ELSEVIER

A recurring theme in protein engineering: the design, stability and folding of repeat proteins

Ewan RG Main[1,2], Alan R Lowe[3], Simon GJ Mochrie[4,5,6], Sophie E Jackson[1] and Lynne Regan[4,7]

Repeat proteins are ubiquitous and are involved in a myriad of essential processes. They are typically non-globular structures that act as diverse scaffolds for the mediation of protein–protein interactions. These excitingly different structures, which arise from tandem arrays of a repeated structural motif, have generated significant interest with respect to protein engineering and design. Recent advances have been made in the design and characterisation of repeat proteins. The highlights include re-engineering of binding specificity, quantitative models of repeat protein stability and kinetic studies of repeat protein folding.

Addresses
[1] Cambridge University Chemical Laboratory, Cambridge University, Lensfield Road, Cambridge CB2 1EW, UK
[2] Chemical Biology Unit, Department of Chemistry, University of Sussex, Falmer BN1 9QG, UK
[3] MRC Cancer Cell Unit, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK
[4] Departments of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA
[5] Department of Physics, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA
[6] Department of Applied Physics, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA
[7] Department of Chemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA

Corresponding author: Regan, Lynne (lynne.regan@yale.edu)

Current Opinion in Structural Biology 2005, **15**:464–471

This review comes from a themed issue on
Engineering and design
Edited by Sophie E Jackson and Lynne Regan

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2005.07.003

Introduction

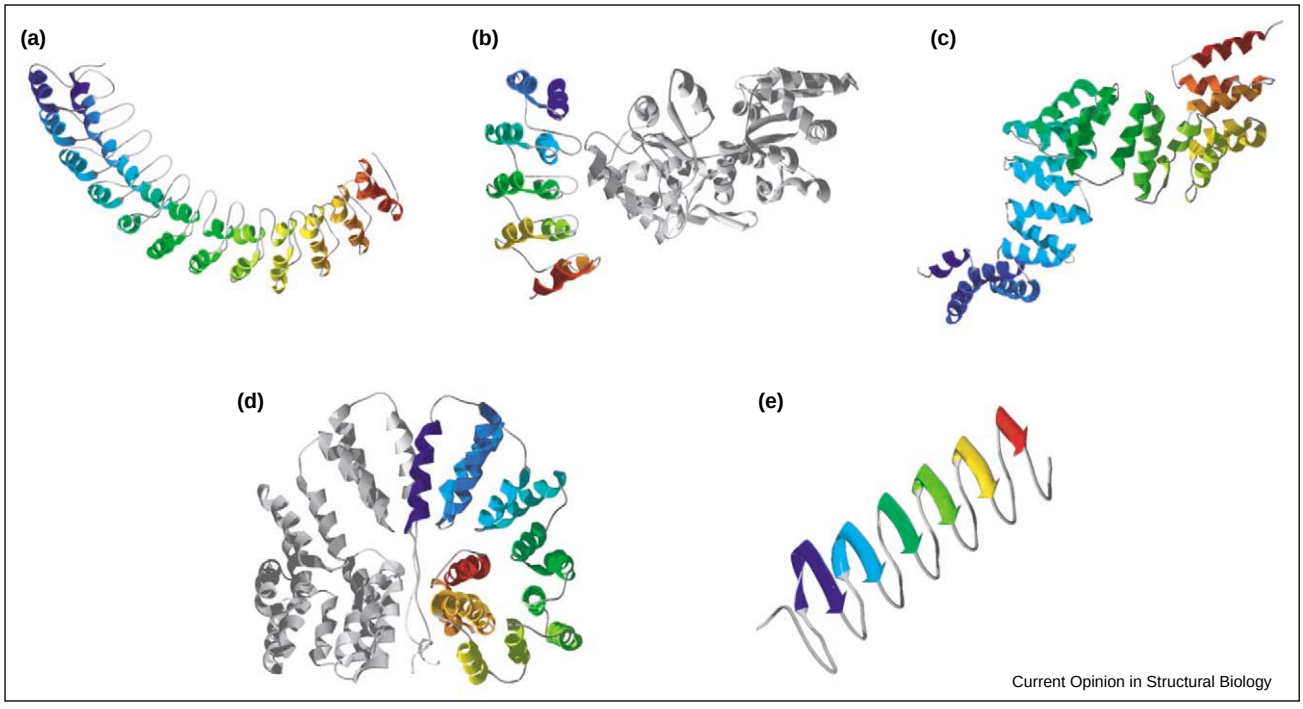
It now seems that, in a map of the 'protein universe', there may be less than 10 000 structural archetypes, or folds [1]. Of these, there are several extremely abundant non-globular folds, or repeat proteins — five families in the top twenty of the Protein Family Database (PFAM) [2,3]. These ubiquitous proteins are characterised by successive homologous structural motifs, or repeats, which stack together to form elongated structures. This construction causes repeat proteins to have shapes, structures and properties that are very different from those of typical globular proteins (Figure 1). Specifically, repeat proteins are dominated by short-range and regularised interactions, whereas globular proteins tend to have complex topologies and are stabilised by numerous long-range interactions. The modular architecture of repeat proteins may be key to their evolutionary success [4], allowing them to evolve not only via point mutations but also by insertion, deletion, duplication or rearrangement of repeat units. This capacity has enabled their involvement in a large and diverse range of protein–protein interactions [5–8]. The combination of non-globular fold and simplistic architecture makes repeat proteins an exciting new, and perhaps more tractable, system in which to investigate protein folding, stability, design and function.

