doi:10.1016/j.jmb.2007.11.046 J. Mol. Biol. (2008) 376, 241–257

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins

Svava K. Wetzel[1], Giovanni Settanni[2], Manca Kenig[1], H. Kaspar Binz[1] and Andreas Plückthun[1]*

[1]*Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*

[2]*MRC Centre for Protein Engineering, Hills Road, Cambridge CB2 0QH, UK*

Received 3 July 2007; received in revised form 3 October 2007; accepted 16 November 2007
Available online
22 November 2007

Full-consensus designed ankyrin repeat proteins were designed with one to six identical repeats flanked by capping repeats. These proteins express well in *Escherichia coli* as soluble monomers. Compared to our previously described designed ankyrin repeat protein library, randomized positions have now been fixed according to sequence statistics and structural considerations. Their stability increases with length and is even higher than that of library members, and those with more than three internal repeats are resistant to denaturation by boiling or guanidine hydrochloride. Full denaturation requires their heating in 5 M guanidine hydrochloride. The folding and unfolding kinetics of the proteins with up to three internal repeats were analyzed, as the other proteins could not be denatured. Folding is monophasic, with a rate that is nearly identical for all proteins ($\sim$400–800 s^{-1}), indicating that essentially the same transition state must be crossed, possibly the folding of a single repeat. In contrast, the unfolding rate decreases by a factor of about 10^4 with increasing repeat number, directly reflecting thermodynamic stability in these extraordinarily slow denaturation rates. The number of unfolding phases also increases with repeat number. We analyzed the folding thermodynamics and kinetics both by classical two-state and three-state cooperative models and by an Ising-like model, where repeats are considered as two-state folding units that can be stabilized by interacting with their folded nearest neighbors. This Ising model globally describes both equilibrium and kinetic data very well and allows for a detailed explanation of the ankyrin repeat protein folding mechanism.

Edited by F. Schmid

Keywords: protein folding; Ising model; ankyrin repeat proteins

Introduction

Repeat protein architecture does not rely on interactions between residues that are distant in sequence (long-range interactions), but stabilizing and structure-determining interactions are formed within a repeat and between neighboring repeats. This special feature, the modular nature of repeat proteins, makes them fundamentally different from globular proteins and, thus, interesting for testing experimental and theoretical views that have emerged from the study of globular proteins. Moreover, since repeat proteins are the only class of proteins that can be extended in size while still constituting a contiguous domain, unique questions about how folding and stability change with the number of repeats can be asked.

Repeat proteins constitute, next to immunoglobulins, the most abundant natural protein classes specialized in binding.[1,2] Ankyrin repeat (AR) proteins consist of repeating structural units (repeats) that stack together to form elongated nonglobular repeat domains. The AR is one of the most common

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

Present addresses: M. Kenig, Novartis Lek Pharmaceuticals, Kolodvorska 27, S1-1234 Menges, Slovenia; H.K. Binz, Molecular Partners AG, Grabenstrasse 11a, CH-8952 Zürich-Schlieren, Switzerland.

Abbreviations used: AR, ankyrin repeat; DARPin, designed ankyrin repeat protein; GdnHCl, guanidine hydrochloride; TPR, tetratricopeptide repeat; RCO, relative contact order.

protein sequence motifs. This 33-residue motif consists of a β-turn, followed by two antiparallel α-helices and a loop that reaches the turn of the next repeat[3] (see Fig. 1a).

Stability and kinetic folding studies of mostly natural AR proteins have been performed. The tumor-suppressor protein p16[6,7] unfolds in a sequential manner; first, both N-terminal repeats unfold, followed by the two C-terminal repeats. Molecular dynamics simulations have been carried out to study this in more detail.[8] The tumor-suppressor protein p19 shows an equilibrium intermediate, as well as three folding phases.[9] In a more detailed kinetic study, an on-pathway intermediate was detected, as well as a suggestion for its structure was made using NMR hydrogen/deuterium exchange.[10] Similarly as observed for the p16 protein, both N-terminal repeats 1 and 2 unfolded first, while repeats 3–5 were still folded. When dissecting the Notch receptor ankyrin domain from *Drosophila melanogaster*,[11,12] constructs from four to seven repeats were made, in which multiple repeats were deleted from either end or both ends, resulting in the finding that stability increased with repeat number. The longest construct has been used for kinetic folding studies,[13,14] and it was found that refolding and unfolding kinetics are best described by a sum of two exponential phases. Equilibrium and kinetic studies were also conducted with myotrophin, a small four-repeat AR protein.[15,16] While the kinetics of the Notch ankyrin domain could be fitted by a sequential three-state model, myotrophin kinetics were assigned to a two-state model. Further analysis with single and double mutants showed that myotrophin follows parallel pathways, where folding is initiated either by the C-terminal repeat or by the N-terminal repeat.[17]

All these studies showed that the folding of AR proteins is not simply a cooperative process, but intermediate states do occur. However, they have all been carried out with natural proteins containing repeats of different sequences and stabilities. Hence, many results only describe the particular protein under study, and they can only partially and qualitatively test the effects of protein length on kinetics and thermodynamics. In addition, they give no indication on the intrinsic properties of the consensus AR.

We therefore intended to examine the folding and unfolding of designed ankyrin repeat proteins (DARPins) with identical repeats as a function of repeat number. The consensus AR represents an "average structure" of all of the natural ARs and will eliminate properties that only come about with particular sequences of individual repeats. The DAR-

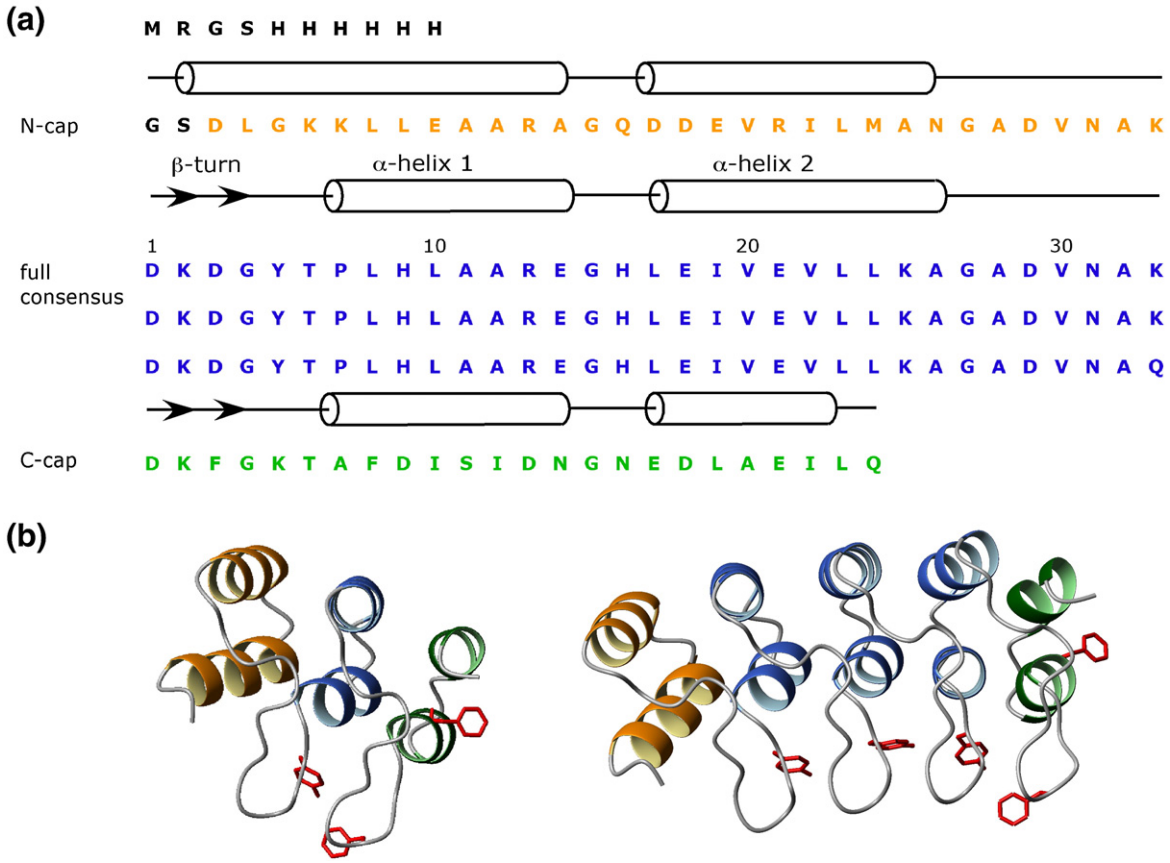

Fig. 1. (a) Amino acid sequence of NI_3C. The numbering of the internal consensus repeat is shown as used in the text. The N-terminal His tag, connected by a Gly-Ser linker, is shown in black. (b) Ribbon model representation (MOLMOL[4]) of NI_1C and NI_3C based on the structure of E3_5.[5] The colors are chosen to distinguish between the different repeats. The N-terminal repeat is depicted in orange, the internal consensus repeat is depicted in blue and the C-terminal repeat is depicted in green. The aromatic amino acids phenylalanine and tyrosine are represented as stick mode in red.

Pins can thus be considered as generalized examples for the study of AR protein folding. By characterizing the thermodynamic and kinetic parameters of these consensus proteins using circular dichroism (CD) and fluorescence spectroscopy, the dependence of stability, as well as of folding and unfolding rate constants, on repeat number was investigated. Moreover, we intended to gain mechanistic insight into the folding pathway of the three smallest proteins consisting of three to five AR repeats.

Results

Design of the consensus sequence

The "full-consensus" AR was based on a repeat module designed previously for a library of DARPins.[18] In the previous work, 7 out of 33 amino acids were allowed to vary in order to bind to target molecules. While the 26 fixed residues of the library repeat module were used without changes in the present study, defined residues had to be assigned to the six randomized potential interaction residues (positions 2, 3, 5, 13, 14 and 33) (Fig. 1a) and to the remaining randomized framework residue (position 26). For this purpose, we used the consensus analyses of sequence and structure described previously.[18] Positions 2 and 33 were defined as lysines, since this was the most frequent amino acid in these positions in our alignment (position 2: Lys, 19%; Arg, 13%; Asn, 13%; Ser, 12%; position 33: Lys, 21%; Arg, 16%; Gln, 16%) and since both were found in unselected library members, indicating their suitability to support the fold.[18] Position 3 was defined as aspartate, since it is one of the most frequent amino acids in this position in our alignment (Asn, 22%; Asp, 19%; Glu, 15%). It was preferred to Asn, as the negative charge might better compensate for the high number of positive charges introduced by the two lysines at positions 2 and 33. Position 5 was defined as tyrosine, in order to introduce one chromophore per consensus repeat for fluorescence measurements and since tyrosine and other large residues are most frequently found in this position in our alignment (Leu, 16%; Tyr, 14%; Phe, 13%; Arg, 12%; Trp and Asn, 9%). We chose glutamate for position 14 due to its most frequent occurrence (Glu, 22%; Asn, 13%). Even though arginine is not the most frequent amino acid in our alignment at position 13 (Gln, 22%; Lys, 19%; Arg, 18%), positively charged residues, when taken together, are most abundant. The first α-helices of each AR are in close proximity in juxtaposed repeats, and thus arginine should compensate for the negative charge of position 14. In addition, arginine has a high α-helical propensity.[19] The framework position 26 was defined as alanine, since this is the most frequently found residue in this position (40%) and since it could successfully be used repetitively for a previous full-consensus design.[20] The resulting consensus sequence is shown in Fig. 1a.

