3$C_Mut6 (green), and NI$_1$C_Mut4 (red). Blue bars indicate α-helices. To put all average *B*-factors on the same scale, we expressed them as *Z*-scores, which are defined by the equation $Z = (\langle B_n \rangle - \langle B_{all} \rangle)/\sigma_B$ with $\langle B_n \rangle$ being the average *B*-factor for residue number *n*, $\langle B_{all} \rangle$ the average *B*-factor for all amino acids, and σ_B the standard deviation of *B*-factors.

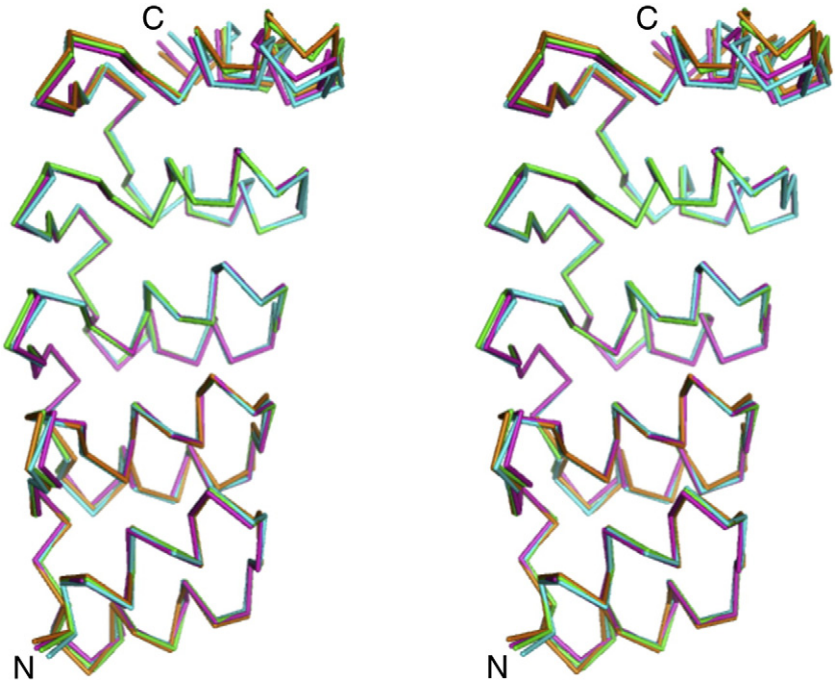

Fig. 2. Walleye stereo plot of the superposition of NI$_3$C with the original C-cap (green), NI$_3$C_Mut5 (blue), NI$_3$C_Mut6 (magenta), and NI$_1$C_Mut4 (orange).

(VdW) distance to the side chain of Val73 from the preceding internal repeat (Fig. 4a).

NI_1C_Mut4 shows a markedly different side-chain packing at the interface between internal and C-terminal repeats (Fig. 4b). In NI_1C_Mut4, the side chain of His52 forms a salt bridge with the side chain of Asp77, whereas in NI_3C, the equivalent His118 interacts with Thr115. The movement of His52 causes the rotation of the side chains of Ile86, Ser87, and Leu95, since they are all at VdW distance to one another. In NI_1C_Mut4, Ser87-OG forms a hydrogen bond with Ala83-O (3.04 Å). Since none of the rotated side chains are in direct contact with the Lys101-Ala102-Ala103 extension of helix B, the observed movements could be the consequence of different crystallization conditions. NI_1C_Mut4 was crystallized at pH 5.0 and in the presence of 30% methanol, whereas NI_3C with the original cap was crystallized at pH 8.5 and in the absence of methanol. Neither the extension of helix B nor the different side-chain conformations of His52, Ile86, Ser87, and Leu95 have a significant effect on the conformation of the capping repeat main chain. Thus, the spatial orientations of the C-caps relative to the last internal repeats are almost identical in the structures of NI_3C with the original C-cap and NI_1C_Mut4. Thus, the movement of His52 is not linked to any structural changes of the backbone (see Discussion).

Structures of NI_3C_Mut5 and NI_3C_Mut6

NI_3C mutants 5 and 6 (NI_3C_Mut5 and NI_3C_Mut6) crystallized with four and three molecules in the asymmetric units (AUs) and were refined at 2.1 and 1.8 Å resolution, respectively. The structures of the N-caps and the internal repeats of both mutants were almost identical with the corresponding repeats of NI_3C with the original C-cap [Protein Data Bank (PDB) ID: 2QYI] (Fig. 2). Superposition of C^{α} atoms of residues 13 to 76 from NI_3C[16] onto NI_3C_Mut5 and NI_3C_Mut6 revealed rmsds of 0.31±0.01 and 0.24±0.04 Å, respectively. However, larger structural changes are seen at the interfaces between the last internal repeats and the C-caps. Superposition of residues 110 to 166 (corresponding to the last internal repeat and the C-cap) revealed rmsds of 0.78±0.01 and 0.68 ±0.07 Å for NI_3C_Mut5 and NI_3C_Mut6, respectively. In both mutants, the C-caps have moved as rigid bodies towards the last internal repeats (shown for NI_3C_Mut5 in Fig. 5a and b). The largest movements are seen for residues 154 to 160, which form the loop between helices A and B of the C-caps. These residues have moved approximately 2 to 3 Å towards the last internal repeat.