In the 2003 'Engineering and design' section of *Current Opinion in Structural Biology*, we reviewed our then-current knowledge concerning the folding and stability of repeat proteins, and highlighted three successful design strategies [9]. Since then, many new publications have furthered our understanding of this class of protein. In this review, we will briefly recap our knowledge of repeat proteins up to and including our previous review, before discussing those new studies that further explore and expand our knowledge of repeat protein folding, design and structure/function relationships.

Repeat protein architecture

Each repeat protein is composed of tandem repeats of a basic structural motif of approximately 20–40 amino acids. The motifs can range from simple linked structural elements, such as a pair of α helices in the tetratricopeptide (TPR) motif [5,9], to more complex arrangements, such as the β-propeller structure of the 40-residue Trp-Asp (WD40) repeat [10] (Table 1). With few exceptions, tandem arrays of repeats stack to form elongated non-globular structures. These structures are stabilised by hydrophobic interactions, both within a repeat and between adjacent repeats (Figure 1). Thus, in general, sequentially distant residues in the structure do not interact with each other. An exception is Nlp1, a prokaryotic protein that is completely composed of TPR motifs and forms a globular protein [11•] in which the repeating units curl into each other to create a central hydrophobic core (Figure 1d). Many structures also have special repeats located at the N and C termini of the repeat protein. These seem to 'cap' the domains and thus shield hydrophobic residues from solvent [12–14].

Figure 1



Ribbon illustrations of various repeat proteins. **(a)** The 12 ANK stack from human ankyrinR, demonstrating the curved interaction surfaces [39]. **(b)** A complex showing the interaction between the designed ANK repeat protein off7 and MBD (grey) [19••]. **(c)** The superhelical TPR repeat domain of *O*-linked GlcNAc transferase, which comprises 11.5 repeats [40]. **(d)** The globular TPR protein NlpI from *Escherichia coli* K-12. Only one of the monomeric units of NlpI is coloured for clarity [11•]. **(e)** The *Tenebrio molitor* beetle antifreeze protein (TmAFP) [41]. Repeat proteins are coloured by secondary structure, progressing from the N terminus (blue) to C terminus (red). Figure prepared using Swiss-PDB viewer v3.7 SP5 (http://www.expasy.ch/spdbv) and POVRay v3.5 (http://www.povray.org), with PDB coordinates 1N11 (ankyrinR), 1SVX (off7), 1W3B (*O*-linked GlcNAc transferase), 1XNF (NlpI) and 1EZG (TmAFP).

Importantly, although each repeating sequence is highly degenerate (even within subfamilies), with no position invariant, there is always a consistent pattern of key residues that are essential for the structural integrity of the fold. These tend to form many of the inter- and intra-repeat packing interactions, and thus define both the secondary and tertiary characteristics of the fold, and the overall stability.

Consensus designed repeats

Their relative simplicity and modular nature have made repeat proteins exciting targets for design and redesign efforts. In our last review, we discussed the successful design of ankyrin (ANK) and TPR motifs [9]. Since then, the design of a leucine-rich repeat (LRR) has also been reported [15].

In each of these examples, a single consensus repeat was designed and then repeated in tandem to give folded, and in some cases functional, repeat proteins. In addition, both the ANK and LRR designs from the Plückthun laboratory, and the TPR designs from the Regan group had N- and C-capping repeats or helices, which were incorporated to avoid solvent exposure of hydrophobic residues at the ends of the repeat. The inclusion of these features has been shown to enhance the solubility and stability of the designed proteins. All the designs produced a series of highly stable repeat proteins, with differing numbers of tandem repeat motifs. The biophysical properties of the designed proteins are summarised in Table 2.

It is interesting to note that, in these designs, covariance among individual amino acids was not explicitly defined or used. However, when this type of analysis was performed on the ANK repeat motif, the highest covariation was observed among the most highly conserved residues and thus was implicitly incorporated into the designed sequence [16]. Recently, Magliery and Regan [17•] have further explored consensus design, with particular attention on TPR proteins. They have described how analysis of statistical free energies can provide a more quantitative description of variability, or conservation, at each position. Moreover, they have shown how covariance analysis can pinpoint interactions between poorly conserved residues, which can then be incorporated as an additional design feature, to modulate stability for example ([17•]; L Regan, TJ Magliery, unpublished).

Table 1

Examples of commonly occurring repeat protein architectures and their interactions.