When comparing the full-consensus AR proteins investigated here with those described earlier,[20] the most important difference is the presence of capping repeats, which are identical with those we have published earlier[18] for our library. These appear to be essential for correct folding *in vivo* and full reversibility of folding *in vitro* at neutral pH.[21] Additionally, our sequence of the "internal" repeats is different in several positions from that of Mosavi *et al.*[20] In our numbering scheme[3,18] (cf. Fig. 1a), we start with the β-turn; thus, our position 3 corresponds to position 1 of the sequence from Mosavi *et al.* (see Supplementary Fig. 1).[20] At position 3, we have introduced Asp for reasons described above, while Mosavi *et al.* used Asn. Our chromophoric Tyr at position 5 was Arg in their work; at position 14, we used Glu instead of Asn for reasons described above. Four additional differences are in helix 2, which are "framework" positions already defined previously.[18] In position 19, Ile had been chosen (instead of Val) in our work because of its higher helical propensity; in position 21, Glu had been chosen (instead of Lys) because of its occurrence in the structure of GABPβ1; in position 22, Val had been chosen (instead of Leu) for the same reason; and, in position 25, Lys had been chosen (instead of Glu) for its opposite charge to Glu21.

The resulting protein designed here shows an almost perfectly alternating pattern of surface charges, as shown in the crystal structure of NI_3C,[22] and may help to partially explain the unusually high stability of these full-consensus AR proteins.

Construction, expression and characterization of full-consensus DARPins

The full-length proteins were termed NI_1C–NI_6C, where I represents the full-consensus repeat, the subscript represents the number of identical full-consensus repeat modules, and N and C correspond to the N- and C-terminal capping repeats (resulting in proteins with three to eight repeats in total). The capping repeats differ slightly from the full-consensus repeat in size and sequence (Fig. 1a) and have a hydrophilic surface. The C-capping repeat has been examined in detail.[21] The protein sequence of NI_3C and the model structure of NI_1C and NI_3C are shown in Fig. 1.

The six proteins were expressed in *Escherichia coli* in soluble form. Purification was done by immobilized metal-ion affinity chromatography. Gel filtration with multiangle light scattering [NI_1C: measured molecular mass, 12 kDa (expected molecular mass, 10.8 kDa); NI_2C: 14.9 kDa (14.4 kDa); NI_3C: 17.4 kDa (17.8 kDa); NI_4C: 22.4 kDa (21.4 kDa); NI_5C: 25.8 kDa (25 kDa); NI_6C: 27.8 kDa (28.5 kDa)] showed that the proteins were monomeric and had the correct size. SDS-PAGE analysis confirmed their purity, and mass spectrometry (data not shown) confirmed their correct composition and lack of degradation. The CD spectra of the six proteins are superimposable on the spectra of natural and other designed ankyrin proteins (data not shown).

Guanidine-hydrochloride-induced equilibrium unfolding

The denaturation of the six full-consensus DARPins NI_1C–NI_6C was monitored using both CD at 222 nm and tyrosine fluorescence. The stabilities increase with increasing repeat number, and this is observed both in guanidine hydrochloride (GdnHCl)-induced unfolding (Fig. 2) and in temperature-induced unfolding (Fig. 3). In GdnHCl-induced unfolding, we observe that only NI_1C, NI_2C and NI_3C can be unfolded (defined by the disappearance of the CD signal at 222 nm), while the longer proteins (NI_4C–NI_6C) can only partially be unfolded. NI_4C reaches a major, but not completed, transition (loss of about 50% of the initial signal) at 8 M GdnHCl. All proteins from NI_3C to NI_6C show a small transition at around 4 M GdnHCl (see below).

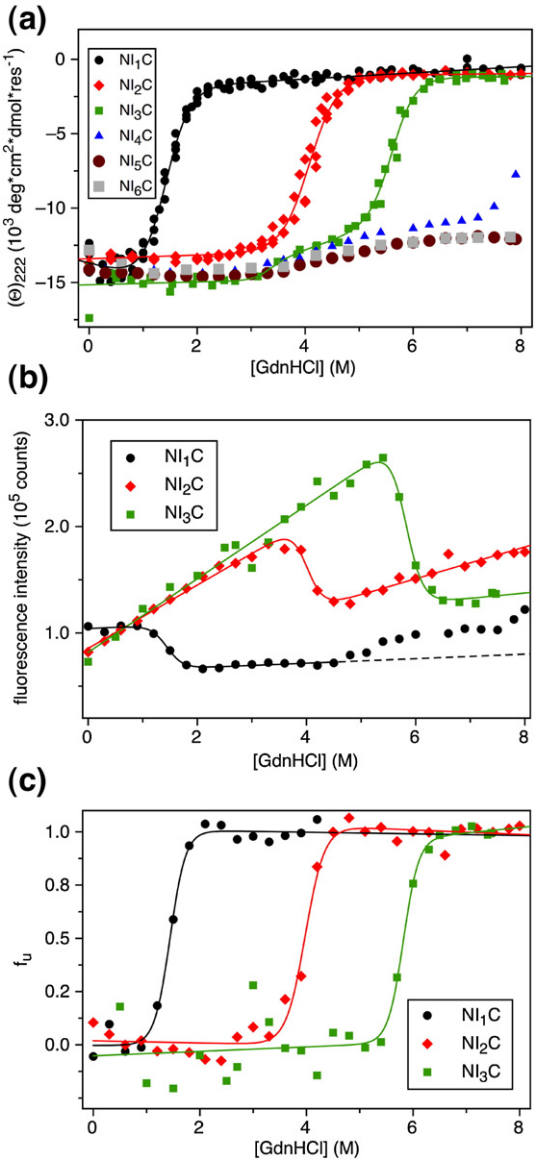

Fig. 2. GdnHCl-induced equilibrium unfolding of the six proteins NI_1C–NI_6C at 20 °C followed by (a) CD spectroscopy and (b and c) tyrosine fluorescence (b) plotted with fluorescence intensity units and (c) converted to fraction unfolded f_U (see Materials and Methods). The line represents the least-squares fit to a two-state model (except NI_3C, whose CD signal was fitted to a three-state model). The parameters of the fits are summarized in Table 1. The protein concentration was 10 μM.

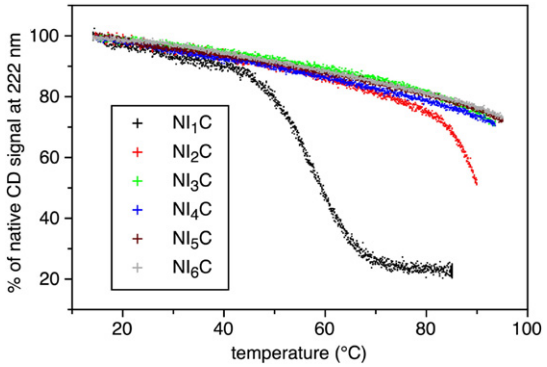

Fig. 3. Thermal melting of NI_1C–NI_6C followed by CD spectroscopy between 15 °C and 95 °C (see Materials and Methods). The derived midpoints of denaturation (melting temperature) are summarized in Table 1. The protein concentration was 10 μM. Melting was only 70% reversible for all proteins.

The CD data, which show a single transition midpoint for NI_1C and NI_2C, can be described by a two-state model (Eq. (1)), which, of course, does not prove a fully cooperative system (see below). NI_3C, for which two transitions are visible, is best described by a sequential three-state model (Eqs. (3)–(9)). The first transition at 3.7 M GdnHCl would correspond to $N \rightleftharpoons I$, and the second transition $I \rightleftharpoons U$ occurs at 5.6 M. This second CD transition coincides with the single fluorescence transition (Fig. 2b and c). We interpret this equilibrium intermediate as one in which the C-terminal capping repeat has unfolded[21] (see below). It appears that NI_4C, NI_5C and NI_6C also show this intermediate but do not denature further.

We also measured the GdnHCl-induced transitions by tyrosine fluorescence. Although the tyrosine residues are not in the hydrophobic core and the nature of the change in fluorescence signal during the unfolding transition is not completely clear, we could observe unfolding transitions with tyrosine fluorescence. The unfolding curves show a single transition (Fig. 2b and c) to lower fluorescence at about the same GdnHCl concentration as the main CD transition and a steep slope of the pre-transition and posttransition baselines (Fig. 2b) whose origin is currently not clear.

The thermodynamic parameters calculated for the only three proteins that can be fully unfolded, which are summarized in Table 1, were analyzed by a classical cooperative folding model. The m value measured for the almost globular NI_1C ($m = 2.6 \pm 0.2$ kcal mol^{-1} M^{-1}; Table 1) is consistent with the expectations derived from the buried surfaces[23] ($m_{\Delta ASA} = 2.7$ kcal mol^{-1} M^{-1}), while the m value determined for NI_2C is much smaller than would be expected for a typical globular protein of this size.

Table 1. Thermodynamic parameters for GdnHCl-induced and thermal unfolding of the proteins NI_1C, NI_2C and NI_3C

			Equilibrium data[b]		Kinetic data[c]	
Protein	T_m [°C]	D_m [M][a]	m [kcal mol^{-1} M^{-1}]	ΔG_0 [kcal mol^{-1}]	m [kcal mol^{-1} M^{-1}]	ΔG_0 [kcal mol^{-1}]
NI_1C	60	1.4	2.6±0.2	3.7±0.3	4.0±0.4	5.4±0.6
NI_2C	90	4.1	2.3±0.2	9.2±0.7	4.4±1.0	16.7±3.9
NI_3C[d]	>100	3.7/5.6	1.7±0.9/3.0±0.6	19.7±4.6	–	–
NI_4C	>100	≈8	–	–	–	–
NI_5C	>100	>8	–	–	–	–
NI_6C	>100	>8	–	–	–	–

[a] Denaturation midpoint (M GdnHCl) at 20 °C.
[b] Values obtained by two-state (NI_1C and NI_2C) or three-state (NI_3C) fitting.
[c] Values obtained from kinetic three-state fit using $K^0=((k^0_{NI}k^0_{IU})/(k^0_{UI}k^0_{IN}))$ and $m=(-m_{UI}+m_{IU}-m_{IN}+m_{NI})$.
[d] Values for both transitions are given (for details, see the text).

