doi:10.1016/j.jmb.2007.11.047 J. Mol. Biol. (2008) 376, 232–240

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Stabilizing Ionic Interactions in a Full-consensus Ankyrin Repeat Protein

Tobias Merz, Svava K. Wetzel, Susan Firbank, Andreas Plückthun, Markus G. Grütter and Peer R. E. Mittl*

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

Received 4 July 2007; received in revised form 28 September 2007; accepted 16 November 2007
Available online
22 November 2007

Full-consensus designed ankyrin repeat proteins (DARPins), in which randomized positions of the previously described DARPin library have been fixed, are characterized. They show exceptionally high thermodynamic stabilities, even when compared to members of consensus DARPin libraries and even more so when compared to naturally occurring ankyrin repeat proteins. We determined the crystal structure of a full-consensus DARPin, containing an N-capping repeat, three identical internal repeats and a C-capping repeat at 2.05 Å resolution, and compared its structure with that of the related DARPin library members E3_5 and E3_19. This structural comparison suggests that primarily salt bridges on the surface, which arrange in a network with almost crystal-like regularity, increase thermostability in the full-consensus NI_3C DARPin to make it resistant to boiling. In the crystal structure, three sulfate ions complement this network. Thermal denaturation experiments in guanidine hydrochloride directly indicate a contribution of sulfate binding to the stability, providing further evidence for the stabilizing effect of surface-exposed electrostatic interactions and regular charge networks. The charged residues at the place of randomized residues in the DARPin libraries were selected based on sequence statistics and suggested that the charge interaction network is a hidden design feature of this protein family. Ankyrins can therefore use design principles from proteins of thermophilic organisms and reach at least similar stabilities.

Keywords: repeat protein; protein stability; salt bridge; thermodynamic stability; X-ray crystallography

Edited by F. Schmid

Introduction

Repeat proteins consist of repeating structural units that stack together to form elongated non-globular domains.[1,2] In contrast to globular proteins, they are not stabilized by interactions between residues that are very distant in sequence; instead, the stabilizing and structure-determining interactions are formed within a repeat and between neighboring repeats. Repeat proteins can be extended in size while still constituting a contiguous domain, making them unique targets for protein engineering. Repeat proteins constitute, next to immunoglobulins, the most abundant natural protein classes specialized in binding.

Because of their abundance and the multiple occurrences of repeats within one protein sequence, a statistical analysis of thousands of sequences can be carried out to design consensus repeats. This has been reported for ankyrin repeat (AR) proteins, tetratricopeptide repeat proteins and leucine-rich repeat proteins.[3–6] The AR is one of the most common protein sequence motifs. This 33-residue motif consists of a β-turn, followed by two antiparallel α-helices and a loop reaching the turn of the next repeat.[7]

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

Present address: S. Firbank, Institute of Cell and Molecular Biosciences, University of Newcastle, Framlington Place, Newcastle upon Tyne NE2 4HH, UK.

Abbreviations used: DARPin, designed ankyrin repeat protein; AR, ankyrin repeat; PDB, Protein Data Bank; SC, shape complementarity; HB, hydrogen bonds; Gdn·HCl, guanidinium hydrochloride.

Two independent approaches used to apply consensus design strategies to AR proteins have been reported so far.[3,4] Both employ a similar but not identical consensus version of the 33-residue AR.[7] In one approach, large libraries of AR proteins,[3] in which only 26 of the 33 amino acids were specified while 7 were allowed to vary in order to bind to different target molecules, were made.[8]

The potential target interaction residues of a single AR (2, 3, 5, 13, 14 and 33), henceforth named randomized residues, are located on the concave side of designed ankyrin repeat proteins (DARPins).[3] Highly specific binders for a number of different target proteins have been selected using the DARPin libraries[8], and structures of ankyrin–target protein complexes have been determined.[9–12] In this approach, N- and C-capping repeats flanking the randomized repeats that shield the hydrophobic core were employed. Recently, it was shown experimentally that indeed the capping repeats are required for soluble expression in *Escherichia coli*[13] and for shielding of the protein against aggregation.

In another approach, a full consensus was derived from sequence statistics.[4] In these studies, no capping repeats were employed, and the resulting proteins were only soluble at acidic pH.[4] The introduction of positive charges in the C-terminal consensus repeat then allowed the protein to be soluble at neutral pH, but it still had to be produced by *in vitro* refolding from inclusion bodies made in *E. coli*.[5] However, the gain in solubility was accompanied by a significant loss in stability at pH 4.