Repeat type[a]	Architecture	Example interaction	PDB codes
Antifreeze protein (AFP)	A 12-residue motif forming a regular α helix, which resembles a rectangle in cross-section	Prevents ice crystals forming in plant/animal tissues [42]	1EZG [41]
ANK repeat	A 33-residue motif forming a helix-loop-helix-α turn motif, which is L-shaped in cross-section	Amongst the best-characterised interactions are those of the INK4 proteins with cyclin-dependent kinases [43,44]. Further examples include the interaction of IκB with NF-κB [45]	1A5E [46], 1BI7, 1BI8 [43], 1G3N [43], 1NFI [45]
Armadillo repeat (ARM)	An ~40-residue motif forming a three-helix bundle	The armadillo domain of β-catenin interacts with the cytosolic domain of E-cadherin [47]	1BK5 [48], 1I7X [47]
HEAT repeat	A 37- to 47-residue motif; each module comprises a pair of antiparallel helices	Importin-β interacts with sterol regulatory element binding protein (SREBP-2) [49]	1F59 [50], 1UKL [49]
Hexapeptide repeat	A hexapeptide motif comprising a β strand and loop, which forms a continuous β helix resembling an equilateral prism in cross-section	Galactoside acetyltransferase in complex with coenzyme A and β-galactoside [51]	1KRV [51]
LRR	A 20- to 29-residue motif forming a β strand-loop-helix structure	Human placental RNase inhibitor, an LRR protein, binds to human angiogenin with high affinity [52]	1FO1 [53], 1A4Y [52]
TPR	A 34-residue motif; each module comprises a pair of antiparallel helices	The adaptor protein Hop contains two TPR regions, which bind Hsp70 and Hsp90 [54]	1NAO [12], 1W3B [40], 1ELW [54]
WD40 repeat	A 40- to 50-residue motif forming a four- stranded β sheet	The β subunit of the G protein heterotrimer Gi $\alpha_1\beta_1\gamma_2$ contains a sevenfold WD40 β propeller, which interacts with both α and γ subunits [55]	1QHU [56], 1GP2 [55]

[a] Many of the repeat proteins listed can be further split into subfamilies [3].

Studies of designed repeat proteins have progressed in two main areas: the introduction of novel binding specificities [18•,19••,20••], and the exploration and characterisation of the fundamental properties of repeat proteins [21•,22•,23••]. These recent studies are discussed in detail below.

Novel binding specificities introduced onto repeat protein frameworks

Plückthun and co-workers [13,15,19••,24,25] have created libraries of ANK repeat proteins that have the potential to bind to any protein. Their approach was to use a consensus ANK framework and to randomise potential ligand-binding residues (identified by analyzing the structures of natural ANK protein–ligand complexes). As proof of principle, they used ribosome display to select members of their ANK library that bind with nanomolar affinity to maltose-binding protein (MBD) and to two eukaryotic kinases [19••]. A crystal structure of the complex between a selected ANK protein and MBD showed that the designed ANK proteins bind in a very similar manner to their natural counterparts (Figure 1b). An interesting extension of this work was to select for ANK variants that function, both *in vitro* and *in vivo*, as kinase inhibitors [20••]. The possible applications of these and other functional designs are far reaching and very exciting.

Cortajarena *et al.* [18•] redesigned a consensus TPR protein (CTPR3) to bind the C-terminal peptide of Hsp90 (a natural and well-studied binding partner). Using a database of the subset of natural TPR proteins that bind Hsp90, the residues in direct contact with ligand were defined and incorporated into the design. Furthermore, by manipulating the charge on the back face of the TPR protein, it proved possible to create variants with a range of binding affinities for the Hsp90 peptide. Certain designs not only bound to Hsp90 with higher affinity than the natural TPR partners, but also exhibited greater discrimination against non-cognate ligands (AL Cortajarena, L Regan, unpublished).

Table 2

Biophysical characteristics of naturally occurring and designed repeat proteins.

Protein	Repeat	Number of repeats	*m* (kcal mol^{-1}M^{-1})[a]	ΔG_{D-N} (kcal mol^{-1})[a]	T_m (°C)	k_f^{H2O} (s^{-1})[a]	References
Naturally occurring							
p16C[b]	ANK	2	–	1.7	31	–	[57]
p16	ANK	4	1.7	3.1	44	0.8	[57,58]
p18	ANK	5	3.8	3.0	–	–	[59]
p19	ANK	5	–	–	52	–	[34]
Myotrophin	ANK	4	1.6	5.3	53	95	[60] (f)
Notch	ANK	7	2.8	8.0	45	–	[61,62]
Nank1–6*[c]	ANK	6	1.9	4.0	39	–	[61,62]
Nank1–5*	ANK	5	1.7	3.7	–	–	[61,62]
Nank1–4*	ANK	4	0.9	–0.4	–	–	[26••]
Nank4–7*	ANK	4	2.1	–0.1	–	–	[26••]
Nank3–7*	ANK	5	1.7	1.7	–	–	[26••]
Nank2–7*	ANK	6	2.4	5.1	–	–	[26••]
Nank2–6*	ANK	5	1.7	2.0	–	–	[26••]
Nank2–5*	ANK	4	1.3	1.5	–	–	[26••]
Gankyrin	ANK	7	–	11.0	–	–	(g)
Internalin B_{248}[d]	LRR	7	7.9	5.4	42	–	[28]
Internalin B_{321}[d]	LRR	7	9.6	9.9	49	–	[28]
Tom70$^{51–617}$ [d]	TPR	7	1.2[e]	3.2[e]	41	–	[29]
Designed							
CTPR1	TPR	1	1.9	1.5	49	–	[12,22•]
CTPR2	TPR	2	2.6	6.8	74	19 860	[12,22•]
CTPR3	TPR	3	3.1	10.7	83	35 032	[12,22•]
3ANK	ANK	3	–	–	69	–	[16]
4ANK	ANK	4	–	–	81	–	[16]
2ANK TALR	ANK	2	1.3	0.8	–	–	[27]
4ANK TALR	ANK	4	1.8	12.3	–	–	[27]
E2_5	ANK	4	3.3	11.4	79	–	[24,25]
E2_17	ANK	4	3.1	9.5	70	–	[24,25]
E3_5	ANK	5	3.0	14.8	>85	–	[24,25]
E3_19	ANK	5	3.3	9.6	66	–	[24,25]
E4_8	ANK	6	5.0	21.1	79	–	[24,25]