This rigid-body movement of the C-cap is a consequence of the new interface design. NI_3C_Mut5 deviates from the NI_3C with the original C-cap by a three-amino-acid extension of helix B (as in NI_1C_Mut4, see above) plus five additional point

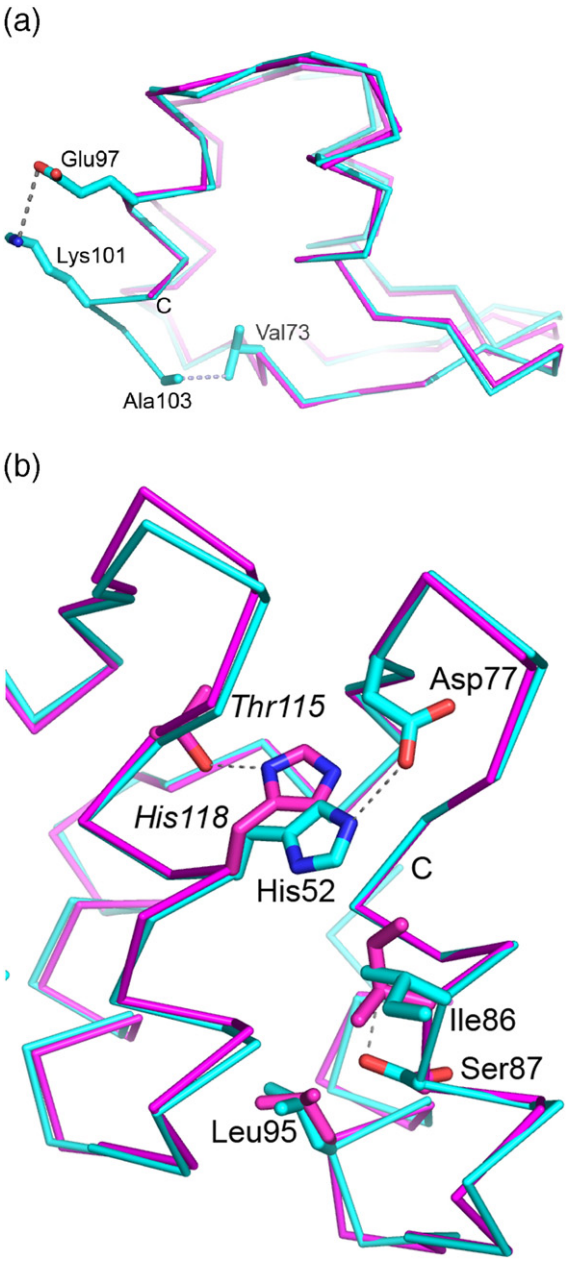

Fig. 4. Superposition of NI_1C_Mut4 (light blue) on NI_3C with the original C-cap (magenta, PDB ID: 2QYI). (a) Extension of the C-terminal helix B introduces a salt bridge between the side chains of Glu97 and Lys101. The C-terminus of NI_3C with the original C-cap is indicated. (b) Depicted are the rotation of the His52 side chain and rotation of side chains in the interface between the last internal repeat and the C-cap. Amino acid numbering refers to NI_1C with the exception of Thr115 and His118 (numbering from 2QYI), which are shown in italics. The C-terminus of NI_1C_Mut4 is indicated.