As described in detail elsewhere in this issue,[14] a full-consensus version of the DARPins, in which randomized residues of the library have now been defined such that every internal repeat has exactly the same sequence, was designed. We denote these proteins as NI_xC, where "N" and "C" refer to the N- and C-terminal capping repeats, respectively, "I" refers to the internal full-consensus repeat and the subscript x gives the number of identical internal consensus repeats. The design of NI_xC proteins and their thermodynamic and kinetic folding properties have been investigated.[14] The N- and C-terminal capping repeats were originally taken[3] from GABP β1 (guanine-adenine binding protein β1) (PDB code 1AWC).[15]

Here we report the crystal structure of the full-consensus DARPin NI_3C at 2.05 Å resolution and compare its features with the structures of two related consensus DARPins E3_5 and E3_19 [Protein Data Bank (PDB) entries 1MJ0 and 2BKG, respectively]. All three molecules were designed using the same framework residues. In contrast to NI_3C, the internal repeats of E3_5[16] and E3_19[17] do not have the same sequences because they contain different residues at the randomized positions. From our structural analysis, we propose that a highly regular array of salt bridges, the overall charge distribution and the binding of sulfate ions significantly contribute to the thermodynamic stability of NI_3C. This study is intended to extend our understanding of stabilizing charge effects in AR proteins.

Results

Overall structure

The final model of the full-consensus NI_3C DARPin structure at 2.05 Å resolution contains all amino acids and side chains of the three identical internal repeats and the flanking N- and C-terminal capping repeats. The capping repeats expose a hydrophilic surface and shield the hydrophobic core of the internal repeats from the solvent to prevent aggregation. As in all AR proteins, the repeats adopt an L-shaped arrangement, which is caused by two almost antiparallel α-helices and a β-turn forming the interrepeat connection (Fig. 1a). Residues positioned at the concave side of the DARPin library (randomized positions) normally mediate specific interactions with selected targets. Based on the statistics of naturally occurring AR protein sequences, they were designed to be highly charged in the full-consensus version.[14] As a consequence, three sulfate ions from the crystallization buffer were found to bind to this site (Fig. 1b).

E3_5/E3_19 consensus *versus* NI_3C full consensus

Sequence differences within repeats between the full-consensus molecule NI_3C and the unselected library members E3_5 and E3_19 occur at positions 2, 3, 5, 13, 14, 26 and 33 (for repeat numbering, see Fig. 2). Residues at these randomized positions depend on the selection process and typically differ from one repeat to the next. In contrast to E3_5 and E3_19, all internal repeats in NI_3C are identical (Fig. 2). As discussed in detail elsewhere in this issue,[14] amino acid types were selected by sequence statistics and structural considerations.

Briefly, at positions 2 and 3, lysine and aspartic acid residues were introduced, respectively. While lysine is the most prevalent amino acid at position 2, aspartic acid is the second most abundant amino acid at position 3 (after asparagine). A negatively charged amino acid was selected for position 3 to balance the positive charges that were introduced at positions 2 and 33. For position 5, which is quite variable, tyrosine was introduced as a spectroscopic probe. In position 13, the second most abundant arginine (after glutamine) was used to compensate for negative charges, and the most abundant glutamic acid was selected for position 14. Alanine was inserted at position 26 because it is most abundant at this position and possesses high helical propensity. For position 33, the most abundant lysine was selected. Figure 2 shows the correspondence between the numbering within a single repeat and the protein sequence.

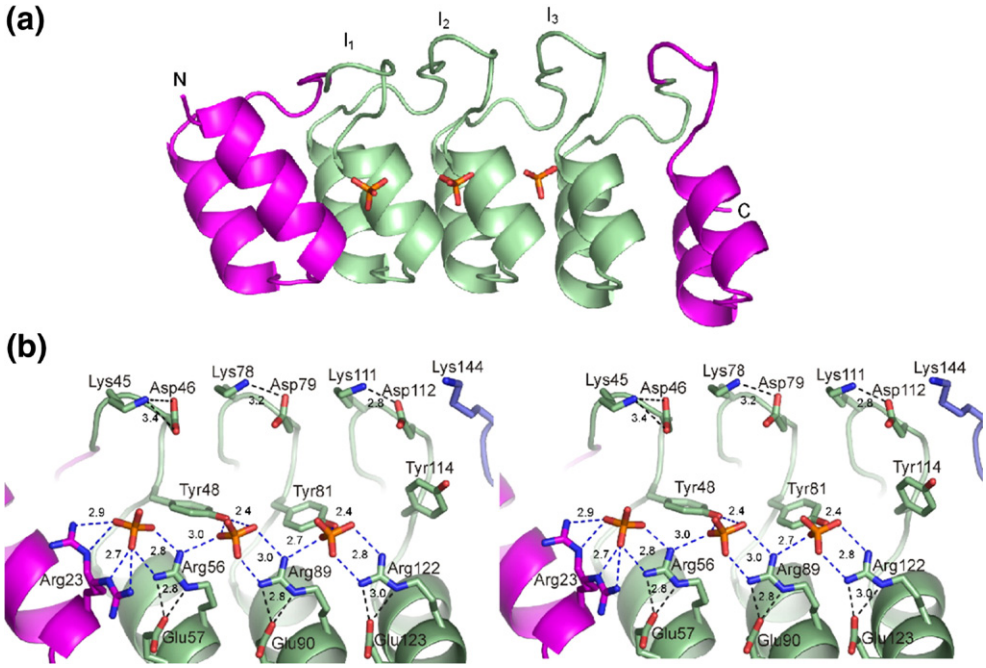

Fig. 1. (a) Ribbon diagram of the NI_3C structure. The terminal and internal repeats are in magenta and green, respectively. (b) Stereo view of the interactions involving the three sulfate ions and the randomized residues. Dashed lines in black denote HB between lysine and aspartate residues at positions 2 and 3, as well as HB between arginine and glutamate residues at positions 13 and 14. Dashed lines in blue denote all interactions of the sulfate ions.