[a]Equilibrium *m*-values, free energies of unfolding and folding rate constants were all determined by chemical denaturation/renaturation using either urea or guanidine hydrochloride. [b]A construct of p16 containing only the two C-terminal repeats. [c]Constructs of Notch representing truncations, for example, repeats 1–6. [d]Multidomain proteins. [e]Tom70 denaturation curves monitored by fluorescence can be fitted to a two-state transition, but those monitored by CD must be fitted to a multistate transition. [f]AR Lowe, LS Itzhaki, unpublished. [g]RD Hutton, LS Itzhaki, personal communication.

Magliery and Regan (unpublished) have analyzed the distribution of amino acids in different classes of repeat proteins. They found that, in all examples studied, the 'hypervariability' of certain residues defines the ligand-binding site. Experimental studies have confirmed these assignments, which were based on sequence information alone. The authors draw the comparison with antibodies, which use the same basic framework on which to present the hypervariable antigen-binding residues. Such analyses provide important guidelines for the functional design, even without structural information, of different classes of repeat proteins.

Fundamental properties of repeat proteins

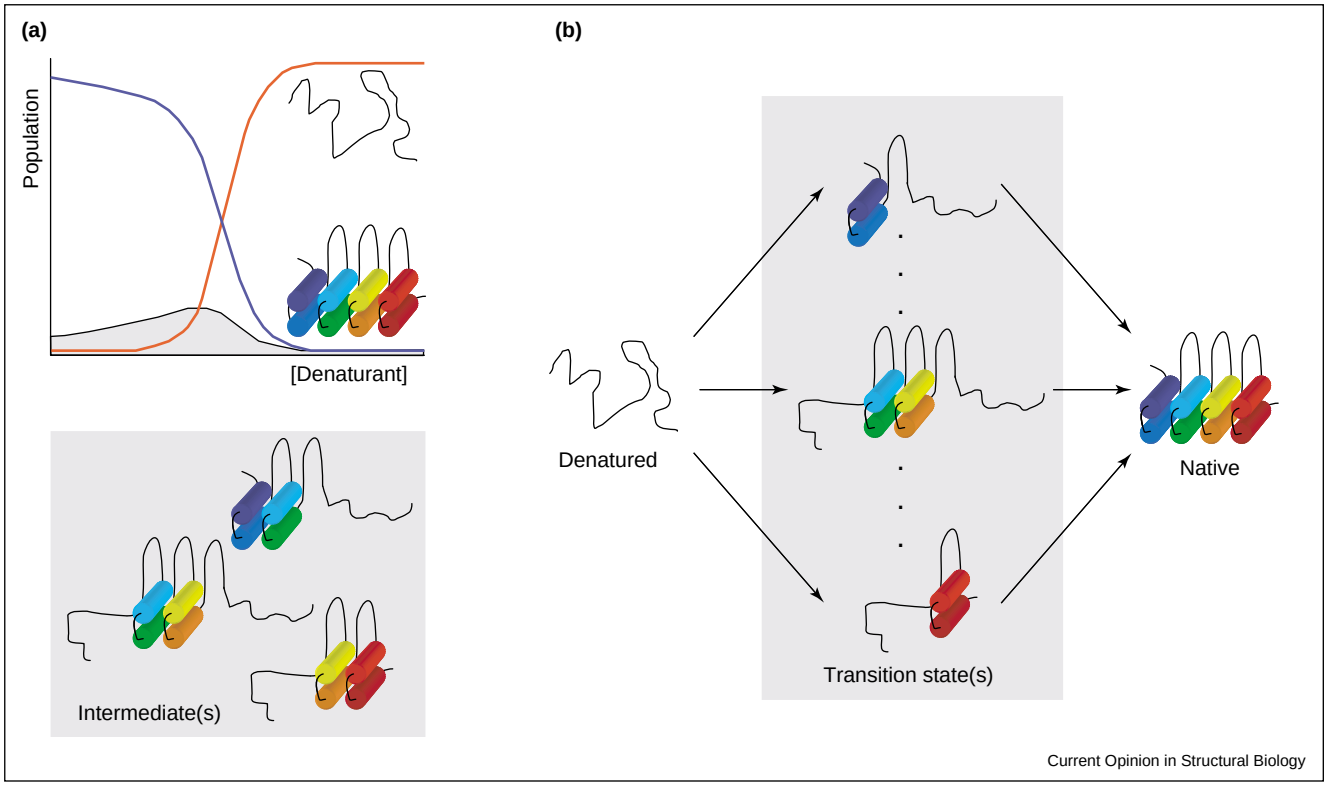
Because of their distinctive modular nature, several questions about the stability and folding of repeat proteins are currently being addressed. Are repeat proteins intrinsically more or less stable than globular proteins of the same size? Are their folding and unfolding transitions more or less cooperative than those of globular proteins (Figure 2a)? Do the folding mechanisms of repeat proteins in some way reflect their modular architecture? Because their topologies are simple and dominated by short-range interactions, is their folding unusually rapid? Also, given the structural equivalence of many sites between the repeats, do repeat proteins have multiple folding pathways (Figure 2b)? Here, we limit our discussion to ANK, LRR and TPR motifs, whose biophysical properties have been studied the most extensively.