This already suggests that, despite the fact that only one transition can be seen, the protein is not well described by a fully cooperative two-state model as typical for small proteins.

In a separate study,[21] we have shown experimentally that the small transition of NI_3C at about 4 M GdnHCl is consistent with a selective denaturation of the C-terminal capping repeat, whose different sequence, shorter C-terminal helix and fewer inter-repeat interactions may explain this phenomenon. Molecular dynamics calculations are also consistent with this interpretation.[21,24] In a different study, where consensus ARs were introduced between repeats 5 and 6 of the Notch AR domain, an equilibrium intermediate has been found as well. This intermediate disappeared when removing the C-terminal repeats 6 and 7.[25]

For the bigger proteins, evidence for some soluble aggregate formation was found at the highest GdnHCl concentration. This could be detected by measuring light scattering with a fluorimeter by recording the emission of light at 360 nm as a function of GdnHCl. Some onset of scattering was observed for NI_4C, NI_5C and NI_6C at 6–7 M GdnHCl, while no detectable aggregates were formed for NI_3C under these conditions (data not shown). NI_3C appeared to be fully denatured under these conditions (see Fig. 2a), and it is thus not the denatured species that partially aggregates. Taken together with the fact that the CD signal hardly changes for the larger molecules, this suggests that the unfolding of the C-terminal capping repeat no longer protects the molecules from some form of soluble aggregate formation at very high GdnHCl concentrations and that essentially native-like species (devoid of the solubilizing C-terminal capping repeat) start to associate under these conditions, somewhat reminiscent of a salting-out effect of native proteins. It is possible that these native-like species stabilize each other during this process (see below). Note that the design of even more stable C-terminal capping repeats has been successfully undertaken.[21]

Ising model describing the stability of full-consensus DARPins

Besides classical cooperative two-state or three-state folding models as described above, we have also analyzed the folding of the full-consensus DARPins by an Ising model. Briefly, this model does not assume a cooperative folding of the whole protein, but considers every repeat as an individual folding unit. The free energy of each repeat is assumed to be linearly dependent on denaturant concentration, characterized by an m value, while the interaction energy between neighboring repeats is considered to be constant, as long as both are folded. Similar models have been proposed previously for the study of repeat proteins.[11,26]

Because of the different stabilities of the capping repeats, we had to use a model that includes them as a separate unit with different ΔG_0 and m values. Even though we had found experimental evidence, triggered by molecular dynamics simulations,[21,24] that it is the C-terminal capping repeat that denatures at the lowest denaturant concentrations, we do not know whether the N-terminal capping repeat is also of somewhat lower stability than the internal repeats. To keep the model as simple as possible, we therefore considered both N-terminal and C-terminal capping repeats as having the same ΔG_0, which is lower than that of the internal repeats. This was done as the fits were significantly better (data not shown) than treating only the C-terminal capping repeats separately.

The Ising model, as used here, thus contains five parameters (see Eqs. (18)–(21) in Materials and Methods) describing all repeat proteins: the coupling energy J between repeats, the free energy of an isolated repeat ΔG_0, its denaturant dependence m, the free energy of isolated capping repeats $\Delta G_0'$ and their denaturant dependence m'. To determine these parameters, the equilibrium denaturation data of NI_1C, NI_2C and NI_3C, as measured by CD (Fig. 2), were globally fitted to Eq. (21), as shown in Fig. 4. This approach assumes that the CD signal is proportional to the number of folded repeats across the whole population, rather than, as in a two-state model, to the percentage of molecules with all repeats folded (for details, see Materials and Methods). The determined parameters of the Ising model are summarized in Table 2.

According to the Ising model, besides the completely folded and completely unfolded configurations, partially folded states are also populated. Figure 5a summarizes the most important possible

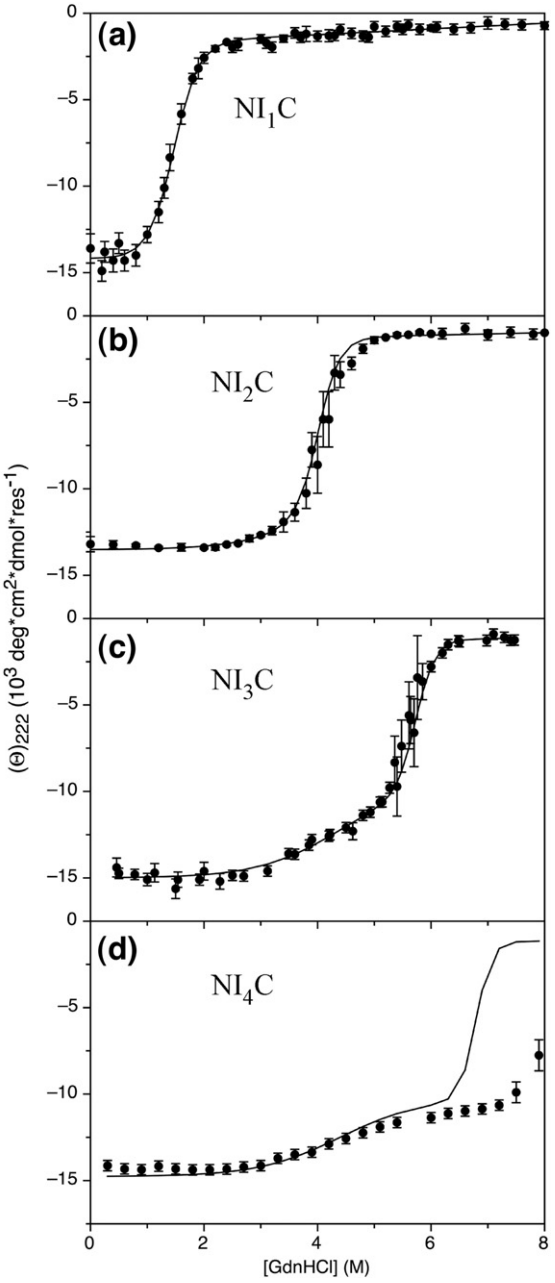

Fig. 4. Ising model fit of NI_1C, NI_2C and NI_3C data, and prediction of NI_4C. Equilibrium GdnHCl-induced unfolding of the three proteins NI_1C (a), NI_2C (b) and NI_3C (c) at 20 °C followed by CD spectroscopy. The line represents the fit to the Ising model (see Materials and Methods). The parameters are summarized in Table 2. (d) Prediction using the Ising model fit and experimental data for NI_4C.

states for NI_1C, NI_2C and NI_3C, where one to four repeats are folded. Note that any folded repeats will always be adjacent, as the population of other states is insignificant. In addition, a lower free energy is obtained if the folded states are all internal than if they include a capping repeat. The free energy of the states of the protein as a function of the GdnHCl concentration [D] (Fig. 5b–d) provides a coherent interpretation of the experimental data. According to this model, the configurations with only one repeat folded are those maximally unstable at any denaturant concentration, as the free energy of folding only one repeat is always unfavorable, and no stabilization from interaction with the neighbors can be provided. Indeed, isolated single ARs were experimentally found not to be stable.[20] This state with exactly one repeat folded thereby separates the two regions of minimal free energy, corresponding to the unfolded states on one side and the folded states on the other side (Fig. 5b).

At zero (or low) denaturant, the most stable state is, as expected, the one in which all repeats are folded. At high denaturant, for the three proteins shown here, the most stable state is the one in which all repeats are denatured, also as expected. However, at high denaturant concentration, the most stable among the native states are those with one and/or both unfolded terminal repeats but folded internal repeats. Thus, under highly denaturing conditions, these states become more stable than the completely folded configuration for all the DARPins examined with this model (Fig. 5c and d). In other words, the model predicts a denaturant-dependent shift of the most stable state in the native free-energy basin. This behavior results in an overall nonlinear dependence of the stability $\Delta G_L(D)$ on [D], that is, the free-energy difference between the most stable conformer in the native basin and the unfolded state (Fig. 6). The kinks observed in $\Delta G_L(D)$ are due to the abovementioned shift in the native state of the protein as [D] increases.

With the parameters determined from the fit of the equilibrium denaturation data measured by CD for NI_1C–NI_3C, it is possible to extrapolate the Ising model to describe the behavior of the larger proteins NI_4C–NI_6C (Figs. 4d and 6), which can no longer be experimentally unfolded (except by heating in high GdnHCl; Fig. 7). The slope of the stabilities $\Delta G_L(D)$ in proximity to the predicted transition midpoint (intersection with the x-axis of Fig. 6) provides m_L values (Table 3). We report these to allow a comparison with standard two-state fits of the data (even if these are only possible for NI_1C and NI_2C) (Table 1). In Table 3, predicted transition midpoints and stabilities at 0 M GdnHCl are also reported. In the Ising model, the stability at 0 M GdnHCl depends linearly on the number of repeats L. In the homogeneous Ising model (all repeats are identical; not shown here), the m_L value is also linear with the number of repeats. Because of the presence of terminal repeats of different intrinsic stabilities in the model presented here, m_L tends to be linear only for large L ($L \geq 5$), where the relative influence of the terminal repeats becomes smaller.

Table 2. Parameters obtained by fitting the equilibrium data of NI_1C, NI_2C and NI_3C to the Ising model (see Fig. 4)

J	-14.2 ± 0.7 kcal mol^{-1}
ΔG_0	3.3 ± 0.2 kcal mol^{-1}
m	1.1 ± 0.1 kcal mol^{-1} M^{-1}
$\Delta G_0'$	10.6 ± 0.6 kcal mol^{-1}
m'	0.83 ± 0.04 kcal mol^{-1} M^{-1}

A direct comparison of model predictions with experimental data for NI_4C (Fig. 4d) and larger constructs (Table 3) shows that the model apparently underestimates the stability of these proteins. This led us to propose a possible hypothesis for the discrepancy. The long constructs, at high denaturant concentration, have unfolded terminal repeats, according to the Ising model, but are otherwise still essentially folded and may be stacking together due to the now-exposed hydrophobic surfaces at the edge of the consensus repeats. These elongated complexes are more stable than the monomeric molecule because of their larger effective length. As reported above, light-scattering experiments confirmed that the DARPins NI_4C–NI_6C begin to form soluble aggregates at high denaturant concentration. The presence of such soluble, essentially native-like, aggregates might explain the discrepancy between the predictions from the Ising model and the experimental data for the large constructs. Further experiments outside the scope of the present work will be needed to confirm the exact nature of the species giving rise to increased light scattering at very high GdnHCl concentrations.