Structural comparison with other AR proteins

The overall structures of NI_3C, E3_5 and E3_19 are very similar. The RMSD values of the pairwise comparisons of Cα atoms between NI_3C/E3_5, NI_3C/E3_19 and E3_5/E3_19 are 0.62 Å, 0.50 Å and 0.60 Å, respectively. Superpositions of the three molecules NI_3C, E3_5 and E3_19 revealed the largest structural differences in the C-terminal capping repeats, which move, in a first approximation, as rigid bodies. This is illustrated by a significant reduction in the overall RMSD values when omitting

```
SEQUENCE          .......20........*.........40..
      NI3C          DLGKKLLEAARAGQDDEVRILMANGADVNAK  :43
N-CAP E3_5          DLGKKLLEAARAGQDDEVRILMANGADVNAT  :43
      E3_19         DLGKKLLEAARAGQDDEVRILMANGADVNAE  :43

POSITION          .........10.........20.........30..
SEQUENCE          ......*.........60.........*......
      NI3C        DKDGYTPLHLAAREGHLEIVEVLLKAGADVNAK  :76
1st   E3_5        DNDGYTPLHLAASNGHLEIVEVLLKNGADVNAS  :76
      E3_19       DTYGDTPLHLAARVGHLEIVEVLLKNGADVNAL  :76

POSITION          .........10.........20.........30..
SEQUENCE          ...80.........*.........100.......
      NI3C        DKDGYTPLHLAAREGHLEIVEVLLKAGADVNAK  :109
2nd   E3_5        DLTGITPLHLAAATGHLEIVEVLLKHGADVNAY  :109
      E3_19       DFSGSTPLHLAAKRGHLEIVEVLLKYGADVNAD  :109

POSITION          .........10.........20.........30..
SEQUENCE          *.........120.......*.........140
      NI3C        DKDGYTPLHLAAREGHLEIVEVLLKAGADVNAQ  :142
3rd   E3_5        DNDGHTPLHLAAKYGHLEIVEVLLKHGADVNAQ  :142
      E3_19       DTIGSTPLHLAADTGHLEIVEVLLKYGADVNAQ  :142

SEQUENCE          .......*.........160....
      NI3C        DKFGKTAFDISIDNGNEDLAEILQ           :166
C-Cap E3_5        DKFGKTAFDISIDNGNEDLAEILQ           :166
      E3_19       DKFGKTAFDISIDNGNEDLAEILQ           :166
```

Fig. 2. Sequence alignment and numbering of NI_3C, E3_5 and E3_19. The N-terminal capping repeat includes a His_6 tag (MRGSHHHHHHGS; sequence not shown). Identical residues are shown against a black background, whereas randomized residues at positions 2, 3, 5, 13, 14, 26 and 33 in the first, second and third internal repeats are shown against a white background. POSITION: numbering within one repeat; SEQUENCE: numbering throughout the sequence.

the C-terminal repeats. The RMSD values for Cα atoms of residues 1–139 for the pairs NI_3C/E3_5, NI_3C/E3_19 and E3_5/E3_19 are 0.29 Å, 0.39 Å, and 0.36 Å, respectively. Shape complementarity (SC) values were analyzed between internal repeats and fragments thereof (see Materials and Methods). They were, in most cases, higher between internal consensus repeats than between the capping repeats and the adjacent internal consensus repeats (Table 1). This also illustrates the success of consensus design, as the internal repeats can apparently be very well stacked.

Hydrogen bond network

The total numbers of intramolecular hydrogen bonds (HB) in NI_3C, E3_5 and E3_19 are 152, 149 and 152, respectively, excluding those involved in binding sulfate ions. Interrepeat HB are least frequent in NI_3C, with 14 HB, compared to 16 and 17 HB in E3_5 and E3_19, respectively. Due to the high sequence identity, the hydrogen-bonding pattern is very similar in all three structures, as indicated by 119 common HB. The main-chain hydrogen-bonding pattern of the conserved TPLH sequence motif (residues 6–9) at the beginning of the first α-helix of every repeat is identical in NI_3C, E3_5 and E3_19. Small structural differences between different types of amino acids at the randomized positions enable the formation of specific HB. In NI_3C, E3_5 and E3_19, there are 22, 17 and 19 specific HB, respectively.

A detailed analysis of the total number of HB at the randomized positions reveals that 10 out of 34 HB in NI_3C, 5 out of 30 HB in E3_5 and 6 out of 30 HB in E3_19 are formed by side chains, again excluding those involved in binding sulfate ions. In contrast, the majority of the randomized residues contribute with their backbone atoms to the overall hydrogen-bonding network.