Stability and cooperativity

Repeat proteins lack the long-range tertiary contacts that are the hallmark of globular proteins. In our previous review, we discussed how, in most cases, the stability and cooperativity of the denaturation/renaturation of repeat proteins nevertheless seems very similar to that of globular proteins [9] (Table 2).

In contrast to globular proteins, however, studies of designed and natural repeats have shown that they can

Figure 2



Schematic diagrams of hypothetical equilibrium and kinetic folding pathways. **(a)** Schematic picture of an equilibrium unfolding transition for a repeat protein. Note the partially unfolded conformations that may be present during the unfolding transition (shown in the grey box).The red line indicates the population of unfolded species, the blue line indicates the population of folded species and the shaded grey area indicates the population of partially folded structures. Derived from [26••]. **(b)** Schematic picture of a repeat protein with multiple folding pathways.

be easily lengthened or shortened by the addition or removal of individual repeat motifs [12,15,16,22•,23••, 24,25]. Moreover, the overall stability of the protein and the sharpness of its denaturation transition increase as the number of repeats increases [22•,23••,24,25,26••,27]. For both TPR proteins and ANK proteins, it has been shown that the increase in stability with increasing number of repeats is mainly manifest in a decrease in the rate of unfolding and only a marginal increase in the rate of folding ([22•]; A Plückthun *et al.*, personal communication).

Even though repeat proteins can be easily lengthened or shortened by repeat motifs, most of the repeat proteins that have been studied so far have displayed apparent two-state equilibrium unfolding behaviour. These include the larger repeat proteins gankyrin (RD Hutton, LS Itzhaki, personal communication), the LRR domain of internalin B [28] and the TPR domains of Tom70 [29], and also the different ANK and TPR consensus designs [22•,24,25,27]. However, there is evidence that the seven ANK repeat domain of Notch undergoes some fraying of the C-terminal repeats before undergoing complete unfolding [30] and that the designed LRR proteins undergo very broad unfolding transitions [15].

Recently, the question of stability versus repeat number and how cooperativity is maintained across distant repeats has been explored further by studies of both TPR [22•,23••] and ANK [26••,31••] repeat proteins. Monitoring the unfolding of a particular repeat protein by measuring its CD/fluorescence signal as a function of denaturant does not allow one to unambiguously distinguish folding that results from the collective behaviour of several units from two-state folding. However, the variation in the folding behaviour of different numbers of identical coupled subunits is a textbook signature of collective effects [32].

Kajander *et al.* [23••] characterised the thermodynamics of folding/unfolding for each member of a series of six consensus TPR proteins, for which the number of helices varies between 5 and 21. As the number of repeats increases, there is systematic variation of both the transition mid-point, which occurs at increasing guanidine hydrochloride concentrations for increasing numbers of helices, and the slope of the unfolding curve, which also increases for increasing numbers of repeats. Representing the proteins as helices that can be either folded or unfolded, and that interact only with nearest-neighbour

helices (i.e. as an Ising model), Kajander *et al.* directly fitted the Ising model predictions to their experimental data, using the same parameters for all members of the series [32,33]. The Ising model fits provide an excellent description of the experimental data for all designed TPRs in the series. Moreover, the Ising model is predictive, because thermodynamic data for just two different consensus TPRs in a series are sufficient to quantitatively predict the behaviour of all additional consensus TPRs in that series.

Barrick and colleagues have performed extensive studies that have characterised the thermodynamics of folding/unfolding by duplicating or deleting internal ANK repeats [31••], or removing the terminal ANK repeats [26••] of the seven ANK repeat containing Notch protein. A particularly interesting result from these studies was the observation of multistate folding when more than one duplicated internal ANK repeat is inserted into the protein [31••]. The inter-repeat interfaces were not optimised in these designs, which could explain the observed reduced stability and cooperativity. However, neither the consensus ANK designs of Plückthun and colleagues [24,25] or Peng and colleagues [27], nor the consensus TPR designs of Regan and colleagues [22•,23••] included a specific consideration of inter-repeat packing. Further studies of these 'mixed' ANK proteins will be of great interest. In conjunction with this study, Mello and Barrick [26••] measured the stability of a series of Notch proteins in which individual or multiple ANK repeats were deleted. These studies allowed both the intrinsic stability of different repeats and the coupling energy between repeats to be calculated. An Ising-like treatment of the data indicated that the cooperativity of this system can be explained by favourable interaction energies between the ANK units, in spite of their low intrinsic stability.