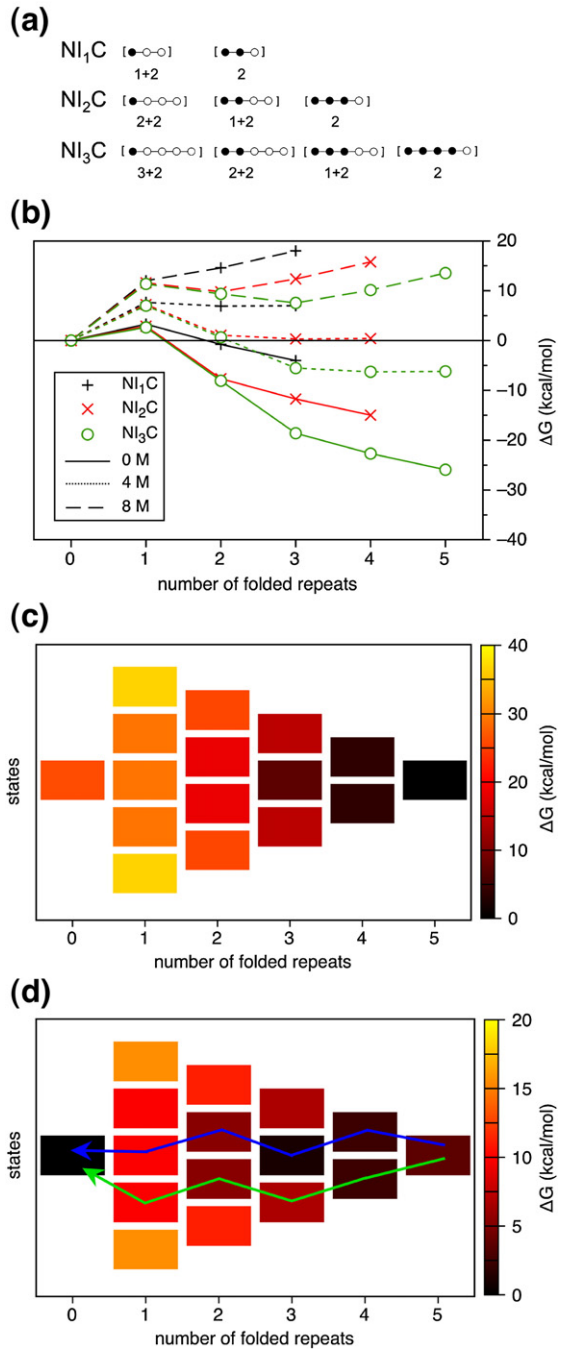

Fig. 5. Possible states and average free energy of the states of NI_1C, NI_2C and NI_3C. (a) Cartoon of possible configurations in NI_1C, NI_2C and NI_3C, according to the Ising model. Folded (●) and unfolded repeats (○) are displayed with different symbols. For states that contain more than one folded repeat, only those in which folded repeats are adjacent to each other are shown (and counted here). For example, for NI_1C, the state (●-○-●) is hardly populated, such that it can be neglected next to (○-●-●) and (●-●-○), explaining the number "2." Since the folded internal repeats have a different energy than the folded terminal ones, their respective numbers are separated by a (+) sign. For example, in NI_1C, there is only one state with a folded internal repeat (○-●-○), while there are two with terminal repeats (●-○-○) and (○-○-●), explaining the numbers "1+2." (b) The free energies of NI_1C, NI_2C and NI_3C as a function of the number of folded repeats plotted at three different GdnHCl concentrations. Free energies obtained at 0 M GdnHCl (solid lines), 4 M GdnHCl (short dashed lines) and 8 M GdnHCl (long dashed lines) are plotted. (c and d) Free energy of NI_3C considering only states with contiguous folded repeats at 0 M GdnHCl (c) and 6 M GdnHCl (d). Each box represents a state, and its color indicates its free energy at the GdnHCl concentration of the plot. States with the same number of contiguous folded repeats are arranged vertically in the figure and ordered, from bottom to top, such that the configurations with folded repeats that include the N-terminal repeat are represented by the lower boxes, those with a C-terminal folded repeat are represented by the upper boxes and those with only consensus (middle) repeats folded are represented by the middle boxes. At 0 M GdnHCl (c), as expected, the state with all the repeats folded is the most stable, while at 6 M (d), the most stable state is the completely unfolded one, but another one that is almost as stable coexists in the native free-energy basin, which has three central repeats folded and the terminal repeats unfolded. The blue pathway contributes to the slowest phase, while the green pathway may give rise to a faster phase. All of the pathways have to progressively cross the diagram from right to left, losing a single (i.e., not more than one) folded repeat at every step.

Thermal stability

The thermal unfolding of all six proteins was monitored by the change in CD signal at 222 nm (Fig. 3). The NI_1C transition is sigmoidal with a midpoint $T_m = 60$ °C. Melting of NI_2C was not complete, but T_m can be estimated to be 90 °C (Table 1). The bigger proteins, NI_3C–NI_6C, could not be thermally unfolded at all. Melting was only possible by heating them in 5 M GdnHCl or by heating them in a buffer of pH 3.5 (data not shown). Only by using these unusually strong denaturing conditions could the proteins NI_3C, NI_4C, NI_5C and NI_6C be unfolded (Fig. 7). For NI_3C and NI_4C, the pretransition

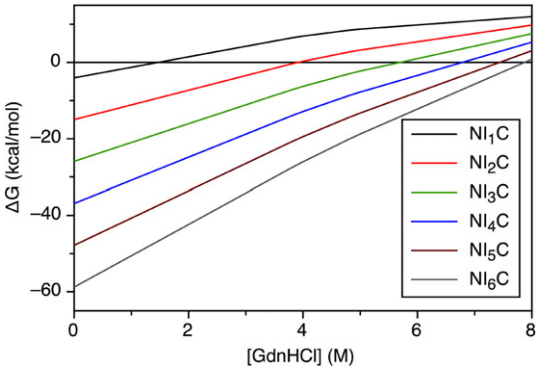

Fig. 6. Stability as a function of GdnHCl for the six consensus DARPins of different lengths, as obtained by the Ising model. Stability is defined as the energy difference between the lowest energy state among the folded ones (at the corresponding denaturant concentration) and the lowest energy state among the unfolded ones, which is the state with all repeats unfolded. This quantity plotted equals the difference between the activation free energy of unfolding and the activation free energy of folding, as determined from the barrier heights in Fig. 5b.

baselines start at values that are significantly below the value of 100% native protein, as these proteins are already partially denatured in 5 M GdnHCl at room temperature (see also Fig. 2a). The transition midpoints T_m in 5 M GdnHCl were approximately 56 °C for NI_3C, 76 °C for NI_4C, 88 °C for NI_5C and approximately 96 °C for NI_6C. The melting was only 70% reversible, as judged from the recovery of the CD signal after cooling, and the thermal denaturation could thus not be used to derive ΔG values. This experiment underlines the extraordinary stability of the proteins. We believe that the consensus design,[18] which has "idealized" the protein structure by introducing many critical interactions in every single repeat, combined with the choice of the residues in the previously randomized positions (see above), which lead to a "checkerboard" arrangement of charges,[22] with maximization of attractive interactions, contributes to this high stability.

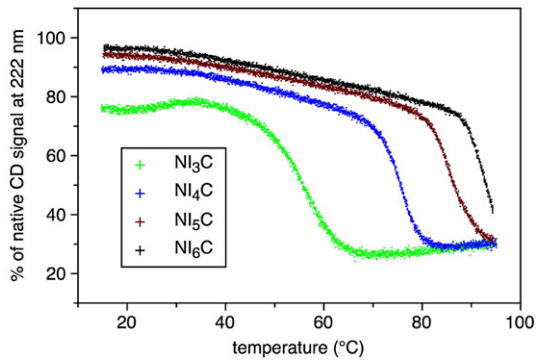

Fig. 7. Thermal melting of NI_3C, NI_4C, NI_5C and NI_6C in the presence of 5 M GdnHCl. The proteins NI_4C–NI_6C can only be denatured by elevated temperatures in 5 M GdnHCl. The stabilities of NI_4C, NI_5C and NI_6C can be distinguished qualitatively, but due to nonreversibility, no thermodynamic parameters can be calculated.

Table 3. Determined and extrapolated $D_{m,L}$, m_L and ΔG_L^0 values for the DARPins investigated in the Ising model

DARPins	$D_{m,L}$ [M][a]	$m_L(D_{m,L})$ [kcal mol^{-1} M^{-1}][b]	ΔG_L^0 [kcal mol^{-1}][c]
NI_1C[d]	1.47±0.08	2.7±0.2	−4.03±0.54
NI_2C[d]	3.96±0.07	2.8±0.2	−15±0.8
NI_3C[d]	5.56±0.18	1.9±0.8	−25.9±1.1
NI_4C[e]	6.65±0.1	2.4±0.7	−36.9±1.4
NI_5C[e]	7.35±0.08	3.3±0.95	−47.8±1.6
NI_6C[e]	7.8±0.08	4.3±0.6	−58.8±1.9

[a] Denaturation midpoint (M GdnHCl).
[b] Slope of ΔG_L at the transition midpoint.
[c] ΔG_L at 0 M GdnHCl.
[d] Experimentally determined.
[e] Extrapolated as described in the text.

The experiment in Fig. 7 also shows that the distances between the T_m values of NI_3C–NI_6C decrease with increasing repeat number. A dependence of stability on length has also been observed for other consensus repeat proteins[25,27] and dissected versions of the natural *Drosophila* Notch receptor.[11] However, such an enormous stability increase has not been observed yet.

Unfolding and refolding kinetics of NI_1C, NI_2C and NI_3C

Since only the shorter AR consensus proteins could be denatured at room temperature, the folding and unfolding kinetics of the DARPins NI_1C, NI_2C and NI_3C were studied as a function of denaturant concentration and are represented as chevron plots. The folding and unfolding rates were determined by monitoring the changes in ellipticity at 225 nm after dilution out of, or into, GdnHCl at 20 °C using a stopped-flow instrument. Slow reaction phases (>100 s) were recorded by manual mixing.

It is immediately apparent that the folding rates are equally fast for all proteins, while the unfolding rates are slow and get slower with increasing numbers of repeats (Fig. 8a).