The rather modest number of side-chain HB from randomized residues in E3_5 and E3_19 shows that these residues are more relevant to the binding of the target molecules rather than to the formation of extended hydrogen-bonding networks on the surface of the molecules (Table 1). The side chains at the previously randomized positions specified in the full-consensus NI_3C molecule are involved in highly regular charge–charge interactions. This charge network is extended by sulfate ions from the crystallization buffer by bridging four arginine and two tyrosines residues in the first two repeats (Fig. 1b).

Charge network

Previous reports on DARPins[3,4,16] emphasized the highly regular hydrogen-bonding patterns between "framework" parts (i.e., contributed by the constant part of the sequence as important for stability). The designed ankyrins displayed a significantly higher thermodynamic stability compared to naturally occurring AR proteins,[3,4] which do not have such regular hydrogen-bonding networks and show greater variability.

The full-consensus AR NI_3C possesses, in addition, a highly regular charge distribution, which spans about half of the protein surface and causes an even higher thermodynamic stability compared to other library members. NI_3C contains a regular array of lysine and aspartic acid residues at positions 2 and 3 in the β-turns of internal repeats, as well as arginine and glutamic acid residues at positions 13 and 14 at the C-terminal ends of helices in the concave binding region. Arg23 positioned in the N-terminal capping repeat and Lys144 in the β-turn of the C-terminal capping repeat extend the charge network. The electron density of Arg23 suggests a double conformation. Both conformations allow the formation of salt bridges with a first sulfate ion. In the concave binding region, three equally spaced sulfate ions form salt bridges with Arg56, Arg89 and Arg122 (corresponding to position 13 of the internal repeat) and HB with Tyr48 and Tyr81 (position 5 in the internal repeat). In addition, these arginine residues participate in salt bridges with Glu57, Glu90 and Glu123 (corresponding to position 14 of the internal repeats) (Fig. 1b).

Thermal denaturation measurements

The thermal denaturation of NI_3C was monitored at 222 nm using CD spectroscopy. Under physiological conditions, NI_3C, unlike E3_5 and E3_19,[3,17] did not show a transition in thermal denaturation experiments.[14] In order to denature the protein

Table 1. Summary of the structural analysis between NI_3C, E3_5 and E3_19

Property	Ankyrins	NI_3C	E3_5	E3_19	
Identity (%)	NI_3C	100	89	88	
	E3_5	88	100	88	
	E3_19	87	87	100	
SC (shape complementary)	NI_3C	0.714 (SC_{N-1})	0.839 (SC_{1-2})	0.803 (SC_{2-3})	0.705 (SC_{3-C})
	E3_5	0.722 (SC_{N-1})	0.762 (SC_{1-2})	0.805 (SC_{2-3})	0.798 (SC_{3-C})
	E3_19	0.702 (SC_{N-1})	0.812 (SC_{1-2})	0.769 (SC_{2-3})	0.739 (SC_{3-C})
HB (overall)	Total	152	149	152	
	Unique	22	17	19	
	Common				119
HB (randomized residues)	Total	34	30	30	
	Side chain	10	5	6	
	Main chain	24	25	24	

completely and to reach the posttransition baseline, thermal denaturation measurements were performed in the presence of 4 M guanidinium hydrochloride Gdn·HCl. Denaturation experiments of NI_3C in Gdn·HCl had revealed a complex unfolding mechanism, which was interpreted as a partial unfolding of the C-terminal capping repeat prior to the main transition.[13,14]

A remarkable shift towards higher transition temperatures in the thermal melting curves of NI_3C was found with increasing sulfate concentrations (Fig. 3a and b). The increase in melting temperature is much more modest with increasing sodium chloride concentrations. The analysis of T_m as a function of ionic strength revealed that T_m increases linearly with ionic strength, which is in agreement with previous studies,[18] but the slope of the regression line is significantly higher for sulfate (0.0092 °C mM^{-1}; $R^2=0.9884$) than for phosphate (0.0064 °C mM^{-1}; $R^2=0.9767$) or chloride (0.0021 °C mM^{-1}; $R^2=0.9268$) ions (Fig. 3c). In contrast, E3_19, which does not have specific sulfate-binding sites and whose limited stability has been proposed to be a consequence of local repulsions,[17] is also stabilized by sodium chloride and sulfate (Fig. 3d). In this case, the melting point depends only on ionic strength and is independent of salt type.