The Ising model description of the behaviour of repeat proteins [23••,26••] requires a new microscopic picture. In the usual two-state transition, a protein is essentially always either completely folded or completely unfolded, with only brief transient behaviour. By contrast, the Ising description implies that, near the transition mid-point, partially folded configurations occur with significant probability. Hydrogen/deuterium exchange studies lend support to this picture. Main *et al.* [22•] measured the native state hydrogen exchange for the consensus TPR proteins CTPR2 and CTPR3, and showed that the distribution of protection factors along the polypeptide backbone was consistent with the 'end fraying' predicted by the Ising model. NMR unfolding studies of the ANK protein p19, followed by two-dimensional ^{1}H-^{15}N HSQC, are also consistent with the existence of partially folded species at equilibrium [34].

Because of the overall structural similarities among different repeat proteins — repetition of a basic structural motif and sequentially localised contacts — it seems likely that an Ising-model-type treatment will be widely applicable to repeat proteins in general and thus represents a new folding paradigm.

Folding

For small, globular proteins with two-state folding kinetics, an approximate, inverse correlation has been observed between contact order and the rate of folding [35,36]. The designed repeats of the all-helical TPR proteins fold very rapidly, in line with what would be expected from contact order [22•]. By contrast, the folding rates of the natural repeat protein myotrophin (AR Lowe, LS Itzhaki, unpublished) and a designed three-repeat ANK protein [21•] are many orders of magnitude slower than predicted from contact order. To date, the folding pathways of three repeat proteins have been mapped in detail using Φ-value analysis: p16 [37], myotrophin (AR Lowe, LS Itzhaki, unpublished) and gankyrin (RD Hutton, LS Itzhaki, personal communication). Interestingly, the rate-determining transition state structures for folding were all found to be primarily polarised at the C-terminal repeats. The other striking feature, apparent for both myotrophin and gankyrin, although not yet probed for p16, is that at least two folding pathways are accessible, with transition state structures polarised at one or the other end of the molecule. Although the concept of pathway heterogeneity (Figure 2b) evolves naturally from the 'new view' of protein folding based on energy landscapes, such heterogeneity has been detected for few two-state folding globular proteins (e.g. [38]). The observation of folding heterogeneity for both of the ANK proteins studied so far suggests that the availability of more than one route to the native state may be a general feature of the repeating nature of these structures — leading to multiple potential nucleation sites of similar energy throughout the protein (Figure 2b). Further folding studies, on different classes of repeat proteins, will clearly be of great interest.

Conclusions

In this review, we have described exciting recent studies of repeat proteins. It is clear that repeat proteins are extremely well suited as scaffolds on which to introduce novel binding specificities and constitute excellent systems in which to study structure/sequence relationships. In addition, their equilibrium unfolding behaviour has been shown to be well described by remarkably simple quantitative models. Finally, their modular nature also seems to be reflected in their folding kinetics, with examples whereby more than one folding pathway can be accessed that has transition state structures with very different nucleation points.

Acknowledgements

We thank Tamara Chiba, Aitziber Lopez Cortajarena, Richard Hutton, Tommi Kajander, Thomas Magliery, Dorina Saro, Chris Wilson and Fang Yi for sharing results, and for extensive and enlightening discussions of repeat protein structure, function and stability. Ewan Main is funded by a Girton Research Fellowship, University of Cambridge. Alan Lowe

is funded by the Medical Research Council, UK. Both Sophie Jackson and Ewan Main acknowledge the support of the Welton Foundation.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Vendruscolo M, Dobson CM: **A glimpse at the organization of the protein universe**. *Proc Natl Acad Sci USA* 2005, **102**:5641-5642.

2. Chukkapalli G, Guda C, Subramaniam S: **SledgeHMMER: a web server for batch searching the Pfam database**. *Nucleic Acids Res* 2004, **32**:W542-W544.

3. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL *et al.*: **The Pfam protein families database**. *Nucleic Acids Res* 2004, **32**:D138-D141.

4. Andrade MA, Perez-Iratxeta C, Ponting CP: **Protein repeats: structures, functions, and evolution**. *J Struct Biol* 2001, **134**:117-131.

5. D'Andrea LD, Regan L: **TPR proteins: the versatile helix**. *Trends Biochem Sci* 2003, **28**:655-662.

6. Kobe B, Kajava AV: **The leucine-rich repeat as a protein recognition motif**. *Curr Opin Struct Biol* 2001, **11**:725-732.

7. Kobe B, Kajava AV: **When protein folding is simplified to protein coiling: the continuum of solenoid protein structures**. *Trends Biochem Sci* 2000, **25**:509-515.

8. Sedgwick SG, Smerdon SJ: **The ankyrin repeat: a diversity of interactions on a common structural framework**. *Trends Biochem Sci* 1999, **24**:311-316.

9. Main ERG, Jackson SE, Regan L: **The folding and design of repeat proteins: reaching a consensus**. *Curr Opin Struct Biol* 2003, **13**:482-489.

10. Jawad Z, Paoli M: **Novel sequences propel familiar folds**. *Structure* 2002, **10**:447-454.

11. • Wilson CGM, Kajander T, Regan L: **The crystal structure of N1pl. A prokaryotic tetratricopeptide repeat protein with a globular fold**. *FEBS J* 2005, **272**:166-179.
An interesting study showing how a protein completely composed of TPR motifs can fold into a globular structure.