All kinetic traces of NI_1C (folding and unfolding) could be fitted to a single exponential function (Eq. (10)). While NI_2C refolding was monophasic, NI_2C unfolding was fitted to a double exponential function (Eq. (11)) representing two unfolding phases, with amplitudes decreasing from 70% to 10% with GdnHCl concentration for the slow phase and increasing from 30% to 90% for the fast phase. Both unfolding phases were recorded after fast mixing, and the slower unfolding phase was also confirmed by manual mixing. NI_3C refolding was monophasic, too, while the unfolding reaction could be divided into three phases (triple exponential). The slowest unfolding phase was recorded by manual mixing (60% of the total amplitude), while both fast unfolding phases were recorded by fast mixing (together 40% of the total amplitude: 30% for the faster phase and 10% for the slower phase). Representative kinetic traces of NI_2C and NI_3C are shown in Supplementary Fig. 2. In addition to the

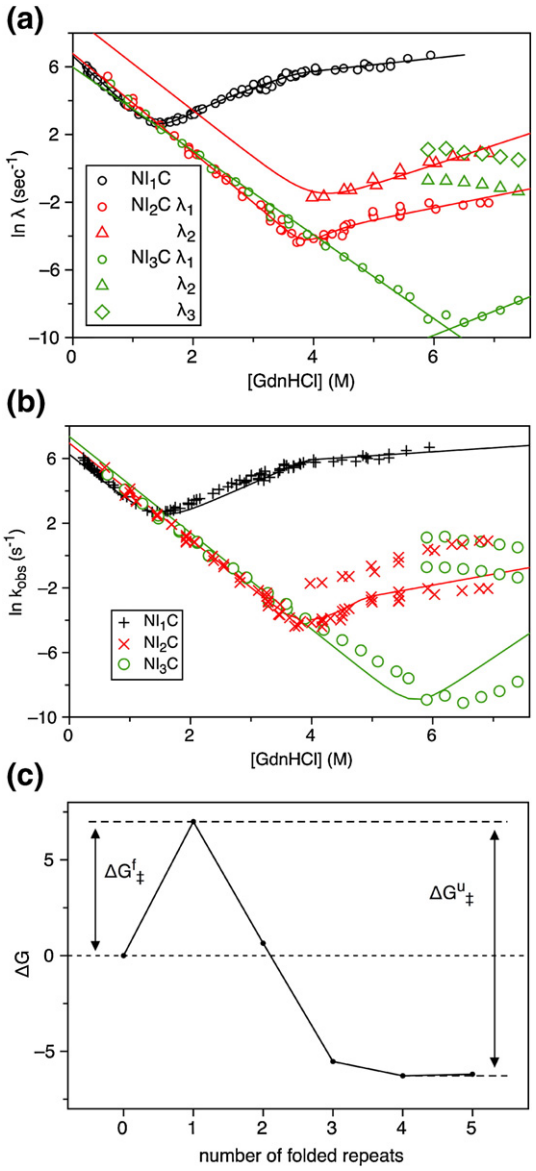

Fig. 8. Chevron plots of NI_1C, NI_2C and NI_3C, and prediction of the kinetics by the Ising model. (a) Experimental chevron plots of NI_1C, NI_2C and NI_3C at 20 °C, followed by CD spectroscopy. The left limb corresponds to the GdnHCl-dependent rate constants for refolding, and the right limbs correspond to the unfolding rate constants. The lines represent the best fit to the data using the kinetic on-pathway intermediate three-state model.[28] The rate constants are summarized in Table 4. (b) Theoretical chevron plots of NI_1C, NI_2C and NI_3C (lines) obtained by using the height of the folding/unfolding barriers from Fig. 5b to estimate the folding and unfolding rates and their dependence on GdnHCl and protein length. The preexponential factor a has been chosen as $a = 1.3 \times 10^5\ s^{-1} \exp(-1.15\ [D])$ to best match the experimental data in the chevron plot [superimposed in the figure as symbols; same data and symbols as in (a)]. (c) Graphical explanation of the procedure used to obtain the height of the folding or unfolding barriers using the data from Fig. 5b. We show, as an example, the NI_3C free-energy profile at 4 M GdnHCl, where the activation free energy for folding is 7 kcal mol^{-1}. Substituting this number into the formula for the rate $k = a \exp(-\Delta G_‡/kT)$ with the preexponential factor a at 4 M GdnHCl, we obtain the rate that was then plotted in (b) for NI_3C at 4 M GdnHCl.

CD data discussed above, folding kinetics have also been measured by tyrosine fluorescence (data not shown). While folding rates were found to be very similar, the fastest unfolding phase of NI_3C was 3-fold faster, suggesting that, in this case, loss of tertiary structure precedes loss of secondary structure. In the unfolding of NI_2C, however, all rates measured by CD and tyrosine fluorescence were found to be very similar.

Plotting the fitted rates as a function of denaturant concentration revealed clear deviations from the classical V-shaped two-state chevron plot for all proteins. NI_1C unfolding and refolding are monophasic, but the unfolding limb exhibits a curvature (Fig. 8a). This curvature is even more pronounced at a lower temperature (5 °C; data not shown). One way to explain this behavior in a classical framework is to assume a sequential three-state model with a metastable high-energy intermediate.[29] In this model, the metastable intermediate is not populated enough to cause a second observable folding or unfolding phase, but it does influence the denaturant dependence of the unfolding rate. Such a behavior can be seen in the nonlinear slope of the unfolding limb of the NI_1C chevron plot. Another curved chevron plot has also been found for the ankyrin protein myotrophin[16] that consists of one repeat more than the DARPin NI_1C.

NI_2C kinetics are monophasic in the refolding limb, while the unfolding reaction is best described by a double exponential (Supplementary Fig. 2). The NI_2C unfolding limb of the slower unfolding phase shows a curvature, too. The NI_2C chevron plot could be well fitted by a sequential three-state model (Eq. (12)), where the unfolding intermediate state is more populated than for NI_1C and appears to be formed before reaching the fully unfolded state (on-pathway intermediate). Therefore, a second kinetic unfolding phase is observed (Fig. 8a).

Two unfolding phases were also found for the Notch ankyrin domain with seven ARs.[13] However, the Notch ankyrin domain also reveals a second refolding phase, which has been attributed to prolyl isomerization in the unfolded state, and that protein is much less stable and folds much more slowly than the DARPin NI_2C, which is three repeats shorter and has only a single folding phase. The Notch domain Nank1–7Δ has four *trans*-prolines and one disordered proline, while the tumor-suppressor protein p19 has even seven *trans*-prolines. In contrast, NI_1C has one *trans*-proline, NI_2C has two *trans*-prolines and NI_3C has only three *trans*-prolines. The bigger number of proline residues in the Notch domains might explain why slow folding phases were detectable, while they were not detectable in our case.

The kinetics of NI_3C are more complex: the refolding reaction is again monophasic as observed for NI_1C and NI_2C, while unfolding was separated into three phases (two phases monitored by stopped-flow mixing, one phase observed by manual mixing). A three-state model is not able to describe such complex kinetics, and models with more than three states are very difficult to apply

Table 4. Unfolding and refolding rates of NI_1C, NI_2C and NI_3C

DARPins	k^0_{UI}	k^0_{IU}	k^0_{IN}	k^0_{NI}	k^0_u	k^0_f
NI_1C	776±43	3.6±1.4	3589±3417	66±17	3.4	638
NI_2C	892±103	$1.1\times10^{-3}\pm7\times10^{-4}$	$6961\pm4.6\times10^{4}$	$1.4\times10^{-3}\pm7\times10^{-4}$	6.3×10^{-4}	790
NI_3C	n.d.	n.d.	n.d.	n.d.	1×10^{-8}	400

The rates (s^{-1}) at 20 °C, extrapolated to native buffer, were obtained by fitting a sequential three-state model with on-pathway intermediate.[28] k^0_f (Eq. (16)) and k^0_u (Eq. (17)) of NI_1C and NI_2C were calculated from the fitting parameters k^0_{UI}, k^0_{IU}, k^0_{IN} and k^0_{NI}.

quantitatively as they exhibit too many parameters, making a converging fit very difficult.

The kinetic data thus indicate a more complex model than the two-state model for NI_1C and NI_2C. In contrast, the equilibrium data can be fitted well to a two-state model, and there was thus no hint for a deviation from that fit alone. We therefore also report the ΔG_0 and m values calculated from kinetics (Table 1) using the four rate constants k_{ij} according to the sequential three-state model. A global fit to both kinetic and thermodynamic data of NI_1C and NI_2C using the three-state model (with metastable and stable intermediates, respectively), which fits the data well, gives values rather similar (data not shown) to the ΔG_0 values derived from kinetics alone (Table 1). We assume that this similarity is largely an effect of the kinetics dominating the parameter fit, since equilibrium measurements offer looser constraints on the parameters of the model than kinetic data because of the additional degrees of freedom provided by the choice of the baselines. However, even the kinetics themselves do not fully determine all the parameters of the three-state model, as some of them are also rather insensitive to the global fit, as can be seen by the size of some of the error margins in Table 4. In summary, when incorporating all observations, these proteins are not consistent with a two-state system. Instead, their folding behavior is consistent with three-state models, even if some of their parameters remain rather poorly constrained, resulting in significant error margins.

The kinetic parameters of NI_1C, NI_2C and NI_3C are summarized in Table 4. The first important observation is that the folding rates do not vary much between all three proteins. Folding is monophasic, and the folding process thus needs to cross essentially the same transition state, independent of the number of repeats in the protein. This is intuitively appealing, as it suggests that a very small number of repeats, in the simplest case one, must be folded in the transition state, which is common to all proteins investigated here.

In sharp contrast, the unfolding rates decrease tremendously (a factor of 10^4 from NI_1C to NI_2C and from NI_2C to NI_3C) with increasing repeat number. In order to compare the folding and unfolding rates of the three proteins, the refolding rate and the slowest unfolding rate of NI_3C were used. With the longer proteins, this unfolding rate could not be measured, as they do not fully denature at all under these conditions, as explained above. If the slowest unfolding rates are considered, unfolding rates of 0.6 s^{-1}, 1.3×10^{-8} s^{-1} and 1.2×10^{-16} s^{-1} (corresponding to a half-life of 1.8×10^8 years) are extrapolated using the Ising model (see Folding and Unfolding Pathways According to the Ising Model) to zero denaturant for NI_1C, NI_2C and NI_3C, respectively. The extrapolated values for NI_4C, NI_5C and NI_6C would even be predicted to be a factor of 10^8 slower for each repeat added, but, of course, we have no possibility to test the validity of these extrapolations at this point. The stability is thus directly reflected in their (predicted) half-lives in native buffer. As the folding rate is certainly not slow and does not vary between the three proteins, the very slow denaturation rates are thus a direct reflection of the very low free energy of the folded state, which raises the barrier to unfolding correspondingly.

Folding and unfolding pathways according to the Ising model

In experiments probing the kinetics of folding and unfolding, the proteins are expected to follow the minimum free-energy pathways, connecting the initial state with the equilibrium state under the new conditions (i.e., under different denaturant concentrations). The Ising model allows for the reduction of a complex protein folding problem into a simple reaction occurring between a finite set of states. In this simplified conformational space, the slowest relaxation rate is determined by the largest free-energy barrier found along the minimum free-energy pathway connecting the initial state and the final state. Given a number of folded repeats, the Ising model assigns the minimum free energy (i.e., the maximal probability of being occupied as computed by Eq. (19)) to the state where the folded repeats are contiguous. Assuming that the repeats fold and unfold one at a time, according to the Ising model, the largest barrier between the fully folded state and the fully unfolded state corresponds to the state with a single repeat folded (Fig. 5b). From the height of the free-energy barrier separating the folded state from the unfolded state, it is possible to obtain folding and unfolding rates by providing a valid preexponential factor† (defined by an intercept and denaturant dependence) (Fig. 8c). In the present case, by choosing a preexponential

† The preexponential factor is the parameter a in the formula connecting the folding rate to the activation free energy: $k = a\ \exp(-\Delta G_{\ddagger}/kT)$.

factor of $a = 1.3 \times 10^5\ s^{-1}\ \exp(-1.15[D])$, it is possible to draw a chevron plot for the slowest relaxation rates, using the height of the folding/unfolding barriers in Fig. 5b as activation free energies $\Delta G_{\ddagger}$ (Fig. 8b). The model captures the bending of the unfolding arm of NI_1C and NI_2C. The distance between the unfolding arms of the different proteins is also predicted quite accurately.