Discussion

Thermal and chemical stabilities are highly relevant to the biotechnological and biomedical applications of proteins and are thus an important design goal. Consensus design exploits sequence statistics and is based on the assumption that the most frequently occurring residues in a first approximation are correlated with molecules of high thermal stability.[19] As most random mutations are destabilizing, only those protein variants that have stabilizing residues elsewhere can tolerate them. The consensus of all naturally occurring sequences would be expected to reflect these favorable residue combinations. A structural inspection is still required to avoid designing mutually exclusive residues. Compared to naturally occurring ankyrin proteins with similar numbers of repeats, consensus-designed AR proteins showed higher thermodynamic stabilities and increased stabilities towards chemical denaturants.[5,16]

The previous consensus design of DARPins has been limited to "framework" positions. The original goal was to create a library with the potential to recognize a wide range of target molecules. Consequently, residues at the randomized positions were allowed to vary and exerted attractive and repulsive charge interactions. Nonetheless, key residues specifying intrarepeat and interrepeat HB were present in every one of the internal DARPin repeats, but not necessarily in natural AR proteins, thereby partially explaining the greater stability of the designed library molecules.[3,16]

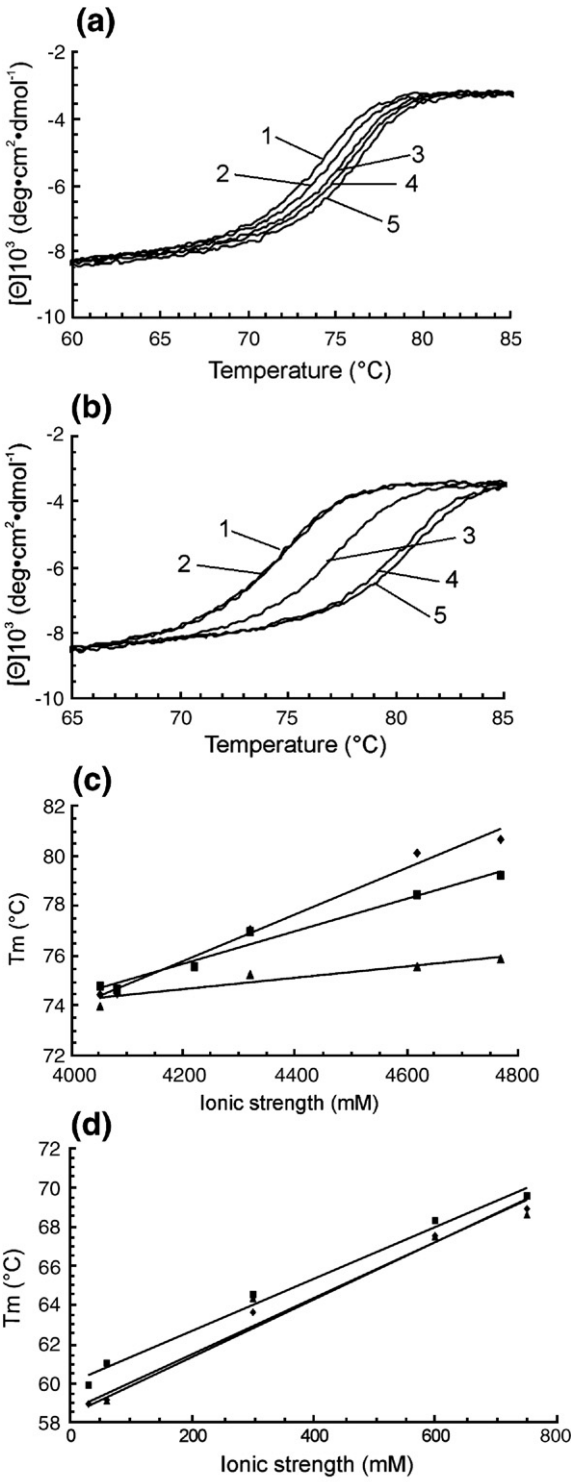

Fig. 3. Thermal melting curves of NI_3C in 20 mM Hepes, 4 M Gdn·HCl (pH 7.4) and different salt conditions: (a) sodium chloride: 30 mM (curve 1), 60 mM (curve 2), 300 mM (curve 3), 600 mM (curve 4) and 750 mM (curve 5); (b) sodium sulfate: 10 mM (curve 1), 20 mM (curve 2), 100 mM (curve 3), 200 mM (curve 4) and 250 mM (curve 5). (c) Melting points (T_m) of NI_3C plotted as a function of ionic strength for different sodium chloride (▲), sodium sulfate (♦) and sodium phosphate (■) concentrations. (d) Melting points of E3_19 as a function of ionic strengths in the same buffers as for NI_3C, but omitting Gdn·HCl. The estimated errors on T_m are ±0.5 °C.

In the present work, we investigated the structural consequences of a full-consensus design where the sequence variability of internal repeats has been eliminated. In order to correlate structural features with the increased thermal stability of the full-consensus protein, we compared the crystal structures of the full-consensus ankyrin NI_3C with the previously published consensus ankyrins E3_5[16] and E3_19[17] (PDB codes 1MJO and 2BKG, respectively). A comparison of RMSD values (Table 1) and visual inspection of Cα–Cα distance plots (data not shown) indicate no major structural differences between the molecules, except within the last repeat. They have almost the same number of HB, and most of them are common between all three molecules. Contacts between repeat interfaces were measured qualitatively by analyzing shape complementarities.[20] Differences in SC values between internal repeats of NI_3C, E3_5 and E3_19 indicate that the stackings of internal repeats differ due to minor side-chain rearrangements. It is unlikely, however, that these differences explain the higher stability of NI_3C.