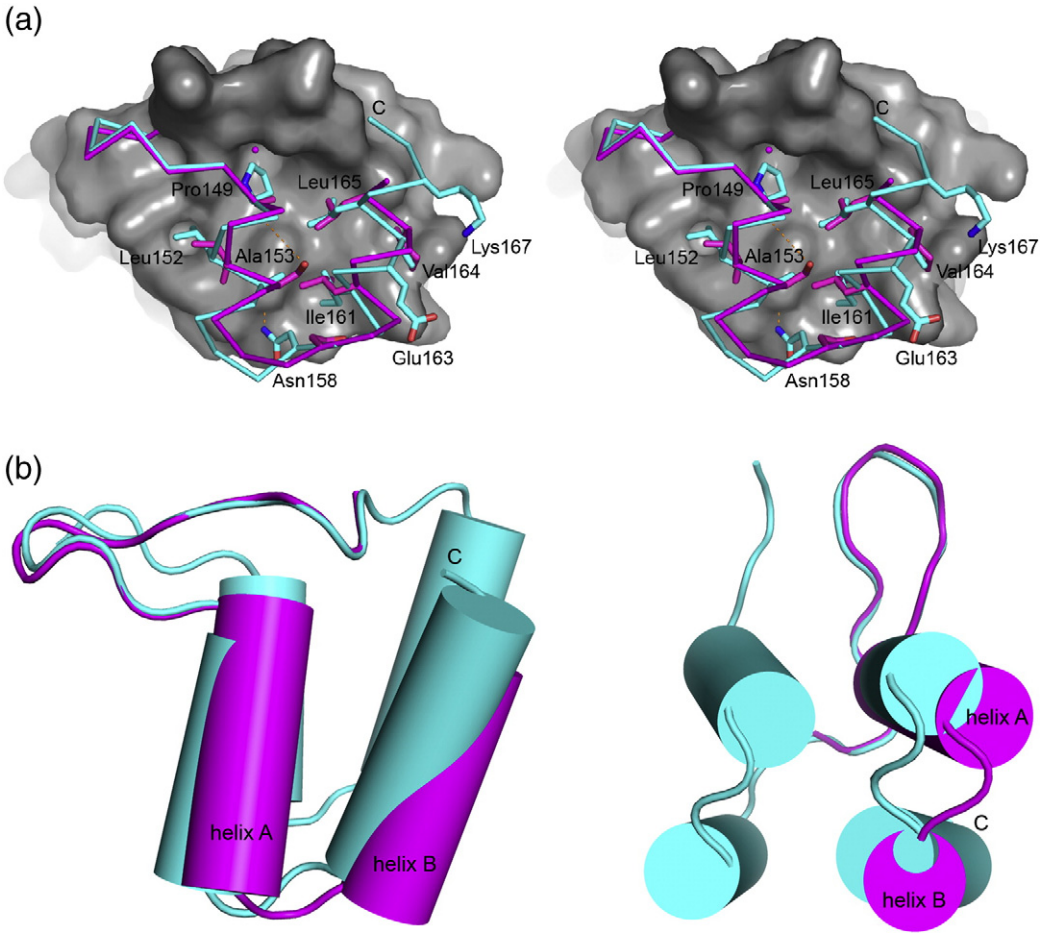

Fig. 5. Superposition of NI_3C_Mut5 (cyan) onto NI_3C with the original C-cap (magenta). (a) Residues 13–142 of NI_3C_Mut5 are depicted as a molecular surface (gray) and the C-caps are depicted as C^{α} traces. Residue nomenclature follows the sequence of NI_3C_Mut5. (b) Sketch illustrating the rigid-body movement of helices A and B. Depicted are the last internal repeat and the C-cap of NI_3C_Mut5 and the original C-cap of NI_3C. The superposition is based on the N-cap and the three internal repeats.

mutations Ala149Pro, Ile152Leu, Ser153Ala, Leu161Ile, and Ile164Val (Table 1) (equivalent to Ala83Pro, Ile86Leu, Ser87Ala, Leu95Ile, and Ile98-Val in Ref. 10 where their design has been discussed). Mutations Ala149Pro, Ser153Ala, and Leu161Ile occupy buried hydrophobic pockets at the interface between the last internal repeat and the C-cap, whereas mutations Ile152Leu and Ile164Val are partially solvent accessible (Fig. 5a). The mutation Ala149Pro is located at the N-terminus of helix A and replaces an internal water molecule in the loop region between helix A and the preceding β-strand. The Ser153Ala mutation was chosen because it increases the helix propensity.[10] The crystal structure of NI_3-C_Mut5 shows that this mutation abrogates the H-bond between Ser153-OG and Phe150-O at the interface between helices A and B. However, this mutation seems to have almost no effect on the distances between helices A and B, which are 5.29 and 5.35 Å in the original C-cap and NI_3C_Mut5, respectively (distances between C^{α} atoms at positions 153 and 162). In NI_3C_Mut5, the side chains of Leu152, Ile161, and Val164 are penetrating deeper into the corresponding cavities, with the consequence that the number of inter-repeat hydrophobic contacts has increased significantly (Table 1) and the C-cap is pulled towards the last internal repeat. This reorientation of the capping repeat allows the formation of the hydrogen bonds Asn158-ND2...Ala121-O (2.81±0.05 Å) and Asn156-ND2...Leu152-O (2.88±0.14 Å, in chains A, B, and D) in NI_3C_Mut5, which are not seen in the structure with the original C-cap (Fig. 6a). Only in molecule C does Asn156 participate in the formation of a crystal contact to Gly91-O of molecule A.