It may be worth underlining that this is not a conventional fit of the slowest kinetic traces of the proteins. Here, only the data from the equilibrium denaturation curves have been fitted using the Ising model, and the latter has been used to derive the kinetic properties using no additional parameters apart from the preexponential factor and its denaturant dependence. The preexponential factor, in this case, includes information on the rate of structure formation/structure loss of a single repeat. This term depends on denaturant concentration. This information is not present in the equilibrium data and thus has to be provided separately. The two parameters describing the preexponential factor are the minimal set of data needed to translate the height of the free-energy barriers of the Ising model into rates. Indeed, visually, the choice of the two parameters defining the preexponential factor is made by superimposing a straight segment of the folding arm of one of the theoretical chevron plots to the corresponding experimental data by suitably adjusting these two parameters. All the other features of all the chevron plots are then fixed. That means that the transition midpoints, the kinetic m_U values and the location (on the [GdnHCl] axis) and size of the change in m values of *all* the chevron plots are fixed once the folding arm of a *single* chevron plot is fixed. The Ising model preexponential factor, which reports the rate of structure formation/structure loss of the single repeat, at 0 M GdnHCl provides a lower boundary for the classical folding preexponential factor known from the literature (p. 558 of Fersht[30]), which reports the speed of diffusion at the top of the free-energy barrier.

The discrepancies found with the experimental kinetic data for NI_3C at high denaturant concentration might be related again to transient "salting-out" phenomena that slow the complete unfolding of the protein. The high consistency of data interpretation provided by the model is summarized below.

Discussion

The folding behavior of a series of six full-consensus DARPins has been analyzed by experimental thermodynamic and kinetic experiments. We examined the dependence of stability on repeat number, as well as the correlation between the folding and unfolding rates and the repeat number. Moreover, we propose a description of the folding mechanism of the three full-consensus DARPins NI_1C, NI_2C and NI_3C using two different models, which also allow us to predict the behavior of other members of this series of proteins.

Consensus DARPins are very stable proteins

The first observation is that our consensus design yielded extremely stable proteins. Equilibrium experiments showed that adding repeats increases the thermodynamic stability of the proteins, and, with NI_6C, an ankyrin protein with eight repeats, we still have not reached a stability limit, as judged from the distances between the T_m in 5 M GdnHCl. The proteins NI_4C, NI_5C and NI_6C, which are not completely denaturing as judged from their CD signal but show an increase in light scattering at 6–7 M GdnHCl, may associate via the exposed interface after one or both caps have denatured.[21] This soluble aggregate state of quasi-native proteins may provide the proteins with an even larger stability than the monomeric state. Experimentally, neither high temperature nor 8 M GdnHCl is able to unfold them. Only by combining heat and GdnHCl could we unfold these molecules and distinguish their stabilities. The stability increases with increasing repeat number; however, due to the extremely high thermodynamic stability and the inability to fully denature the larger proteins (Figs. 2 and 7), it is difficult to quantitate this for the whole series.

In Fig. 7, it is apparent that the distances between the main transition midpoints become smaller with increasing repeat number. This raised the question as to whether we can calculate the equilibrium parameters for NI_4C, NI_5C and NI_6C, as well as a maximal asymptotic denaturation midpoint of DARPins NI_xC with large x. Calculations using the Ising model provide theoretical values for the large DARPins (for NI_4C, NI_5C and NI_6C, see Table 2). However, these values, being the product of an extrapolation process, should be considered as only indicative. The assumed linearity of both stabilities and m_L for large repeat numbers L (see Results) leads to an asymptotic value for the transition midpoint of the repeat proteins. In the case of the DARPins, this value (~11 M GdnHCl) is well above the experimentally accessible range.

Unfolding rates decrease enormously with increasing repeat number

Second, we analyzed the dependence of the folding and unfolding rates on the repeat number. The dependence of the stability on the number of repeats is thus determined by a change in the unfolding rate, while the refolding rates of all three proteins are very similar (Table 4). At the molecular level, we imagine the following: in a molecule with many repeats, repeat interactions broken during unfolding by fluctuations will usually reseal and reanneal, since the remaining folded fragments do not denature. Only if the remaining folded fragment is small (consisting of one or two repeats) will it denature. The more repeats there are in the molecule, therefore, the less likely it will generate folded fragments that are small enough to continue complete denaturation. In the case of NI_3C, we have two strong interactions by the three highly compatible

full-consensus interfaces (Fig. 1); in NI_1C, there is none; and, in NI_2C, there is only one such strong interaction. This explains the decrease of k_u with repeat number and the very slow unfolding rate of NI_3C.

In the language of Φ-value formalism,[30] the addition of repeats would be assigned a formal value of 0 for the folding reaction, as this addition would not change the transition state in folding but would only contribute interactions on the native side of the transition state. This formalism only conceptualizes the role of the additional repeats in stabilizing the native state, but of course makes no statement about the nature of the transition state within a single repeat.

Interpretation using classical cooperative folding models

None of the studied proteins can be interpreted using the classical two-state model of protein folding as the chevron plots exhibit nonlinearity in the unfolding limbs and, for the larger proteins, a multiphasic behavior. This can be treated by a formalism,[28] according to which folding proceeds through transiently populated and partially folded intermediates that are separated by major free-energy barriers along a reaction coordinate. The intermediates can be considered as local minima between the global minimum of the native state and the ensemble of unfolded states. The simplest case for such complex kinetics is the sequential three-state mechanism with folding through an on--pathway high-energy intermediate.[29] When several kinetic phases can be detected, the intermediate state is more populated. In our case, we have three states: N, U and an on-pathway intermediate I.[28]

The rates of NI_1C are all monoexponential, yet due to the curvature in the unfolding limb, we have to assume a deviation from a simple two-state model for this protein as well. Indeed, the data can be fitted using a sequential three-state mechanism with high-energy intermediate. Consistent with these results, differential scanning calorimetry experiments revealed a deviation of an NX_1C library member from a two-state model at a higher temperature, while in the kinetic unfolding at 5 °C, no intermediates were detected.[31] NI_2C can also well be fitted to the on-pathway three-state mechanism, with the intermediate being more populated than in NI_1C. The kinetics of NI_3C, however, are even more complex as we detected three separated phases for unfolding. This can no longer be quantitatively fitted, as there would be too many parameters.

Interpretation using an Ising model

According to a second model, based on the classical one-dimensional Ising chain, the DARPin constituent repeats can fold independently and interact via nearest-neighbor coupling. The model provides a very good fit to the equilibrium denaturation data (Fig. 4) and an appealing rationalization of the kinetic experiments. Namely, the model fitted only on equilibrium data offers a prediction for the free energy of partially unfolded states (Fig. 5) that allows for the drawing of a chevron plot for each DARPin (Fig. 8b). The theoretical chevron plot is in good agreement with the experimental data and provides a validation for the Ising model of the DARPins. This model, in fact, provides a comprehensive explanation for some experimental data summarized in the following.

The Ising model fitted to DARPins predicts that the activation free energy of folding is practically the same for all the proteins, irrespective of their length (Fig. 5b): after the nucleation step, provided by the folding of one of the consensus repeats, the other repeats condensate on this folding nucleus. In the same way, according to the Ising model, the activation free energy of unfolding increases with the number of repeats because any folded repeat added to the folding nucleus provides a further decrease in the free energy of the native state of the molecule (Fig. 6). The kink in the slowest unfolding limb of the chevron plot of NI_1C and NI_2C is ascribed to the stabilization of partially unfolded states, with one (for NI_1C) or two (for NI_2C) terminal repeats unfolded (Fig. 5b). In other words, while the transition state remains the same as [D] increases, the most stable conformer in the native free-energy basin changes with [D], causing changes in the major unfolding free-energy barrier.

In more detail, the appearance of multiple phases in the unfolding arm of the chevron plots of NI_2C and NI_3C could be explained by several mechanisms. One possibility that is very plausible in the classical model is that the fast phases may correspond to the fast unfolding of the less stable terminal repeats at high GdnHCl concentration. This rate is limited by the intrinsic rate of the structure disruption of the terminal repeats. A further splitting of the observable phases may be caused by the different natures of the N- and C-terminal repeats.[21]

In principle, multiple pathways may exist from the fully folded state to the fully unfolded state, some of which may not reach the lowest energy intermediate with the unfolded termini. It would be difficult, however, to determine the relative amplitudes of such pathways using the classical model.

In contrast, in the Ising model, we obtain an energy landscape as a function of GdnHCl. This landscape predicts an intermediate with the C-cap and the N-cap unfolded at 6 M GdnHCl for NI_3C (Fig. 5d). The slowest of the three unfolding phases of NI_3C at 6 M GdnHCl (Fig. 5d) has been thus associated with the pathways passing through these most stable intermediates (three central repeats folded; state in dark brown), while the two fast phases may correspond to the pathways passing through the less stable intermediates (two with three repeats folded, including a terminal one). We illustrate this in Fig. 5d.

Figure 5c shows a similar pattern as in the free-energy landscape described by Mello and Barrick,

where the fully folded state and the fully unfolded state are free-energy minima, the states with only one folded repeat have the highest (most unfavorable) free energy and the other states have a free energy that decreases roughly linearly with the number of folded repeats.[11] In this work, using full-consensus designed proteins with identical internal repeats, we have, however, removed most of the heterogeneity present in the study of Mello and Barrick.[11] As a consequence, we have a large degree of degeneracy due to the presence of many states with the same free energy.

Comparison to the folding of other proteins

Comparing the rates with the natural AR proteins p16 and p19,[6,9] our DARPins fold with faster rates and unfold much more slowly at 20 °C.

Rates for another class of consensus repeat proteins have been measured for tetratricopeptide repeat (TPR)[27] and a hybrid ankyrin construct[25] consisting of natural and consensus ARs. Our result—unfolding rates slowing with the number of repeats —is in agreement with the kinetic study of Main *et al.*[27] In their study, the unfolding rate decreased by a factor of 36 when comparing the two-repeat CTPR2 with the three-repeat construct CTPR3 (0.35 s^{-1} and 0.01 s^{-1} for the two-repeat and three-repeat constructs, respectively),[27] while the folding rates were similar (≈20,000 s^{-1} and 35,000 s^{-1} for CTPR2 and CTPR3, respectively). Compared to DARPins, these proteins unfold extremely fast (10^2- to 10^7-fold faster than the DARPins) and show the classical two-state behavior; therefore, possible intermediate states cannot be detected with standard experimental methods. In another study using an Ising model to describe a series of consensus TPR proteins,[26] a stabilizing energy of about 4 kcal mol^{-1} per repeat was obtained, while our DARPins gained 11 kcal mol^{-1} in stability per additional repeat (see Fig. 6). This illustrates why adding ARs decreases the unfolding rate much more than the addition of TPR.