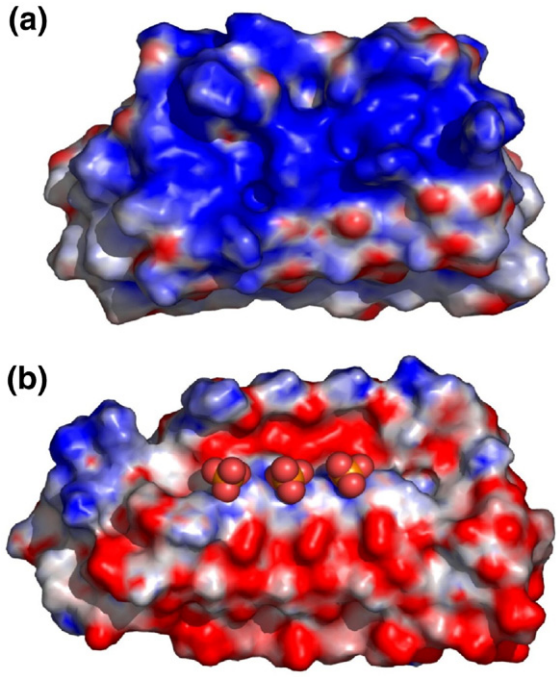

Fig. 4. Electrostatic surface potentials, calculated using the Adaptive Poisson–Boltzmann Solver[21] of (a) 4ANK (PDB code 1N0R) and (b) NI_3C. Both representations show the concave binding site with the N- and C-termini on the right and left sides, respectively. Bound sulfate ions in the NI_3C molecule are shown as spheres. Both surfaces were in blue and red for positive and negative electrostatic charge densities (same scale), respectively.

Charges and charge networks

The major difference between NI_3C, E3_5 and E3_19 is observed in the surface charge distribution. As explained above, the NI_3C design was based on sequence statistics, leading to a charge network that must have been selected during the evolution of AR proteins. One may speculate that residues involved in stabilizing charge networks are most strongly selected and thus dominate sequence statistics of surface residues, whereas residues that are required for target recognition cancel out in the alignments of whole sequence families.

In the work of Mosavi *et al.*[4], a different full-consensus sequence had been derived. In their work, positions 3 and 14 were chosen as uncharged (both asparagine), and an additional charge was introduced by an arginine at position 5 (tyrosine in this work). Additionally, there were two charge reversals in the framework of helix 2, which are lysine at position 21 (here glutamic acid because of its occurrence in GA-binding protein β1) and glutamic acid at position 25 (here lysine because of the negative charge at position 21). In the design of Mosavi *et al.*[4] (PDB entries 1N0Q and 1N0R), these residues generated a cluster of positively charged residues on the concave side of the molecule that might be electrostatically unfavorable (Fig. 4a). In contrast to this, the surface potential of the NI_3C molecule is more balanced in this area (Fig. 4b), especially when considering the protein without the sulfate ions.

Nature has used a multitude of strategies for the adaptation of proteins to life at high temperatures.[22] Surface charges were, for quite some time, considered as rather unimportant for protein stability. It was argued that the high dielectric constant of the solvent would decrease the strength of charge–charge interactions.[23] Recently, it was shown that computational redesign of surface charges can improve protein stability significantly,[24] and structural genomics data on proteins from *Thermotoga maritima* revealed a significant increase in the density of salt bridges in proteins from this organism compared to their mesophilic counterparts.[25] In contrast, the importance of oligomerization order, HB and secondary structure was found to be less pronounced than previously assumed. Similar results were found in another comprehensive structural bioinformatics study[26] and in more specialized studies on individual protein families.[27–30] The high number of salt bridges in hyperthermophilic proteins was explained by the diminishing desolvation penalty for salt bridges at increasing temperatures.[31] Therefore, the contribution of salt bridges to protein stability becomes more important at higher temperatures. NI_3C has the smallest and most balanced negative charge excess among all three molecules studied here and, thus, essentially fewer local repulsions. A higher charge imbalance on E3_19 compared to E3_5 was proposed to be the major cause of its decreased stability.[17,32]

Another possible explanation for the extraordinarily high temperature stability, even in the absence of sulfate ions, is the flexibility of the salt bridge network, giving it an entropic advantage. Although the surface-exposed ion pair network looks rigid in

the present NI_3C structure, only modest side-chain rearrangements would be necessary to swap between intrarepeat and interrepeat salt bridges. We suggest that the sum of all favorable charge interactions (and the absence of unfavorable ones) of the full-consensus ankyrin NI_3C is responsible for its significantly increased thermal stability. It is remarkable that the sequence statistics, at least in part, mirror these favorable properties.

Sulfates in the charge network

Three regularly arranged sulfate ions form the core of the salt bridge network on the concave side of the molecule. Thermal melting experiments in the presence and in the absence of sulfate ions clearly confirm the contribution of sulfate to the thermal stability of the full-consensus NI_3C ankyrin. It should be noted that thermal stability is already very high in the absence of sulfate. Even though the surface potential is balanced (Fig. 4b), the protein can bind three sulfate ions, which increase the total negative charge excess of the molecule from 12 e^- to 18 e^-. Because of the alternating arrangement of the charges, sulfate binding is energetically favored, and the stabilizing effect may be rationalized by the very short hydrogen-bonding distance of 2.4 Å between the sulfate ions and tyrosine residues at position 5 (Fig. 1b).