NI_3C_Mut6 contains all mutations of NI_3C_Mut5 (see above) plus the additional mutations Asp155Arg, Asn156Glu, and Asn158His (equivalent

to Asp89Arg, Asn90Glu, and Asn92His in Ref. 10 where the design is discussed). Mutations Asp155Arg and Asn156Glu had been chosen, because it was expected that the side chains of the newly introduced residues would form a stabilizing salt bridge, as it was seen in the internal repeats of NI_3C.[10] However, in NI_3C_Mut6, the distances between the guanidinium groups of Arg155 and the carboxylate groups of Glu156 were larger than 4 Å, suggesting that no salt bridges were formed. Furthermore, the Asn156Glu mutation abrogates the hydrogen bonds between Asn156-ND2 and Leu152-O or Arg122-O, which were observed in NI_3C_Mut5 (Fig. 6a). In NI_3C_Mut6, Arg122-O is in a suitable position to form a hydrogen bond with the imidazole side chain of His158. This hydrogen bond links the C-cap to the last internal repeat, as it was already seen in NI_3C_Mut5 for Asn158 (Fig. 6b). However, a 180° rotation around the C^{β}–C^{γ} bond would bring the His158 side chain in NI_3C_Mut6 into a suitable orientation to form a similar hydrogen bond with Ala121-O as seen for Asn in NI_3C_Mut5. In general, residues 155 to 158 are rather flexible in the crystal structures of NI_3C_Mut5 and NI_3C_Mut6 and consequently show different conformations and weak electron densities in some non-crystallographic symmetry (NCS)-related molecules.

As a consequence of the redesign of the C-cap, the buried surface area of the interface has increased by 15% and the surface complementarity[19] has increased from 0.705 for the original C-cap to 0.801 for NI_3C_Mut5 (Table 1). The improved fit between the C-cap and the last internal repeat is primarily a consequence of improved stacking interactions between side chains within the interface, because NI_1C_Mut4, which only harbors the extension of helix B but lacks the other mutations (Ala149Pro, Ile152Leu, Ser153Ala, Leu161Ile, and Ile164Val), reveals interface properties that are just slightly improved over the properties of the original C-cap.

The optimized interface also rigidifies the conformation of the C-cap. Although residues 142 to 169 still have larger *B*-factors[20] than the corresponding residues in the internal repeats, the *B*-factor gradient is significantly diminished in NI_3C_Mut5 and NI_3C_Mut6 compared to the original C-cap (Fig. 3). Furthermore, the optimized C-caps reveal *B*-factor profiles within the repeat that are now similar to the profiles of the internal repeats, with low *B*-factors for residues forming α-helices and higher *B*-factors for residues in loop regions.

Impact of precipitating agent on crystal twinning

Some full-consensus DARPins have been found to be prone to form twinned crystal lattices (P.R.E.M., unpublished observation). NI_3C_Mut5 crystallized from sodium citrate/polyethylene glycol conditions, and initial data sets were indexed assuming a

