Dissecting and reprogramming the folding and assembly of tandem-repeat proteins

Pamela J.E. Rowling[1], Elin M. Sivertsson[1], Albert Perez-Riba[1],
Ewan R. G. Main[2*], Laura S. Itzhaki[1*]

[1]University of Cambridge Department of Pharmacology, Tennis Court Road, Cambridge CB2 1PD, UK.

[2]School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK.

[*]To whom correspondence should be addressed: e.main@qmul.ac.uk and lsi10@cam.ac.uk

Studying protein folding and protein design in globular proteins presents significant challenges because of the two related features, topological complexity and cooperativity. In contrast, tandem-repeat proteins have regular and modular structures composed of linearly arrayed motifs. This means that the biophysics of even giant repeat proteins is highly amenable to dissection and to rational design. Here we discuss what has been learnt about the folding mechanisms of tandem-repeat proteins. The defining features that have emerged are: (i) accessibility of multiple distinct routes between denatured and native states, both at equilibrium and under kinetic conditions; (ii) different routes are favoured for folding versus unfolding; (iii) unfolding energy barriers are broad, reflecting stepwise unraveling of an array repeat by repeat; (iv) highly cooperative unfolding at equilibrium and the potential for exceptionally high thermodynamic stabilities by introducing consensus residues; (v) under force, helical-repeat structures are very weak with non-cooperative unfolding leading to elasticity and buffering effects. This level of understanding should enable us to create repeat proteins with made-to-measure folding mechanisms, in which one can dial into the sequence the order of repeat folding, number of pathways taken, step size (cooperativity) and fine-structure of the kinetic energy barriers.

Over the last 15 years the study of naturally occurring and consensus-designed repeat proteins has provided us with detailed insights into the folding mechanism of this distinctive structural class. These studies, predominantly of ankyrin repeats (ANK) and tetratricopeptide repeats (TPR), have revealed (i) how their folding mechanisms are mainly directed by the hierarchy of stabilities across a repeat array, (ii) how the intrinsic instability of individual repeats and the highly stabilising inter-repeat energies can enable very cooperatively folded structures despite the absence of long-range contacts, and (iii) how the modular and symmetrical nature of tandem repeat structures leads to folding landscapes in which many partially folded states have similar energies and which are very sensitive to local details meaning that even small perturbations can reroute the folding transitions. Below we describe the main findings and conclusions reached to date.

Repeat proteins have polarized folding mechanisms

Using a protein engineering approach known as "phi-value analysis" the transition states for folding of many globular and a number of repeat proteins have been mapped [1]. For globular proteins, despite the variety of different native structures, a common folding mechanism has emerged, termed the "nucleation-condensation" mechanism. According to this mechanism folding occurs via a transition state of diffuse structure i.e., all of the protein is involved to some extent. The fact that the whole poplypeptide chain contributes to the transition state structure means that the protein is less likely to undergo partial unfolding, which can lead to misfolding and disease. In contrast to globular proteins, phi-value analysis of repeat proteins indicates that they have polarized folding mechanisms in which a subset of repeats is highly structured in the transition state and other repeats are unstructured [2–6]. It has been shown that the order in which the repeats fold tends to be governed by thermodynamic stability, i.e. the subset of repeats that folds early is the one that is thermodynamically the most stable. Interestingly, certain low stability repeats that fold late or are natively disordered have been shown to be functionally important: the low stability repeats undergo folding transitions upon binding. Examples include ankyrin-repeat protein IkBa [7], TPR protein LrcH [8] and HEAT-repeat protein PR65/A [9].

The shape of the rate-limiting energy barrier reflects the repeat architecture

An essential component of the phi-value approach, and a powerful complement to it, is 'chevron plot' analysis. A chevron plot describes the perturbation in a protein's folding/unfolding kinetics (observed rate constants) when varying concentrations of chemical denaturant are added. Importantly, by measuring the denaturant dependence of the rate constants, one can obtain information on the both the position along the reaction coordinate of the rate-limiting energy barriers and the shape of these barriers (narrow or broad). A number of interesting features have emerged from the comparison of chevron plots from wild type and mutant variants. One feature that is commonly observed is downward curvature in the unfolding arm of the chevron plot. Downward curvature can be explained in terms of sequential transition states that, in the case of repeat proteins, reflect the sequential unravelling of the repeat array, repeat by repeat. Downward curvature tends to be more pronounced the greater the number of repeats in the protein [9,10]. The gradual shift from narrow to broad energy barriers for small *versus* large repeat proteins, resulting in a shift from relatively linear to highly non-linear unfolding arms, is particularly neatly illustrated in the series of consensus-designed TPR proteins, comprising between two and ten repeats, analysed by Javadi and Main [11] (Figure 1).

Parallel routes for folding repeat proteins

One key finding of chevron and phi-value analysis of repeat proteins, predicted by Ferreiro, Komives and Wolynes in 2005, is that there is more than one low energy route between the folded and unfolded states [12]. This was first shown experimentally in 2007 for the ankyrin-repeat protein myotrophin [3]. Parallel pathways can be inferred from the kinetic data in a number of different, mutually non-exclusive ways. First, mutation can shift the flux from one path to another giving rise to upward curvature in the unfolding arm of the chevron plot [9,10]. Second, the rate constants observed for a reaction along parallel pathways will be faster than for a single pathway [3]. A final key signature requires a double mutant analysis (Figure 2A). Mutants are first constructed that direct folding down one path or another. Then a second series are made on top of these and phi values calculated; the second series of mutants will display different phi values [3]. Thus, a repeat protein's folding pathway is balanced on a fulcrum. The heterogeneous distribution of stabilities across their repeat array results in a polarised folding mechanism. However, as the distribution of stabilities are sufficiently closely

balanced small perturbations can shift the folding flux from exclusively one pathway to another or allow flux through parallel pathways [3,12].

What folds first does not necessarily unfold last

One consequence of multiple pathways having similar energies is that what folds first does not necessarily unfold last. This apparently paradoxical behaviour can in fact be rationalised quite easily. Differences between the routes in terms of the position and shape of the rate-limiting energy barriers mean that their sensitivities to denaturant are different. Therefore under folding conditions of low denaturant one pathway may be favoured whereas under strongly denaturing conditions of high denaturant a different pathway is favoured. We have observed this behaviour most strikingly for the 7-ankyrin repeat protein gankyrin (R. Hutton, J. Wilkinson, M. Faccin, A. Pelizzola, A. Lowe, P. Bruscolini, LSI, submitted) (Figure 2B). In isolation the N-terminal subdomain is stable and it forms very rapidly from the denatured state under refolding conditions. However, it is also easiest to unravel the native state from the N-terminus.

Mapping the full breadth of the energy landscape of repeat proteins

The cooperative nature of globular protein architectures makes it difficult to dissect their folding pathways. Frequently only one snapshot of the reaction is accessible by conventional techniques: namely the rate-determining transition state, and the rest of the folding energy landscape is effectively invisible to experimentalists. In contrast, as described above and illustrated by gankyrin (Figure 2B), for repeat proteins we can access many more states and therefore map out the topography of their energy landscapes widely and in much greater detail: First, there can be more than one (un)folding pathway that is accessible, and by careful mutational analysis we can selectively populate and then map all of them. Second, the broad energy barriers often observed in the unfolding kinetics of repeat proteins mean that, again by carefully comparing wild-type and mutant rate constant over a broad range of denaturant concentrations, we can access both early and late parts of the reaction and thereby resolve how the protein gets to the top of the energy barrier and what happens subsequently. Lastly, there is one other tool in our

arsenal, and it is unique to repeat proteins. In phi-value analysis one uses a series of single-site substitutions to perturb the structure and thus free energy of the native state and thereby obtain structural information on intermediate states and transition states; for repeat proteins there is an alternative way of perturbing the energetics other than point mutations, namely the addition or removal of repeats. In a globular protein, removal of a subdomain would tend to unfold the whole structure. However, the modular nature of the structures means that they are strikingly tolerant to such gross modifications. The most elegant way in which this perturbation has been applied is by comparing the folding and unfolding kinetics of a series of consensus-designed repeat proteins of increasing length [11,13,14]. Such an approach has given us additional insights into the folding landscapes, including defining the size of the folding nucleus and identifying misfolded states. Two experimental studies of consensus TPR (CTPRa) and ANK proteins are the most complete to date [11,14]. In both studies, refolding rates increase with repeat number (although in the CTPRa study, the effect becomes smaller and smaller with increasing repeat number) (see Figure 1). For the ANK study, the increase in rate was modeled as multiple parallel folding pathways. In the case of the CTPRa data, the refolding arm of the chevron plots exhibited increased rollover as the proteins became larger. When these data were fitted it suggested that all CTPRa folding begins with the formation of a nucleus that has approximately similar burial of surface area in all constructs. When the proteins become larger, stable substructures form to produce metastable intermediates. Finally, the largest proteins exhibit a decrease in folding rate, caused by the population of a misfolded intermediate (Figure 1).

From small to giant repeat proteins

It is very challenging to map the kinetic folding pathways of large proteins, and indeed detailed analyses using approaches like phi-analysis have been limited to small globular proteins of less than ~100-150 residues. One constraint is that larger globular proteins are usually multi-domain and they often unfold *via* intermediate states that are prone to aggregation, making them difficult to characterise. In contrast, repeat protein's partly folded intermediates do not appear to be as prone to aggregation. This property could be important given the vectorial nature of protein biosynthesis on the ribosome, as it would allow repeat proteins to fold co-translationally without risk of misfolding [15].

A second reason for limiting folding studies to small globular proteins is that the kinetics for large globular proteins tend to be too complex to dissect easily. In contrast, the simple repeating architecture makes it much easier to dissect the folding mechanisms of even very large repeat proteins (up to 600 residues). As one would expect, these studies have given interesting insights through key features becoming more pronounced with increased size [9–11,16,17]. For example, we showed that the unfolding of the 590-residue 15-HEAT repeat protein PR65/A occurs via a double-branched mechanism in which the pathway bifurcates into parallel routes twice in the reaction [9] (Figure 3).

Folding mechanisms made-to-measure

Because of the simplicity and modularity of the structures it should be possible to design repeat proteins with made-to-measure folding mechanisms in which we can accurately tune the various "flavors" of the reaction. Indeed, it is already possible to predict many of the key features of the folding landscape of a repeat protein from the experimentally determined equilibrium stabilities of the individual repeats [18]. We have gone a step further and have been able to rationalise the folding mechanism of a repeat protein from a purely *in silico* analysis of the native structure [9]; using the 15-HEAT repeat protein PR65/A we showed that the stability, and hence folding order, could be predicted from the native contact density. Most recently for the 7 ankyrin-repeat protein gankyrin, using an *in silico* approach we were able to recapitulate many of the key features of its folding landscape, including relative stabilities of subdomains, the order of repeat folding, the presence of alternative pathways, and the shapes of the associated kinetic energy barriers (R. Hutton, J. Wilkinson, M. Faccin, A. Pelizzola, A. Lowe, P. Bruscolini, LSI, submitted).

Relationship between folding and function

The solenoid shapes of repeat proteins suggest that some may have functions such as molecular springs and elastic adaptors. Indeed analysis by atomic force microscopy (AFM) and force molecular dynamics simulations has revealed such elastic behaviour [19–21], consistent with the stretching and twisting motions that are implied by the

diverse structures observed crystallographically for some repeat proteins such as the HEAT-repeat importins [22]. Moreover stepwise folding and unfolding of repeat proteins may be important for translocation across membranes [17,23] to pass through the narrow pore of the proteasome (EMS, J. Wilkinson, LSI, unpublished results).

Outstanding questions

Some important questions remain. One is the nature of the first rate-determining step in the folding of repeat proteins. Protein-engineering phi analysis and consensus approaches may be too coarse-grained to give us enough information about the structure of the folding nucleus. Recently, we inserted loops of increasing length between consensus TPR motifs in order to further probe the initial rate-determining step of folding (AP, ERGM & LSI, unpublished results). We found that the folding rate was strikingly little perturbed even when we insert multiple copies of large loops of up to 25 residues. This result suggests that the rate-determining step involves the formation of individual helices or repeats and that their subsequent coalescence is very rapid.

Another outstanding question is the extent and nature of repeat-protein misfolding, and whether it is similar to that observed for polyproteins comprising tandem immunoglobulin domain for which the sequence identity between adjacent domains is high [24]. Any misfolding might be expected to be more likely for longer repeat proteins and for consensus-designed proteins in which repeats have identical sequences. However, misfolding by domain swapping, as observed for the poly-immunoglobulin domains, would be difficult in the case of repeat proteins because helices cannot swap position and remain folded as they have only very short connecting loops. Instead, misfolded states could involve the misdocking of non-adjacent repeats, although as the intervening repeat would be unfolded this would have a high entropic penalty. We recently proposed that there may be an element of negative design in the evolution of giant repeat proteins such as PR65/A, whereby cryptic misfolding signals within the unstable central repeats are protected by the adjacent high stability repeats [25].

For smaller (<7) naturally occurring repeat proteins and consensus-designed ankyrin-repeat proteins there is no evidence of misfolding in the chevron plots. Perhaps the route to correct folding of helical repeat proteins is too easy and rapid for misfolding to compete. One caveat is that the population of misfolded species may be too low to be detectable by any methods other than single-molecule techniques. Alternatively, misfolded states may be spectroscopically silent and therefore would go undetected in single-jump stopped-flow measurements; for the larger repeat proteins D34 and PR65/A, we have indeed detected, by double-jump experiments, additional intermediate states that only slowly reach the native structure although we have not characterised their structures [9,10]. The study of the CTPR series by Javadi and Main did provide evidence of misfolding [11]. Extreme rollover was observed in the folding arm as the number of repeats was increased, pointing to the formation of misfolded states that need to unfold before correct refolding can occur; moreover, the fluorescent dye 1-anilinophthalene-8-sulfonate (ANS), which binds to patches of hydrophobic residues, was found to bind to these misfolded states.

Lastly, the folding mechanisms of all-beta repeat proteins have not been characterised in any detail to date. It will be interesting to see to what extent and in what ways they differ in their folding properties compared to all-alpha and mixed alpha/beta repeat proteins.

Future directions: repeat-protein design and assembly

It was recognized early in the 2000's that the unique properties of repeat proteins could be readily exploited as non-antibody scaffolds. First, the residues important for binding are distinct from those important for stability of the fold. Hence consensus-designed proteins, with fixed framework positions and variable binding positions, can be used to construct novel binding molecules. Second, repeat proteins are modular and consensus-designed repeats are self-compatible. Thus there are no constraints of the number of modules that one could assemble. Third, proteins constructed from consensus-designed repeats express in very high yields and can be extraordinarily stable, both important characteristics for nanotechnology applications. Exploiting these properties, one can extend the modular design process from individual repeat proteins to the assembly of repeat proteins into novel biomaterials with encoded properties and precisely displayed

functional sites. To date, examples of this exciting field includes using metals, peptide binding, native chemical ligation and disulphide bonding to assemble repeat proteins into gels, films and fibres (Figure 5) [26–28]. One interesting design aspect that is peculiar to linear repeat proteins is the potential to change the shape of the superhelix (curvature, twist, pitch). We have recently shown that this can be achieved in a rational way for the TPR motif by using the HMM plots for different TPR sub-families and making helix-helix packing substitutions that move away from the consensus sequence (AP, G. Fischer, M. Hyvönen, ERGM & LSI, unpublished results). The ability to create repeat-protein building blocks with altered super-secondary structures would allow us to build materials with specific geometrical arrangements of the proteins and to display and organize functional groups within the materials in a precise and predefined way.

Acknowledgements

Figure Legends

Figure 1: Folding of consensus-designed TPR series

(A) Chevron plots for the CTPRa2 to CTPRa10 proteins. CTPRa2 is fitted to a linear 2-state model of folding. CTPRa3 to CTPRa10 fitted to a sequential 3-state on-pathway model. (B & C) Ribbon representation of the crystal structures of (B) CTPR2 [Protein Data Bank (PDB) entry: 1NA3] and (C) CTPR8 (PDB entry: 2AVP). Both were prepared using PYMOL. (D-F) Schematic of the proposed folding pathways of CTPRa proteins as they increase in repeat number. Cylinders are coloured from the N terminus in red and correspond to 1 helix. (D) Folding of CTPRa2: 2-state folding over the conditions studied with a transition state (T.S.) that is ≈50% solvent exposed as the native state. The transition state is drawn, for illustrative purposes, as 3 formed helices arranged as a formed repeat and 1 partially formed helix. (E) Folding of CTPRa proteins greater than 3 repeats: multistate folding through a stable intermediate. Although there is no evidence to support a structure for the intermediate state, the proposed structures (i, ii, and iii) are shown. If the stabilities of helices/repeats obtained from the Ising model are used, this would correspond to the formation of a unit equal to 2.5 repeats. Folding from the intermediate requires a rearrangement that has no change in compaction when passing through the final transition state on route to the native state. This could consist of the docking of preformed modules. (F) Folding of CTPRa proteins of at least 10 repeats: folding is hampered by the population of a misfolded intermediate (drawn as wrongly docked repeats). The protein has to unfold from this state to continue to fold productively to the final native structure [11].

Figure 2

(A) Rerouting and mapping the parallel folding pathways of myotrophin. The folding of myotrophin can be represented by analogy with weighing scales [3]. The two folding pathways are shown schematically: For pathway A folding starts from the C-terminus, pathway B from the N-terminus). Which pathway is followed depends on the relative stability of the two ends of the protein, represented as different sizes of weights on the scales that tip the balance accordingly. Chevron plots are shown for representative mutants, measured using stopped-flow fluorescence. The mutants in brackets (black data points) were designed to shift the flux exclusively through one or other folding pathway; additional mutations, in bold, were then made to probe the transition-state structure for that pathway (data points in colour). These data illustrate how the same mutation, A9G

for example, displays a different phi value depending on the mutant background in which it is made (either A9G or E17V/D20L) because folding is occurring along different pathways (A or B, respectively).

(B) Phi-value analysis gives multiple snapshots of the folding energy landscape of gankyrin. Box inset shows the chevron plot obtained for wild-type gankyrin using stopped-flow fluorescence. Fast and slow phases are observed, corresponding to folding/unfolding of an intermediate and of the native state, respectively (the fast unfolding phase was obtained by interrupted refolding experiments). Upward curvature, indicative of parallel pathways, is evident in the unfolding arm for the slow phase. In red are the unfolding arms of two destabilizing mutants in the N-terminal (left) and C-terminal (right) repeats that cause flux to be shifted exclusively through one or other pathway. Phi-value analysis maps out the structures of the transition states and intermediates for the two pathways, A and B, as shown in the schematic. Under native conditions pathway A is favored, whereas under strongly denaturing conditions (high urea concentrations) pathway B is favored (R. Hutton, J. Wilkinson, M. Faccin, A. Pelizzola, A. Lowe, P. Bruscolini, LSI, submitted).

Figure 3: Folding of a giant repeat protein

Unfolding mechanism of 15 HEAT-repeat protein PR65/A [9]. Box inset shows the kinetic unfolding phases observed for wild type by stopped-flow fluorescence. The data for wild type and mutants could be fitted to the model shown in the top panel in which the native protein unfolds via a series of intermediate states. According to this model, the pathway bifurcates into two parallel routes at two points in the reaction; the unfolding of repeats 3-10 (k_1) is competitive with unfolding of repeats 14-15 (k_2), and unfolding of repeats 1-2 (k_3) is competitive with unfolding of repeats 11-13 (k_4). Note the upward curvature in the urea dependence of the fastest unfolding phase, as the pathway switches from that going via intermediate I to intermediate II when $k_2 > k_1$. Bottom panel shows the proposed relationship between PR65/A folding and function. PR65/A is the scaffold subunit of the heterotrimeric Ser/Thr protein phosphatase, PP2A, shown schematically with catalytic C subunit and regulatory B subunit bound to a multiply phosphorylated intrinsically disordered substrate. The folding analysis showed that the central repeats of PR65/A are the least stable and that they are partially unfolded at physiological temperature; it is proposed that unfolding/refolding of these low-stability central repeats of PR65/A modulates the positioning of the PP2A catalytic subunit relative to the

substrate, opening and closing the interface to coordinate processive dephosphorylation of multiply phosphorylated intrinsically disordered substrates. Such a model was originally postulated based on the elastic behaviour of PR65/A [29].

Figure 4: Designing the self-assembly of repeat proteins

The top row shows different repeat protein components (A). Grey denotes scaffold modules and colour denotes modules with binding capabilities. The different types of repeat protein modules (e.g. ankyrin, TPR, etc) are indicated by different shaped pieces. Examples of published triggers for association are (A & H) chemical-induced polymerization of CTPR modules using native chemical ligation & disulphide bond formation[27,28], (B & H) metal induced polymerisation using minimized β-roll motifs on addition of La^{2+} [30], (C & I) thick film formation after large, rigid superhelical 18 CTPR module proteins were deposited on a teflon surface with a plasticizer and left to dry [rectangles of alternating blue and green of 3 module units denote the extending superhelix [31], (D & J) hydrogel formation of the same CTPR18 combining with multivalent cognate peptide–PEG cross-linker (pink circles on black lines)[26], (E & K) reversible gelation of a chimaera of a leucine zipper (black and red rectangles) with designed minimized β -roll motifs. On addition of Ca^{2+} (yellow circles) the minimized β -roll motifs fold (green squares) from unstructured polypeptides and oligomerize [32]. (F, G & L) In addition to the published routes to assembly and functionalisation, others can be envisaged such as domain swapping [for example, the self assembly of designed β-propeller proteins [33] (F) and insertion of domains between repeat modules (G). By mixing and matching in a combinatorial fashion, a further diverse arrays of nanostructures could be fabricated (L).

References

1   Matouschek, A., Kellis, J. T., Serrano, L. and Fersht, A. R. (1989) Mapping the transition state and pathway of protein folding by protein engineering. Nature **340**, 122–126.

2   Tang, K. S., Fersht, A. R. and Itzhaki, L. S. (2003) Sequential unfolding of ankyrin repeats in tumor suppressor p16. Structure **11**, 67–73.

3   Lowe, A. R. and Itzhaki, L. S. (2007) Rational redesign of the folding pathway of a modular protein. Proc. Natl. Acad. Sci. U. S. A. **104**, 2679–84.

4   Bradley, C. M. and Barrick, D. (2006) The notch ankyrin domain folds via a discrete, centralized pathway. Structure **14**, 1303–12.

5   Courtemanche, N. and Barrick, D. (2008) The leucine-rich repeat domain of Internalin B folds along a polarized N-terminal pathway. Structure **16**, 705–14.

6   DeVries, I., Ferreiro, D. U., Sánchez, I. E. and Komives, E. A. (2011) Folding kinetics of the cooperatively folded subdomain of the IκBα ankyrin repeat domain. J. Mol. Biol. **408**, 163–76.

7   Truhlar, S. M. E., Torpey, J. W. and Komives, E. A. (2006) Regions of IkappaBalpha that are critical for its inhibition of NF-kappaB.DNA interaction fold upon binding to NF-kappaB. Proc. Natl. Acad. Sci. U. S. A. **103**, 18951–6.

8   Singh, S. K., Boyle, A. L. and Main, E. R. G. (2013) LcrH, a class II chaperone from the type three secretion system, has a highly flexible native structure. J. Biol. Chem. **288**, 4048–55.

9   Tsytlonok, M., Craig, P. O., Sivertsson, E., Serquera, D., Perrett, S., Best, R. B., Wolynes, P. G. and Itzhaki, L. S. (2013) Complex energy landscape of a giant repeat protein. Structure **21**, 1954–65.

10   Werbeck, N. D., Rowling, P. J. E., Chellamuthu, V. R. and Itzhaki, L. S. (2008) Shifting transition states in the unfolding of a large ankyrin repeat protein. Proc. Natl. Acad. Sci. U. S. A. **105**, 9982–7.

11   Javadi, Y. and Main, E. R. G. (2009) Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. Proc. Natl. Acad. Sci. U. S. A. **106**, 17383–8.

12   Ferreiro, D. U., Cho, S. S., Komives, E. A. and Wolynes, P. G. (2005) The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. J. Mol. Biol. **354**, 679–92.

13   Interlandi, G., Wetzel, S. K., Settanni, G., Plückthun, A. and Caflisch, A. (2008) Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. J. Mol. Biol. **375**, 837–54.

14   Aksel, T. and Barrick, D. (2014) Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. Biophys. J. **107**, 220–32.

15   Evans, M. S., Sander, I. M. and Clark, P. L. (2008) Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. J. Mol. Biol. **383**, 683–92.

16   Werbeck, N. D. and Itzhaki, L. S. (2007) Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. Proc. Natl. Acad. Sci. U. S. A. **104**, 7863–8.

17   Kloss, E. and Barrick, D. (2009) C-terminal deletion of leucine-rich repeats from YopM reveals a heterogeneous distribution of stability in a cooperatively folded protein. Protein Sci. **18**, 1948–60.

18   Courtemanche, N. and Barrick, D. (2008) Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B. Protein Sci. **17**, 43–53.

19   Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V. and Marszalek, P. E. (2006) Nanospring behaviour of ankyrin repeats. Nature **440**, 246–9.

20   Li, L., Wetzel, S., Plückthun, A. and Fernandez, J. M. (2006) Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. Biophys. J. **90**, L30–2.

21   Serquera, D., Lee, W., Settanni, G., Marszalek, P. E., Paci, E. and Itzhaki, L. S. (2010) Mechanical unfolding of an ankyrin repeat protein. Biophys. J. **98**, 1294–301.

22   Forwood, J. K., Lange, A., Zachariae, U., Marfori, M., Preast, C., Grubmüller, H., Stewart, M., Corbett, A. H. and Kobe, B. (2010) Quantitative structural analysis of importin-β flexibility: paradigm for solenoid protein structures. Structure **18**, 1171–83.

23   Junker, M., Schuster, C. C., McDonnell, A. V, Sorg, K. A., Finn, M. C., Berger, B. and Clark, P. L. (2006) Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. Proc. Natl. Acad. Sci. U. S. A. **103**, 4918–23.

24   Wright, C. F., Teichmann, S. A., Clarke, J. and Dobson, C. M. (2005) The importance of sequence diversity in the aggregation and evolution of proteins. Nature **438**, 878–81.

25   Tsytlonok, M., Sormanni, P., Rowling, P. J. E., Vendruscolo, M. and Itzhaki, L. S. (2013) Subdomain architecture and stability of a giant repeat protein. J. Phys. Chem. B **117**, 13029–37.

26   Grove, T. Z., Osuji, C. O., Forster, J. D., Dufresne, E. R. and Regan, L. (2010) Stimuli-responsive smart gels realized via modular protein design. J. Am. Chem. Soc. **132**, 14024–6.

27   Phillips, J. J., Millership, C. and Main, E. R. G. (2012) Fibrous nanostructures from the self-assembly of designed repeat protein modules. Angew. Chem. Int. Ed. Engl. **51**, 13132–5.

28   Mejías, S. H., Sot, B., Guantes, R. and Cortajarena, A. L. (2014) Controlled nanometric fibers of self-assembled designed protein scaffolds. Nanoscale **6**, 10982–8.

29   Grinthal, A., Adamovic, I., Weiner, B., Karplus, M. and Kleckner, N. (2010) PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. Proc. Natl. Acad. Sci. U. S. A. **107**, 2467–2472.

30   Scotter, A. J., Guo, M., Tomczak, M. M., Daley, M. E., Campbell, R. L., Oko, R. J., Bateman, D. A., Chakrabartty, A., Sykes, B. D. and Davies, P. L. (2007) Metal ion-dependent, reversible, protein filament formation by designed beta-roll polypeptides. BMC Struct. Biol. **7**, 63.

31   Grove, T. Z., Regan, L. and Cortajarena, A. L. (2013) Nanostructured functional films from engineered repeat proteins. J. R. Soc. Interface **10**, 20130051.

32   Dooley, K., Kim, Y. H., Lu, H. D., Tu, R. and Banta, S. (2012) Engineering of an environmentally responsive beta roll peptide for use as a calcium-dependent cross-linking domain for peptide hydrogel formation. Biomacromolecules **13**, 1758–1764.

33   Voet, A. R. D., Noguchi, H., Addy, C., Simoncini, D., Terada, D., Unzai, S., et al. (2014). Computational design of a self-assembling symmetrical β-propeller protein. Proc. Natl. Acad. Sci. U. S. A. **11**, 15102-15107

(A)

(B)

(C)

(D) Denatured   T.S.   Native

(E) Misfolded   (F) Denatured   (i)   (ii)   (iii) Intermediate   Native

A

A9G Φ=0.1

A115G (A9G) Φ=0.9

Pathway A: Folding C to N

Pathway B: Folding N to C

Denatured

Native

A9G (E17V/D20L) Φ=0.8

A115G (E17V/D20L) Φ=0.0

B

TS B1

TS B2

Unfolding: Pathway B

Native

Denatured

TS₁

TS A

Folding: Pathway A

I