Summary and conclusion

The crystal structure and the thermal denaturation studies of the designed full-consensus AR protein NI_3C suggest that the extended salt bridge network positively influences the stability of the protein. This protein can only be denatured by heating in the presence of 4 M Gdn·HCl. NI_3C was designed based mainly on sequence statistics, suggesting that this charge network among previously randomized residues must also be a hidden design feature of natural ARs. In naturally occurring AR proteins, this feature is no longer visible because the sequences have drifted due to functional selection. Designing such charge interaction networks may thus be a strategy for the development of extremely stable proteins for biomedical and other biotechnical applications. However, since many of these residues normally participate in binding, a compromise between function and thermal stability by surface charge interactions will have to be found.

Materials and Methods

Expression and purification

The full-consensus NI_3C ankyrin was expressed in *E. coli* strain XL-1 blue in LB medium supplemented with 1% glucose, 100 μg/ml ampicillin and 15 μg/ml tetracycline. After growing to an OD_{600} of approximately 0.8–1.0 at 37 °C, the cultures were induced with 0.5 mM IPTG, and growth continued for 5 h. After centrifugation and resuspension in 50 mM Tris, 500 mM NaCl and 20 mM imidazole at pH 8.0, the cells were disrupted using a French press. The protein was purified in self-packed columns containing a Ni-NTA matrix (Qiagen) in accordance with the manufacturer's instructions. The protein was further purified by gel filtration on a HiLoad 26/60 Superdex 75 (Pharmacia Biotech) with an ÄKTA Prime system (Amersham, Pharmacia Biotech) in 10 mM Hepes and 10 mM NaCl (pH 7.4) and concentrated to 20 mg/ml using a Centricon (Millipore USA) with 3-kDa molecular mass cutoff.

Crystallization

Crystals for X-ray diffraction data collection were grown in 24-well crystallization plates using the hanging-drop vapor-diffusion method. A 1-μl protein solution was mixed with 1 μl of reservoir solution containing 2.7 M $(NH_4)_2SO_4$ and 100 mM Tris (pH 8.5). The crystals grown under these conditions were soaked in reservoir solution supplemented with 20% glycerol as cryoprotectant and flash-cooled in a nitrogen stream at 100 K. The X-ray diffraction data of one single crystal were collected using CuKα radiation generated by a Nonius FR591 rotating anode generator.

Structure solution and refinement

Diffraction data up to 2.05 Å resolution on a total of 180 frames were recorded within an oscillation range of 1°. The data were processed with program XDS.[33] A Matthews coefficient of $V_M = 2.4$ Å^3 Da^{-1} suggested one molecule in the asymmetric unit. Initial phases were obtained by molecular replacement using the program

Table 2. Statistics for data collection and refinement

Data collection	
Space group	$P6_1$
Cell dimensions	$a = 74.48$ Å, $b = 74.48$ Å, $c = 50.99$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$
Resolution limits (Å)	64.5–2.05
Observed reflections	78,634 (overall); 10,232 (unique)
Completeness (%)	99.8 (98.2)
Redundancy	7.7 (7.1)
R_{sym} (% on I)	7.3 (37.0)
Refinement	
Resolution range (Å)	64.5–2.05
R-factor/R_{free} (%)	18.6/22.6
Ordered water molecules	110
RMSD from ideal geometry	
Bond length (Å)	0.010
Bond angles (°)	1.109
Average B-factor (Å^2)	26.3
Residues in the most favored region (%)	91.9
Residues in the additionally allowed region (%)	8.1
Residues in the generously allowed region (%)	0.0
Residues in the disallowed region (%)	0.0

Numbers inside parentheses refer to the highest resolution shell (2.1–2.05 Å).

AMoRe,[34] with the structure of E3_5 (PDB code 1MJO) as a search model. The structure was refined to final *R*-factors of R_{work}=18.6% and R_{free}=22.6% in the space group $P6_1$ using the program REFMAC5.[35] Model building was performed with program O.[36] Statistics for diffraction data and structure refinement are summarized in Table 2. The program PROCHECK[37] was used to evaluate deviations from standard geometries, and the programs HBPLUS[38] and SC[20] were used to analyze the hydrogen-bonding pattern and calculations of surface complementarities. Although the SC program was originally developed to calculate the surface complementarity between separated subunits, we used this method to investigate intramolecular shape complementarities. For this purpose, the molecule was split into fragments by inserting artificial chain breaks between residues 31/32, 64/65, 97/98 and 130/131. The SC_{N-1}, SC_{1-2}, SC_{2-3} and SC_{3-C} values refer to the interface SC values of neighboring repeats. Figures were generated with PyMOL.[39]

Thermal denaturation experiments

Thermal denaturation experiments were performed with a Jasco J-715 instrument (Jasco, Japan). CD data were recorded at a protein concentration of 40 μM in the temperature range between 5 °C and 95 °C within 90 min at a wavelength of 222 nm (measuring interval, 10 s; bandwidth, 2 nm). The buffers contained 20 mM Hepes, 4 M Gdn·HCl (pH 7.4) and salt concentrations of between 30 and 750 mM sodium chloride or of between 10 and 250 mM sodium sulfate. The same buffers without Gdn·HCl were used for heat denaturation experiments with E3_19. Ionic strengths were calculated as follows: $I=1/2\sum_i c_i z_i^2$, where c=concentration and z=charge of ion.