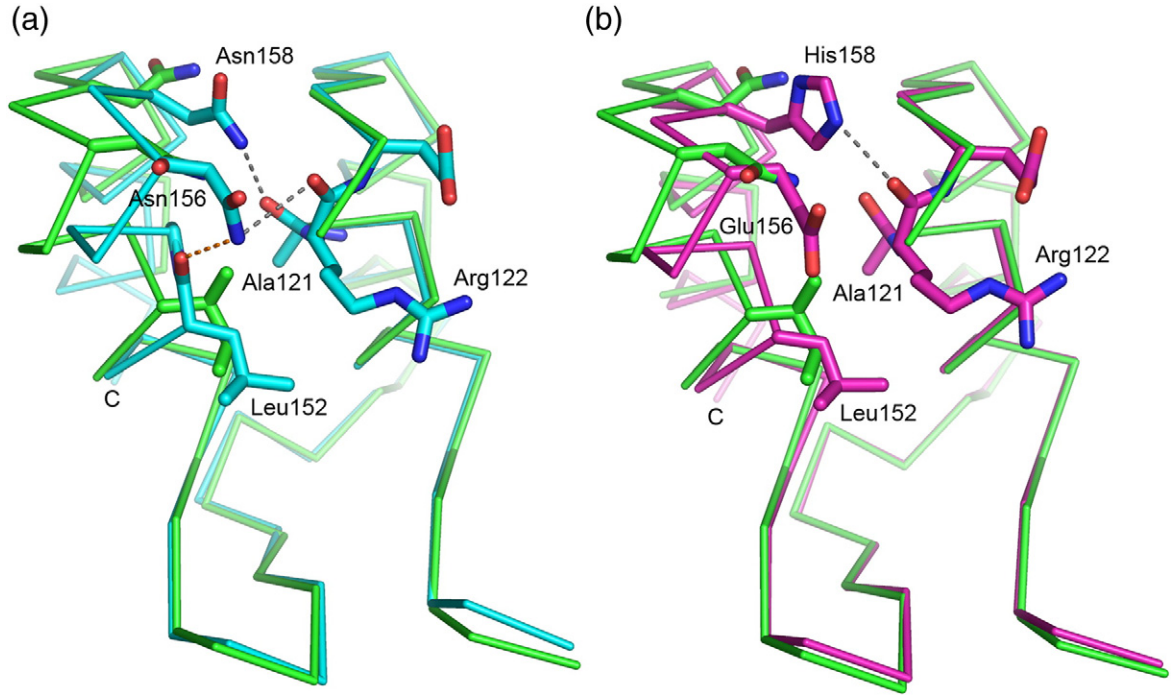

Fig. 6. Superposition of the last internal repeat and the C-cap (residues 121 to 169) of (a) NI_3C with the original C-cap (green carbon atoms) on NI_3C_Mut5 (light blue carbon atoms) or (b) NI_3C_Mut6 (magenta carbon atoms). Gray and orange broken lines indicate inter- and intra-repeat hydrogen bonds, respectively.

hexagonal crystal lattice. Since the structure could not be refined using the hexagonal settings and because data statistics suggested partial merohedral twinning, the structure was refined assuming the trigonal space group $P3_121$, twin law $(-h, -k, l)$ and four molecules in the AU. Twinning is often difficult to assign in the presence of NCS. Therefore, we used the RvR plot[21] to investigate the agreement between the observed diffraction data and the refined structure. Values for R^{obs}_{twin} of 0.17 and for R^{calc}_{twin} of 0.30 confirmed our assumption of partial merohedral twinning in the presence of rotational NCS.

Besides NI_3C_Mut5, also the NI_3C with the original C-caps[16] initially yielded twinned crystals when sodium citrate was used as precipitating agent (P.R.E.M., unpublished observation). For NI_3C, twinning was overcome by replacing sodium citrate with ammonium sulfate. NI_1C_Mut4 and NI_3C_Mut6 also crystallized from ammonium sulfate solutions, and these crystals were not twinned. Analysis of the crystal lattices revealed that sodium citrate causes the formation of crystal contacts involving N-cap/N-cap* and C-cap/C-cap* stacking interactions (where the asterisk refers to the symmetry-related molecule), whereas all crystals originating from ammonium sulfate conditions only show N-cap/C-cap* stacking interactions. In addition, well-defined sulfate ions were identified in the electron density maps during the refinement processes. In NI_1C_Mut4, one sulfate ion was observed in a crystal contact close to helix A from the internal repeat (Fig. 1a), whereas in NI_3C_Mut6, three sulfate ions occupy similar positions like in NI_3C with the original C-caps (Fig. 1b). However, these sulfate ions are not involved in crystal contacts but form salt bridges with Arg23, Arg56, Arg89, and Arg122.

Discussion

The crystal structures of NI_1C_Mut4, NI_3C_Mut5, and NI_3C_Mut6 presented above illustrate the success of the consensus sequence approach for the design of new C-caps.[10] From a pure structural point of view, the extension of helix B by Lys101, Ala102, and Ala103; the replacement of alanine against Pro83 at the N-terminus of helix A; and the introduction of the optimized interface residues Leu86, Ile95, and Val98 contribute significantly to the improved stability of the new C-cap design. Whether the increased helix propensity of the Ser87Ala mutation improves the stability cannot be deduced from the crystal structures alone. Ser87 could still be beneficial for the stability of helix A.

The mutations Asp89Arg and Asn90Glu were introduced based on MD calculations, expecting that the Arg89…Glu90 salt bridge would stabilize the C-cap.[10] Under the given crystallization conditions for NI_3C_Mut6, no such interaction was observed. This is consistent with experimental results from GdnHCl denaturation, where the stability of NI_1C_Mut5 was found to be essentially identical with that of NI_1C_Mut6 and the stability of NI_3C_Mut5 was found essentially identical with that of NI_3C_Mut6.[10] Thus, there is no indication for the formation of this salt bridge in the crystal or in solution, suggesting that this salt bridge formation might have been somewhat overemphasized in the MD calculations. Nonetheless, these calculations have been invaluable in suggesting and computationally verifying the packing and helix extension mutations.[10]