A theoretical study by Plaxco *et al.* has suggested that there may be a linear correlation between the folding rate and the topology [expressed as relative contact order (RCO), the contact order normalized by the length of the protein] of small single-domain proteins exhibiting two-state kinetics.[32] Proteins with low RCO (i.e., containing mainly residues interacting with other residues that are close in sequence distance) fold very fast. Proteins with high RCO fold more slowly. In repeat proteins, which exclusively exhibit local interactions, the RCO does not change with increasing repeat number, and thus no change in the folding rate would be expected either. The folding rates‡ predicted from this model of NI_1C, NI_2C and NI_3C are 8.9×10^6 s^{-1}, 9.3×10^6 s^{-1} and 9.4×10^6 s^{-1}, respectively, while the corresponding measured folding rates at 20 °C range from 450 s^{-1} to 800 s^{-1} (638 s^{-1}, 790 s^{-1} and 450 s^{-1}, respectively; see also Table 4). Although the predicted folding rate is 10^4 times faster than the experimentally observed rates and is thus not in good quantitative agreement, it is interesting to see that the folding rates of all three proteins are very similar to each other, consistent with the model.

The Ising model suggests that folding and unfolding occur along multiple pathways following a nucleation–condensation mechanism, where multiple nucleation sites are possible. The fitting of the data to the model also suggested a possible size for the nucleation site (one consensus repeat for DARPins) and supports the emergence of stable partially folded intermediate states at high GdnHCl concentration.

‡ http://depts.washington.edu/bakerpg/

Conclusions

Our study revealed two major insights: even though many AR proteins show two-state behavior in equilibrium studies, all kinetic studies performed with AR proteins to date have proven that the folding mechanism is more complex, with at least one intermediate state. In addition, our generalized full-consensus DARPin series confirms these findings. Second, the stability of the repeat proteins, characterized by short-range interactions and low contact order, is determined by the unfolding rates. With increasing repeat number, the unfolding rates decrease moderately in TPR proteins, but enormously for our stable consensus DARPins. This behavior can be rationalized, following the Ising model, by considering the folding process as a nucleation process where the formation of a small assembly of repeats (probably one single consensus repeat in the DARPins) triggers the whole folding cascade. The reverse process (i.e., unfolding), on the other hand, requires the progressive disruption of all the "condensed" folded repeats and is thus dependent on protein length.

Materials and Methods

Design and synthesis of DNA-encoding AR proteins

Oligonucleotides were obtained from Microsynth (Balgach, Switzerland), following the assembly strategy described previously:[18]

INT5 (forward) : 5′-TTCCGCGGATCCTAGGAAGACCTGACGTTAACGCT-3′

PRO1 (forward) : 5′-CTGACGTTAACGCTAAAGACAAAGACGGTTACACTCCGCTGCACCTGGC-3′

PRO2 (forward) : 5′-ACTCCGCTGCACCTGGCTGCTCGTGAAGGTCACCTGGAAATCG-3′

PRO3 (reverse) : 5′-ACGTCAGCACCAGCCTTCAGCAGAACTTCAACGATTTCCAGGTGACC-3′

PRO4 (reverse) : 5′-TTTGGGAAGCTTCTAAGGT-CTCACGTCAGCACCAG-3′.

The full-consensus AR was generated by assembly PCR using the oligonucleotides PRO1, PRO2, PRO3, PRO4 and INT5, and Vent® polymerase (5 min at 95 °C; followed by 25 cycles of 30 s at 95 °C, 1 min at 50 °C and 30 s at 72 °C; followed by 5 min at 72 °C; standard Vent® polymerase buffer with a final concentration of 3.5 mM $MgSO_4$). The PCR product was cloned via BamHI/HindIII into pPANK,[18] a pQE30 (QIAgen, Germany) derivative lacking the BbsI and BsaI sites, and sequenced using standard techniques. The resulting plasmid was termed pPRO.

Using the plasmids pPRO, pEWT (a pPANK derivative containing the N-terminal capping AR) and pWTC (a pPANK derivative containing the C-terminal capping AR), the DNA encoding the six AR proteins was generated by a ligation procedure using the type IIs restriction enzymes BpiI and BsaI, as described previously.[18]

Protein expression and purification

The repeat proteins NI_1C–NI_6C were expressed as follows: 50 ml of overnight cultures of *E. coli* XL1-Blue (LB medium, 1% glucose and 100 mg/l ampicillin; 37 °C) was used to inoculate 1-l cultures (LB medium, 1% glucose and 50 mg/l ampicillin; 37 °C). At $OD_{600}=0.7$, the cultures were induced with 500 μM IPTG, and incubation was continued for 4 h. The cultures were harvested by centrifugation at 3300*g* for 10 min at 4 °C, and the resulting pellets were resuspended in 40 ml of 50 mM Tris–HCl (pH 8) and 500 mM NaCl. The cells were lysed using a French press, and the lysate was centrifuged again at 8000*g* for 15 min at 4 °C, and glycerol (final concentration, 10%) and imidazole (final concentration, 20 mM) were added to the resulting supernatant. The proteins were purified over a Ni-NTA column (column volume, 3.8 ml) in accordance with the manufacturer's instructions (QIAgen). The protein was then dialyzed overnight against 50 mM sodium phosphate buffer (pH 7.4) and 150 mM NaCl (PBS_{150}). Purity was checked with 15% SDS-PAGE, the monomeric state was verified by gel filtration combined with multiangle light scattering (miniDAWN, Wyatt, Germany; Astra software) and the correct molecular mass was verified by mass spectrometry.

CD spectroscopy

The CD signal at 222 nm was recorded on a Jasco J-715 instrument (Jasco, Japan) equipped with a computer-controlled water bath, using a cylindrical quartz cell of 1 mm pathlength. To measure denaturant-induced equilibrium unfolding, CD data were collected at 222 nm (measurement intervals, 5 s), 2-nm bandwidth and 4-s response time, and each data point was averaged over 2 min. Thermal unfolding was recorded by continuous heating with a temperature gradient of 0.5 °C min^{-1} from 15 °C to 95 °C. CD data were collected at 222 nm (measurement intervals, 5 s), 2-nm bandwidth and 4-s response time. Reversibility was determined from the recovery of ellipticity after cooling.

All CD experiments were performed in PBS_{150} using 10 μM purified protein, and a baseline correction was made with the buffer. The CD signal was converted to mean residue ellipticity (Θ_{MRW}) using the concentration of the sample determined spectrophotometrically at 280 nm.

Fluorescence spectroscopy

Tyrosine fluorescence was measured by excitation at 274 nm and by recording the emission spectra from 290 nm to 350 nm using a PTI Alpha Scan spectrofluorimeter (Photon Technologies, Inc.). Slid widths of 5 nm were used for both excitation and emission. Samples were prepared as for the CD measurements. After buffer correction, the intensity of the emission maximum at 306 nm was plotted against denaturant concentration.

Equilibrium unfolding

The transitions were monitored using both the CD signal and tyrosine fluorescence. For measuring denaturant-induced equilibrium unfolding curves, the samples were equilibrated overnight at the corresponding GdnHCl concentrations at 20 °C. The GdnHCl concentrations were determined by refractive index.

The data were fitted by assuming both classical cooperative folding models and an Ising model (see below). Where indicated, a two-state model[33] was used according to Eq. (1):

$$S_{obs}(D) = (S_U + m_U[D])f_U + (S_N + m_N[D])f_N \quad (1)$$

where $S_{obs}(D)$ is the observed quantity (e.g., CD signal or fluorescence signal) as a function of denaturant concentration [D], S_U is the signal of the unfolded state (extrapolated to zero [D]), S_N is the signal of the native state at zero [D], and f_U and f_N are unfolded and native fractions, respectively, defined as:

$$f_U = 1 - f_N = K_U/(1 + K_U) \quad (2)$$

where $K_U=[U]/[N]$ is the equilibrium constant of denaturation. Where indicated, a sequential three-state model[34,35] for equilibrium denaturation was used according to Eqs. (3)–(9):

$$U \rightleftharpoons I \rightleftharpoons N \quad (3)$$

with

$$K_{UI} = [I]/[U] \quad (4)$$

$$K_{IN} = [N]/[I] \quad (5)$$

$$f_U = 1/(1 + K_{UI} + K_{UI}K_{IN}) \quad (6)$$

$$f_I = K_{UI}/(1 + K_{UI} + K_{UI}K_{IN}) \quad (7)$$

$$f_N = K_{UI}K_{IN}/(1 + K_{UI} + K_{UI}K_{IN}) \quad (8)$$

$$S_{obs}(D) = (S_U + m_U[D])f_U + (S_I + m_I[D])f_I + (S_N + m_N[D])f_N \quad (9)$$

where U, I and N represent the unfolded, intermediate and native states of the protein, and *f* and *S* are the fractional population and spectroscopic signals of the states indicated by the subscripts, respectively. Both ln (K_{IN}) and ln (K_{UI}) are assumed to be linearly dependent on [D]. Data were fitted using ProFit (Quantum Soft, Switzerland).

Kinetic folding and unfolding

Kinetic experiments were performed with a PiStar-180 stopped-flow instrument (Applied Photophysics, UK). The CD signal was monitored at 225 nm, and tyrosine fluorescence changes were monitored with a 295-nm

cutoff filter. For NI_3C, an optical pathlength of 2 mm was used, and the final concentration was 18 μM. NI_1C and NI_2C were measured with a 10-mm pathlength, and the final protein concentrations were 4–5 μM. All refolding and unfolding reactions were measured in PBS_{150} at 20 °C. Refolding experiments were performed as follows: the protein NI_1C was denatured in 2.5 M, the protein NI_2C was denatured in 5.5 M and the protein NI_3C was denatured in 6.8 M GdnHCl. Refolding was initiated by rapid mixing of 1 vol of denatured protein solution with 10 vol of buffer containing various concentrations of denaturant. Unfolding rates were measured by rapid mixing of 1 vol of native protein solution with 10 vol of denaturing buffer with different GdnHCl concentrations (>1.4 M for NI_1C; >4.1 M for NI_2C; >5.6 M for NI_3C). Reactions with a half-life longer than about 12 s were mixed manually.

Ten to 25 kinetic traces were averaged for each GdnHCl concentration and fitted to either a single exponential function or a double exponential function according to Eqs. (10) and (11):

$$S_{obs}(t) = a\ e^{-\lambda\ t} + b \tag{10}$$

$$S_{obs}(t) = a_1\ e^{-\lambda_1\ t} + a_2\ e^{-\lambda_2\ t} + b \tag{11}$$

where a_1 or a_2 represents the change in spectroscopic signal of phase 1 or phase 2; λ_1 or λ_2 represents the observed rate constants of phase 1 or phase 2; and b represents the spectroscopic signal after the reaction had reached equilibrium.

The observed rate constants λ_i were analyzed as a function of the GdnHCl concentration, according to a kinetic sequential three-state model with on-pathway intermediate:[28]

$$U \underset{k_{IU}}{\overset{k_{UI}}{\rightleftharpoons}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} N \tag{12}$$

where all four microscopic rate constants k_{UI}, k_{IU}, k_{IN} and k_{NI} are defined by the solution of a quadratic equation according to Eqs. (13)–(15):

$$\lambda_{1,2} = \frac{-B \pm \sqrt{B^2 - 4C}}{2} \tag{13}$$

with

$$B = -(k_{UI} + k_{IU} + k_{IN} + k_{NI}) \tag{14}$$

$$C = k_{UI}(k_{IN} + k_{NI}) + k_{IU}k_{NI} \tag{15}$$

and the refolding and unfolding rates are defined according to Eqs. (16) and (17):

$$k_f^0 = k_{UI}^0 k_{IN}^0 / (k_{UI}^0 + k_{IN}^0) \tag{16}$$

$$k_u^0 = k_{NI}^0 k_{IU}^0 / (k_{NI}^0 + k_{IU}^0) \tag{17}$$

For all k_{ij}, the relation $\ln k_{ij} = \ln k_{ij}^0 + m_{ij}[D]$ was assumed, where k_{ij} represents the denaturant-dependent rate constant, k_{ij}^0 is the rate constant in the absence of denaturant and m_{ij} is the denaturant dependence of $\ln k_{ij}$.

For NI_1C, a three-state model with a metastable high-energy intermediate was used,[29] as all measured kinetics were monoexponential. For NI_2C, the single observed refolding phase and both unfolding phases were simultaneously fitted to Eq. (13) using ProFit. For NI_3C, the folding and unfolding rate constants were determined by fitting the denaturant dependencies of the single refolding phase and the slowest unfolding phase, using the relation $\ln k_{ij} = \ln k_{ij}^0 + m_{ij}[D]$.

Ising model for the folding and unfolding of NI_xC

As an alternative to cooperative folding models (see above), an Ising-like model was used to study the dependence of the stability of the consensus DARPins NI_xC on denaturant concentration. According to this model, each repeat of the protein is considered as an independent folding unit with a free energy of unfolding that depends linearly on the denaturant concentration [D]. Adjacent folded repeats interact by a stabilizing potential, whose magnitude is independent of [D] but requires that both repeats be folded.

The effective free energy of a configuration of the repeat protein ΔG_{conf}, where some of the repeats may be folded and some may not, can be written according to Eq. (18):

$$\Delta G_{conf}(\{s_i\}; D) = \Delta G(D) \sum_{i=2}^{L-1} s_i + \Delta G'(D)(s_1 + s_L) + J \sum_{i=2}^{L-1} s_i s_{i+1} \tag{18}$$

where s_i is a variable that describes the state of the ith repeat (s_i=1 when folded; s_i=0 otherwise), and L is the total number of repeats in the protein. $\Delta G(D) = \Delta G_0 + m[D]$ is the free energy of folding of a consensus repeat that depends linearly on the denaturant concentration. A different stability $\Delta G'(D) = \Delta G_0' + m'[D]$ has been introduced to describe the different characteristics of the C- and N-terminal capping repeats with respect to the consensus repeats.

The interaction parameter J has been chosen as independent of [D] to reduce the number of free parameters of the model. More complex models have been considered, but, in our opinion, the improvement of the fit did not justify the increase in the number of free parameters of the model.

The probability of each state of the proteins $P_L(\{s_i\})$ (i.e., the probability of observing a certain set of folded/unfolded repeats $\{s_i\}$ for a protein with L repeats) under any concentration of denaturant [D] is obtained using the standard canonical formalism (Eq. (19)):

$$P_L(\{s_i\}; [D]) = \frac{e^{-\Delta G_{conf}(\{s_i\};[D])/kT}}{\sum_{\{s_j\}} e^{-\Delta G_{conf}(\{s_j\};[D])/kT}} \tag{19}$$

where k is the Boltzmann constant, T is the temperature and the sum in the denominator represents the canonical partition function (extended to all possible states of the protein; i.e., all the possible combinations of folded/unfolded repeats).

Finally, it is possible to compute the average fraction of folded repeats, at each value of [D], according to Eq. (20):

$$f_L([D]) = \sum_{\{s_i\}} P_L(\{s_i\}; [D]) \frac{1}{L} \sum_{i=1}^{L} s_i \tag{20}$$

The parameters of the model have been determined by fitting the denaturant-induced equilibrium denaturation data of proteins NI_1C, NI_2C and NI_3C, which were obtained at room temperature and monitored by CD. Equation (21) was used to fit the observed CD signal:

$$S_{obs} = (S_U + m_U[D])(1 - f_L([D])) + (S_N + m_N[D]) f_L([D]) \tag{21}$$

Note that, in Eq. (21), f_L refers to the fraction of folded repeats, while in Eqs. (1) and (9), f_N refers to the fraction of

folded whole molecules. The baselines of the CD signal for the folded and unfolded states have been considered as independent of [D] where possible (i.e., to have a slope of zero), in order to keep the number of free parameters as low as possible. In the case of NI_1C and NI_2C, the very long posttransition baseline had to be considered as linearly dependent on [D] because the approximation of a slope of zero was clearly not suitable. Different baselines were allowed for the different protein constructs. The least-squares fit has been obtained using the Levenberg–Marquardt algorithm, where both the parameters of the model and the coefficients of the baselines were allowed to vary.

Acknowledgements

We thank Drs. Ilian Jelesarov and Ben Schuler for fruitful discussions, Dr. V. Sathya Devi for help with the stopped-flow instrument, Dr. Christophe Bodenreider (Biozentrum, Basel, Switzerland) for help with data fitting using ProFit and Enrico Guarnera for his valuable help. G.S. would like to thank Drs. Amedeo Caflisch and Gianluca Interlandi for stimulating discussions. This work was supported by the Swiss National Science Foundation and the National Center of Competence in Research Structural Biology.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.11.046

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. 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.
4. Koradi, R., Billeter, M., Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55, 29–32.
5. 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.
6. Tang, K. S., Guralnick, B. J., Wang, W. K., Fersht, A. R. & Itzhaki, L. S. (1999). Stability and folding of the tumour suppressor protein p16. *J. Mol. Biol.* **285**, 1869–1886.
7. Tang, K. S., Fersht, A. R. & Itzhaki, L. S. (2003). Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure (Cambridge)*, **11**, 67–73.
8. Interlandi, G., Settanni, G. & Caflisch, A. (2006). Unfolding transition state and intermediates of the tumor suppressor p16INK4a investigated by molecular dynamics simulations. *Proteins: Struct. Funct. Genet.* **64**, 178–192.
9. Zeeb, M., Rosner, H., Zeslawski, W., Canet, D., Holak, T. A. & Balbach, J. (2002). Protein folding and stability of human CDK inhibitor p19(INK4d). *J. Mol. Biol.* **315**, 447–457.
10. Low, C., Weininger, U., Zeeb, M., Zhang, W., Laue, E. D., Schmid, F. X. & Balbach, J. (2007). Folding mechanism of an ankyrin repeat protein: scaffold and active site formation of human CDK inhibitor p19 (INK4d). *J. Mol. Biol.* **373**, 219–231.
11. Mello, C. C. & Barrick, D. (2004). An experimentally determined protein folding energy landscape. *Proc. Natl Acad. Sci. USA*, **101**, 14102–14107.
12. Zweifel, M. E. & Barrick, D. (2001). Studies of the ankyrin repeats of the *Drosophila melanogaster* Notch receptor: 2. Solution stability and cooperativity of unfolding. *Biochemistry*, **40**, 14357–14367.
13. Mello, C. C., Bradley, C. M., Tripp, K. W. & Barrick, D. (2005). Experimental characterization of the folding kinetics of the notch ankyrin domain. *J. Mol. Biol.* **352**, 266–281.
14. Bradley, C. M. & Barrick, D. (2006). The notch ankyrin domain folds via a discrete, centralized pathway. *Structure*, **14**, 1303–1312.
15. Mosavi, L. K., Williams, S. & Peng Zy, Z. Y. (2002). Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein. *J. Mol. Biol.* **320**, 165–170.
16. Lowe, A. R. & Itzhaki, L. S. (2007). Biophysical characterisation of the small ankyrin repeat protein myotrophin. *J. Mol. Biol.* **365**, 1245–1255.
17. Lowe, A. R. & Itzhaki, L. S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc. Natl Acad. Sci. USA*, **104**, 2679–2684.
18. 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.
19. O'Neil, K. T. & DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, **250**, 646–651.
20. 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.
21. Interlandi, G., Wetzel, S.K., Settanni, G., Plückthun, A., Caflisch, A. (2007). Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. *J. Mol. Biol.* in press [2007 Sep 21; Epub ahead of print]. doi:10.1016/j.jmb.2007.09.042.
22. Merz, T., Wetzel, S.K., Firbank, S., Plückthun, A., Grütter, M., Mittl, P.R.E. (2007). Stabilizing ionic interactions in a full consensus ankyrin repeat protein. *J. Mol. Biol.* [in press].
23. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
24. 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.
25. Tripp, K. W. & Barrick, D. (2007). Enhancing the stability and folding rate of a repeat protein through

the addition of consensus repeats. *J. Mol. Biol.* **365**, 1187–1200.
26. Kajander, T., Cortajarena, A. L., Main, E. R., Mochrie, S. G. & Regan, L. (2005). A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* **127**, 10188–10190.
27. Main, E. R., Stott, K., Jackson, S. E. & Regan, L. (2005). Local and long-range stability in tandemly arrayed tetratricopeptide repeats. *Proc. Natl Acad. Sci. USA*, **102**, 5721–5726.
28. Buchner, J. & Kiefhaber, T. (2005). Protocols—analytical solutions of three-state protein folding models. In *Protein Folding Handbook: Part 1*, vol. 1, pp. 402–406, Wiley-VCH Verlag GmbH and Co. KGaA, Weinheim.
29. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375–386.
30. Fersht, A. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Freeman, New York.
31. Devi, V. S., Binz, H. K., Stumpp, M. T., Plückthun, A., Bosshard, H. R. & Jelesarov, I. (2004). Folding of a designed simple ankyrin repeat protein. *Protein Sci.* **13**, 2864–2870.
32. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
33. Pace, C. N. (1975). The stability of globular proteins. *CRC Crit. Rev. Biochem.* **3**, 1–43.
34. Barrick, D. & Baldwin, R. L. (1993). Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry*, **32**, 3790–3796.
35. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.