Protein Data Bank accession number

Coordinates and diffraction data have been deposited at the PDB under accession number 2QYJ.

Acknowledgement

Financial support from the Swiss National Science Foundation and the Baugartenstiftung (Zurich, Switzerland) is gratefully acknowledged.

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: Struct. Funct. Genet.* **17**, 363–374.
3. 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.
4. 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.
5. Mosavi, L. K. & Peng, Z. Y. (2003). Structure-based substitutions for increased solubility of a designed protein. *Protein Eng.* **16**, 739–745.
6. 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.
7. Sedgwick, S. G. & Smerdon, S. J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316.
8. 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.
9. 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.
10. 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**(8), 1131–1141.
11. Schweizer, A., Roschitzki-Voser, H., Amstutz, P., Briand, C., Gulotti-Georgieva, M., Prenosil, E. *et al.* (2007). Inhibition of Caspase-2 by a designed ankyrin repeat protein: specificity, structure and inhibition mechanism. *Structure*, **15**(15), 625–635.
12. Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grütter, M. G. (2007). Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol.* **5**(1), e7.
13. Interlandi, G., Wetzel, S.K., Settanni, G., Plückthun, A., Caflisch, A. (2007). Molecular dynamics simulations and chemical denaturation experiments. *J. Mol. Biol.* doi:10.1016/j.jmb.2007.09.042.
14. 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.
15. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L. & Wolberger, C. (1998). The structure of GABPalpha/beta: an ETS domain-ankyrin repeat heterodimer bound to DNA. *Science*, **279**, 1037–1041.
16. 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.
17. Binz, H. K., Kohl, A., Plückthun, A. & Grütter, M. G. (2006). Crystal structure of a consensus-designed ankyrin repeat protein: implications for stability. *Proteins: Struct. Funct. Genet.* **65**, 280–284.
18. Tadeo, X., Pons, M. & Millet, O. (2007). Influence of the Hofmeister anions on protein stability as studied by thermal denaturation and chemical shift perturbation. *Biochemistry*, **46**, 917–923.
19. Forrer, P., Binz, H. K., Stumpp, M. T. & Plückthun, A. (2004). Consensus design of repeat proteins. *ChemBioChem*, **5**, 183–189.
20. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.

21. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
22. Jaenicke, R. & Böhm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738–748.
23. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science*, **310**, 638–642.
24. Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T. *et al.* (2006). Protein stability and surface electrostatics: a charged relationship. *Biochemistry*, **45**, 2761–2766.
25. Robinson-Rechavi, M., Alibes, A. & Godzik, A. (2006). Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of *Thermotoga maritima*. *J. Mol. Biol.* **356**, 547–557.
26. Alsop, E., Silver, M. & Livesay, D. R. (2003). Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis. *Protein Eng.* **16**, 871–874.
27. Cheung, Y. Y., Lam, S. Y., Chu, W. K., Allen, M. D., Bycroft, M. & Wong, K. B. (2005). Crystal structure of a hyperthermophilic archaeal acylphosphatase from *Pyrococcus horikoshii*—structural insights into enzymatic catalysis, thermostability, and dimerization. *Biochemistry*, **44**, 4601–4611.
28. Corazza, A., Rosano, C., Pagano, K., Alverdi, V., Esposito, G., Capanni, C. *et al.* (2006). Structure, conformational stability, and enzymatic properties of acylphosphatase from the hyperthermophile *Sulfolobus solfataricus*. *Proteins: Struct. Funct. Genet.* **62**, 64–79.
29. Karlstrom, M., Steen, I. H., Madern, D., Fedoy, A. E., Birkeland, N. K. & Ladenstein, R. (2006). The crystal structure of a hyperthermostable subfamily II isocitrate dehydrogenase from *Thermotoga maritima*. *FEBS J.* **273**, 2851–2868.
30. Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. & Engel, P. C. (1998). Protein thermostability in extremophiles. *Biochimie*, **80**, 933–941.
31. Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**, 489–502.
32. 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: Struct. Funct. Genet.* **65**, 285–295.
33. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
34. Navaza, J., Panepucci, E. H. & Martin, C. (1998). On the use of strong Patterson function signals in many-body molecular replacement. *Acta Crystallogr. Sect. D*, **54**, 817–821.
35. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. Sect. D*, **55**, 247–255.
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. A*, **47**, 110–119.
37. 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.
38. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
39. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, CA, USA.