Redesign of the interface between the last internal repeat and the C-cap significantly improved the stability of full-consensus DARPins[8,10,16] as can be deduced from denaturation experiments. The crystal structures of NI_1C_Mut4, NI_3C_Mut5 and NI_3C_Mut6 reveal the molecular determinants for this stabilization. The mutation Ala149Pro, part of the conserved Thr-Pro-Leu-His motif, places a proline residue at the N-terminus of helix A. At this position, the proline participates in the hydrogen-bonding network of helix A (Pro149-O…Ala153-N, 2.94 ± 0.15 Å in NI_3C_Mut5) and improves helix formation due to its restricted ϕ-value. Furthermore, mutations Ile152Leu, Leu161Ile, and Ile164Val enable a tighter packing between the C-cap and the last internal repeat (Fig. 5a). Altogether, these mutations allow for a reorientation of the capping repeat that can be approximated by a rigid-body movement of the capping repeat towards the last internal repeat by 7° and an equal rotation along the axis of the stacked repeats (Fig. 5b). The additional hydrogen bond between Asn158-ND2 and Ala121-O is a consequence (and not a cause) of this reorientation (Fig. 6a), because the capping repeat adopts the same conformation in NI_3C_Mut6, even though this hydrogen bond is disrupted by the Asn158Glu mutation. Thus, the C-caps of NI_3C_Mut5 and NI_3C_Mut6 provide larger interface areas with the preceding internal repeat, improved surface complementarities, and, consequently, superior shielding of the hydrophobic core against the solvent, compared to NI_3C with the original C-cap derived from a natural protein. The improved stabilities are reflected by significantly lower *B*-factors of the C-caps (Fig. 3 and Table 1). This better packing observed experimentally is thus consistent with the results from MD calculations that inspired the design.[10]

Even though none of the mutations discussed above have been implemented in NI_1C_Mut4, this mutant shows improved structural stability over NI_1C with the original C-cap, as judged by GdnHCl-induced and thermal unfolding experiments[10] and almost no *B*-factor gradient. The improved stability of the capping repeat can be attributed to the extension of helix B, which provides a C-terminal

helix cap (Lys101) and additional VdW contacts with the preceding internal repeat. Whether the reorientation of His52 also contributes to the improved stability remains questionable, because its movement does not seem to be a direct consequence of the extension of helix B, whose orientation does not change. The formation of the salt bridge between His52 and Asp77 could be a consequence of modified crystallization conditions, particularly the addition of methanol. NI_1C_Mut4 was crystallized under acidic conditions (pH 5.0) and in the presence of 30% methanol, whereas all NI_3C crystals were obtained in the absence of methanol and under different pH conditions (NI_3C_Mut5 at pH 4.1, NI_3C_Mut6 at pH 5.5, and NI_3C with the original C-cap at pH 8.5). Since in all NI_3C structures His118 (the equivalent to His52 in NI_1C) forms intra-repeat H-bonds with Thr115 even when the protein was crystallized under acidic conditions, the absence of the inter-repeat salt bridge that is observed in NI_1C_Mut4 is independent of the pH of the crystallization conditions and independent of the cap mutations, pointing to methanol as a possible cause.

In all full-consensus DARPin structures, the N- and C-terminal caps are involved in crystal contacts. In the crystal lattices of non-twinned NI_3C, NI_1C_Mut4 and NI_3C_Mut6 crystal contacts are formed by N-cap/C-cap* stacking interactions (the asterisk denotes the symmetry-related molecule), whereas in the crystal lattice of twinned NI_3C_Mut5, crystal contacts are formed by N-cap/N-cap* and C-cap/C-cap* stacking interactions. The crystal lattices of NI_3C with the original C-cap and NI_3C_Mut6 are almost isomorphic. Crystals of both proteins belong to space group $P6_1$ with a unit cell *c*-axis of approximately 50 Å and similar super-helical crystal packings of NI_3C molecules. Thus, in the case of full-consensus DARPins, the precipitating agent ammonium sulfate causes the formation of crystal lattices involving N-cap/C-cap* stacking interactions, and these lattices seem to be less prone to form twinned crystals. Furthermore, an influence of the crystal lattice on the interface between the last internal repeat and the C-cap (IC interface) can be ruled out, because NI_3C with the original C-cap and NI_3C_Mut6 show significantly different IC interfaces although the crystal contacts are almost identical.