12. Main ERG, Xiong Y, Cocco MJ, D'Andrea L, Regan L: **Design of stable alpha-helical arrays from an idealized TPR motif**. *Structure* 2003, **11**:497-508.

13. Forrer P, Stumpp MT, Binz HK, Pluckthun A: **A novel strategy to design binding molecules harnessing the modular nature of repeat proteins**. *FEBS Lett* 2003, **539**:2-6.

14. Forrer P, Binz HK, Stumpp MT, Pluckthun A: **Consensus design of repeat proteins**. *ChemBioChem* 2004, **5**:183-189.

15. Stumpp MT, Forrer P, Binz HK, Pluckthun A: **Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family**. *J Mol Biol* 2003, **332**:471-487.

16. Mosavi LK, Minor DL Jr, Peng ZY: **Consensus-derived structural determinants of the ankyrin repeat motif**. *Proc Natl Acad Sci USA* 2002, **99**:16029-16034.

17. • Magliery TJ, Regan L: **Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif**. *J Mol Biol* 2004, **343**:731-745.
A comprehensive study describing how multiple alignments can be used to obtain a greater understanding of repeat protein structure/function/stability relationships than mere consensus design alone.

18. • Cortajarena AL, Kajander T, Pan WL, Cocco MJ, Regan L: **Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins**. *Protein Eng Des Sel* 2004, **17**:399-409.
A paper that describes the redesign of a consensus TPR design to obtain a functional protein that binds a cognate ligand.

19. •• Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grutter MG, Pluckthun A: **High-affinity binders selected from designed ankyrin repeat protein libraries**. *Nat Biotechnol* 2004, **22**:575-582.
An exciting study that describes and shows, using crystal structures, how ANK repeat libraries can be selected to bind novel protein targets (in a manner akin to antibodies).

20. •• Amstutz P, Binz HK, Parizek P, Stumpp MT, Kohl A, Grutter MG, Forrer P, Pluckthun A: **Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins**. *J Biol Chem* 2005, **280**:24715-24722.
Following on from the work described in [19••], this study reports the selection of ANK repeat proteins to inhibit intracellular kinases.

21. • Devi VS, Binz HK, Stumpp MT, Pluckthun A, Bosshard HR, Jelesarov I: **Folding of a designed simple ankyrin repeat protein**. *Protein Sci* 2004, **13**:2864-2870.
One of the first papers, together with [22•], to describe the folding of a designed repeat protein.

22. • Main ER, Stott K, Jackson SE, Regan L: **Local and long-range stability in tandemly arrayed tetratricopeptide repeats**. *Proc Natl Acad Sci USA* 2005, **102**:5721-5726.
This interesting study shows the changes in hydrogen/deuterium exchange, stability and folding kinetics when designed repeat proteins increase in size by adding consensus TPR motifs.

23. •• Kajander T, Cortajarena AL, Main ERG, Mochrie SGJ, Regan L: **A new paradigm for repeat protein folding**. *J Am Chem Soc* 2005, in press.
An exciting study showing that the unfolding equilibrium denaturation of a series of designed TPR proteins can be described and predicted using an Ising model.

24. Kohl A, Binz HK, Forrer P, Stumpp MT, Pluckthun A, Grutter MG: **Designed to be stable: crystal structure of a consensus ankyrin repeat protein**. *Proc Natl Acad Sci USA* 2003, **100**:1700-1705.

25. Binz HK, Stumpp MT, Forrer P, Amstutz P, Pluckthun A: **Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins**. *J Mol Biol* 2003, **332**:489-503.

26. •• Mello CC, Barrick D: **An experimentally determined protein folding energy landscape**. *Proc Natl Acad Sci USA* 2004, **101**:14102-14107.
This important paper explores cooperativity within the seven ANK repeat Notch protein. It dissects the energetics into 'intrinsic' free energies of the individual repeats and 'coupling' free energies between them.

27. Mosavi LK, Peng ZY: **Structure-based substitutions for increased solubility of a designed protein**. *Protein Eng* 2003, **16**:739-745.

28. Freiberg A, Machner MP, Pfeil W, Schubert WD, Heinz DW, Seckler R: **Folding and stability of the leucine-rich repeat domain of internalin B from *Listeria monocytogenes***. *J Mol Biol* 2004, **337**:453-461.

29. Beddoe T, Bushell SR, Perugini MA, Lithgow T, Mulhern TD, Bottomley SP, Rossjohn J: **A biophysical analysis of the tetratricopeptide repeat-rich mitochondrial import receptor, Tom70, reveals an elongated monomer that is inherently flexible, unstable, and unfolds via a multistate pathway**. *J Biol Chem* 2004, **279**:46448-46454.

30. Bradley CM, Barrick D: **Limits of cooperativity in a structurally modular protein: response of the Notch ankyrin domain to analogous alanine substitutions in each repeat**. *J Mol Biol* 2002, **324**:373-386.

31. •• Tripp KW, Barrick D: **The tolerance of a modular protein to duplication and deletion of internal repeats**. *J Mol Biol* 2004, **344**:169-178.
An interesting study probing the limits of cooperativity within the seven ANK repeat Notch protein.

32. Nelson P: *Biological Physics: Energy, Information, Life.* New York: WH Freeman; 2003.

33. Zimm BH, Bragg JK: **Theory of the phase transition between helix and random coil in polypeptide chains**. *J Chem Phys* 1959, **31**:526-535.

34. Zeeb M, Rosner H, Zeslawski W, Canet D, Holak TA, Balbach J: **Protein folding and stability of human CDK inhibitor p19(INK4d)**. *J Mol Biol* 2002, **315**:447-457.

35. Plaxco KW, Simons KT, Baker D: **Contact order, transition state placement and the refolding rates of single domain proteins**. *J Mol Biol* 1998, **277**:985-994.

36. Plaxco KW, Simons KT, Ruczinski I, Baker D: **Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics**. *Biochemistry* 2000, **39**:11177-11183.

37. Tang KS, Fersht AR, Itzhaki LS: **Sequential unfolding of ankyrin repeats in tumor suppressor p16**. *Structure* 2003, **11**:67-73.

38. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J: **Parallel protein-unfolding pathways revealed and mapped**. *Nat Struct Biol* 2003, **10**:658-662.

39. Michaely P, Tomchick DR, Machius M, Anderson RG: **Crystal structure of a 12 ANK repeat stack from human ankyrinR**. *EMBO J* 2002, **21**:6387-6396.

40. Jinek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E: **The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha**. *Nat Struct Mol Biol* 2004, **11**:1001-1007.

41. Liou YC, Tocilj A, Davies PL, Jia Z: **Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein**. *Nature* 2000, **406**:322-324.

42. Daley ME, Graether SP, Sykes BD: **Hydrogen bonding on the ice-binding face of a beta-helical antifreeze protein indicated by amide proton NMR chemical shifts**. *Biochemistry* 2004, **43**:13012-13017.

43. Russo AA, Tong L, Lee JO, Jeffrey PD, Pavletich NP: **Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a**. *Nature* 1998, **395**:237-243.

44. Jeffrey PD, Tong L, Pavletich NP: **Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors**. *Genes Dev* 2000, **14**:3115-3125.

45. Jacobs MD, Harrison SC: **Structure of an IκBα/NF-κB complex**. *Cell* 1998, **95**:749-758.

46. Byeon IJ, Li J, Ericson K, Selby TL, Tevelev A, Kim HJ, O'Maille P, Tsai MD: **Tumor suppressor p16INK4A: determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4**. *Mol Cell* 1998, **1**:421-431.

47. Huber AH, Weis WI: **The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin**. *Cell* 2001, **105**:391-402.

48. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J: **Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha**. *Cell* 1998, **94**:193-204.

49. Lee SJ, Sekimoto T, Yamashita E, Nagoshi E, Nakagawa A, Imamoto N, Yoshimura M, Sakai H, Chong KT, Tsukihara T *et al.*: **The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor**. *Science* 2003, **302**:1571-1575.

50. Bayliss R, Littlewood T, Stewart M: **Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking**. *Cell* 2000, **102**:99-108.

51. Wang XG, Olsen LR, Roderick SL: **Structure of the lac operon galactoside acetyltransferase**. *Structure* 2002, **10**:581-588.

52. Papageorgiou AC, Shapiro R, Acharya KR: **Molecular recognition of human angiogenin by placental ribonuclease inhibitor – an X-ray crystallographic study at 2.0 Å resolution**. *EMBO J* 1997, **16**:5162-5177.

53. Liker E, Fernandez E, Izaurralde E, Conti E: **The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain**. *EMBO J* 2000, **19**:5587-5598.

54. Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, Hartl FU, Moarefi I: **Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine**. *Cell* 2000, **101**:199-210.

55. Wall MA, Coleman DE, Lee E, Iniguez-Lluhi JA, Posner BA, Gilman AG, Sprang SR: **The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2**. *Cell* 1995, **83**:1047-1058.

56. Paoli M, Anderson BF, Baker HM, Morgan WT, Smith A, Baker EN: **Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two beta-propeller domains**. *Nat Struct Biol* 1999, **6**:926-931.

57. Zhang B, Peng Z: **A minimum folding unit in the ankyrin repeat protein p16(INK4)**. *J Mol Biol* 2000, **299**:1121-1132.

58. Tang KS, Guralnick BJ, Wang WK, Fersht AR, Itzhaki LS: **Stability and folding of the tumour suppressor protein p16**. *J Mol Biol* 1999, **285**:1869-1886.

59. Yuan C, Li J, Selby TL, Byeon IJ, Tsai MD: **Tumor suppressor INK4: comparisons of conformational properties between p16(INK4A) and p18(INK4C)**. *J Mol Biol* 1999, **294**:201-211.

60. Mosavi LK, Williams S, Peng Z: **Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein**. *J Mol Biol* 2002, **320**:165-170.

61. Zweifel ME, Barrick D: **Studies of the ankyrin repeats of the *Drosophila melanogaster* Notch receptor. 2. Solution stability and cooperativity of unfolding**. *Biochemistry* 2001, **40**:14357-14367.

62. Zweifel ME, Barrick D: **Studies of the ankyrin repeats of the *Drosophila melanogaster* Notch receptor. 1. Solution conformational and hydrodynamic properties**. *Biochemistry* 2001, **40**:14344-14356.