Table 2. Crystallization parameters and refinement statistics

	NI_1C_Mut4	NI_3C_Mut5	NI_3C_Mut6
Data collection			
Space group	*C*2	$P3_121$	$P6_1$
Molecules/AU	1	4	3
Unit cell parameters			
a, *b*, *c* (Å)	77.15, 35.33, 32.40	62.05, 62.05, 270.18	128.91, 128.91, 50.70
α, β, γ (°)	90, 114.43, 90	90, 90, 120	90, 90, 120
Resolution limit (Å)	31–2.2 (2.3–2.2)	50–2.1 (2.2–2.1)	50–1.80 (1.85–1.80)
Observed reflections	4117	35,848	44,397
Completeness (%)	98.8 (98.3)	98.2 (94.7)	99.4 (96.5)
Redundancy (%)	6.10 (6.38)	5.63 (5.13)	6.31 (6.02)
R_{sym} (% on *I*)	5.8 (13.1)	9.1 (39.4)	6.6 (22.7)
Refinement			
Resolution range (Å)	31–2.2	50–2.1	46–1.8
R_{work} (%)	16.8	17.5	21.7
R_{free} (%)	22.2	24.0	25.0
Ordered water molecules	53	175	257
Protein atoms	684	4742	3583
Sulfates	1	0	9
Methanol	1	0	0
Ethylene glycol	1	0	0
rmsd of bond lengths	0.008	0.002	0.004
rmsd of bond angles	0.944	0.460	0.803
Average *B*-factor	23.7	31.7	31.7
Ramachandran plot (%)			
Most favored	95.0	90.0	91.2
additional allowed	5.0	10.0	8.8
generously Disallowed	0.0	0.0	0.0
Crystallization conditions	0.05 M sodium succinate, pH 4.0, 2.9 M ammonium sulfate, 30% methanol	0.1 M sodium citrate, pH 4.1, 1.2 M lithium chloride, 30% polyethylene glycol 6000	0.1 M Hepes, pH 5.5, 0.1 M l-cysteine, 2.9 M ammonium sulfate

Values in parentheses refer to the outermost resolution shell.

Conversely, NI_3C_Mut5 and NI_3C_Mut6 show very similar IC interfaces, although the crystal contacts are significantly different.

In summary, the current work shows, from a structural perspective, the importance of the capping repeats for the stability of the protein, which fulfill the extremely important role of providing a hydrophilic surface to the protein, without which the proteins form aggregates or inclusion bodies *in vivo*. The originally used C-cap, taken from the guanine–adenine binding protein β1, does not pack optimally against the internal repeats, and the redesigned mutants, based on MD simulations and an adaptation to the overall consensus sequence, have greatly improved this packing. Recently, these proteins have also been studied by NMR.[22] Despite the repeating sequence motifs, the resonances could be fully assigned using [^{1}H,^{15}N,^{13}C] triple-labeled proteins and paramagnetic spin labels were attached to either end of these elongated proteins. Deuterium exchange experiments comparing NI_3C and NI_3C_Mut5 indicate that the stability is strongly dependent on the coupling between repeats, as the stabilized C-cap decreases the exchange rate throughout the whole protein. This emphasizes the great importance of the inter-repeat interfaces for the overall stability of the protein, as they mediate the coupling in these extended proteins. These results will thus influence the further design of repeat proteins.

Materials and Methods

Protein purification and crystallization

NI_1C_Mut4, NI_3C_Mut5, and NI_3C_Mut5 were expressed in *E. coli* and purified as described previously.[11] Size-exclusion chromatography was performed in 10 mM Hepes, pH 6.5, and 20 mM NaCl on a Hi-Load 26/60 Superdex-75 column (GE Healthcare) mounted on an Äkta Prime system (GE Healthcare). The proteins were concentrated to 30 mg/ml using Centricon (Millipore, USA) ultrafiltration devices (3 kDa molecular mass cutoff). Crystals were grown using the sitting drop-vapor diffusion method at 20 °C in 24-well crystallization plates. Proteins and reservoir solutions were mixed at 2 μL-to-1 μL ratios. Crystallization conditions are given in Table 2. Crystals were soaked in the reservoir buffers supplemented with 30% glycerol as cryo-protectant and flash-cooled in a nitrogen stream at 100 K.

X-ray data collection and refinement

X-ray diffraction data were measured at the PX6 beamline at the Swiss Light Source (Villigen, CH) using the Pilatus detector. For each structure, 360 frames were recorded with an oscillation range of 0.5° per frame. Data were processed using the program XDS.[23] Twinning of X-ray diffraction data was analyzed using the program PHENIX.XTRIAGE.[24] Initial phases were determined by molecular replacement using the program Phaser[20] with the structures of the full-consensus NI_3C (PDB ID: 2QYJ)[16] or the three-repeat AR protein $3CA_{1A2}$ (PDB ID: 2ZGD)[25] as search models. Structures were refined using the programs PHENIX[24] and Coot.[26] For NI_3C_Mut5, the partial twinning test[27] and Britton plot[28] analysis yielded twinning fractions of 0.338 and 0.351, respectively. Therefore, the structure was refined using the twin law $(-h, -k, l)$ with a twinning fraction of 0.374.

The qualities of the structures were evaluated with the programs MolProbity[29] and PROCHECK.[30] To investigate the intra-molecular shape complementarities, we used the program SC,[19] which was originally developed to calculate the shape complementarities between independent subunits. Therefore, separate coordinate files were prepared by splitting the DARPin structures at the repeat boundaries (between residues 76 and 77 for NI_1C and between residues 142 and 143 for NI_3C). PyMOL[31] was used to prepare figures.

Accession numbers

Coordinates and structure factors have been deposited at the Macromolecular Structure database with accession number 2XEN (NI_1C_Mut4), 2XEE (NI_3C_Mut5), and 2XEH (NI_3C_Mut6).

Acknowledgements

We thank the staff of the beamline X06SA at the Swiss Light Source for support in X-ray data collection. M.A.K. thanks Oliv Eidam for contributing to structure refinement of NI_3C_Mut5. This work was supported by the National Center of Competence in Research for Structural Biology and Swiss National Science Foundation grants to M.G.G. and A.P.

References

1. Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001). Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131.
2. Bork, P. (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins*, **17**, 363–374.
3. Parmeggiani, F., Pellarin, R., Larsen, A. P., Varadamsetty, G., Stumpp, M. T., Zerbe, O. *et al.* (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.
4. Main, E. R., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. (2003). Design of stable alpha-helical arrays from an idealized TPR motif. *Structure*, **11**, 497–508.
5. Stumpp, M. T., Forrer, P., Binz, H. K. & Plückthun, A. (2003). Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J. Mol. Biol.* **332**, 471–487.

6. Mosavi, L. K., Minor, D. L., Jr. & Peng, Z. Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl Acad. Sci. USA*, **99**, 16029–16034.
7. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503.
8. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. & Plückthun, A. (2008). Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *J. Mol. Biol.* **376**, 241–257.
9. Mosavi, L. K. & Peng, Z. Y. (2003). Structure-based substitutions for increased solubility of a designed protein. *Protein Eng.* **16**, 739–745.
10. Interlandi, G., Wetzel, S. K., Settanni, G., Plückthun, A. & Caflisch, A. (2008). Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. *J. Mol. Biol.* **375**, 837–854.
11. Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P. *et al.* (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582.
12. Steiner, D., Forrer, P. & Plückthun, A. (2008). Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *J. Mol. Biol.* **382**, 1211–1227.
13. Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl Acad. Sci. USA*, **100**, 1700–1705.
14. Amstutz, P., Binz, H. K., Parizek, P., Stumpp, M. T., Kohl, A., Grütter, M. G. *et al.* (2005). Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *J. Biol. Chem.* **280**, 24715–24722.
15. Kohl, A., Amstutz, P., Parizek, P., Binz, H. K., Briand, C., Capitani, G. *et al.* (2005). Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure*, **13**, 1131–1141.
16. Merz, T., Wetzel, S. K., Firbank, S., Plückthun, A., Grütter, M. G. & Mittl, P. R. (2008). Stabilizing ionic interactions in a full-consensus ankyrin repeat protein. *J. Mol. Biol.* **376**, 232–240.
17. Yu, H., Kohl, A., Binz, H. K., Plückthun, A., Grütter, M. G. & van Gunsteren, W. F. (2006). Molecular dynamics study of the stabilities of consensus designed ankyrin repeat proteins. *Proteins*, **65**, 285–295.
18. Binz, H. K., Kohl, A., Plückthun, A. & Grütter, M. G. (2006). Crystal structure of a consensus-designed ankyrin repeat protein: implications for stability. *Proteins*, **65**, 280–284.
19. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
20. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
21. Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). Intensity statistics in twinned crystals with examples from the PDB. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **62**, 83–95.
22. Wetzel, S. K., Ewald, C., Settanni, G., Jurt, S., Plückthun, A. & Zerbe, O. (2010). Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR. *J. Mol. Biol.* **402**, 241–258.
23. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
24. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W. *et al.* (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **58**, 1948–1954.
25. Kelly, L., McDonough, M. A., Coleman, M. L., Ratcliffe, P. J. & Schofield, C. J. (2009). Asparagine beta-hydroxylation stabilizes the ankyrin repeat domain fold. *Mol. Biosyst.* **5**, 52–58.
26. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
27. Yeates, T. O. (1997). Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344–358.
28. Fisher, R. G. & Sweet, R. M. (1980). Treatment of diffraction data from crystals twinned by merohedry. *Acta Crystallogr. Sect. A: Found. Crystallogr.* **36**, 755–760.
29. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–383.
30. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067.
31. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA.