The folding and design of repeat proteins: reaching a consensus

Ewan RG Main*†, Sophie E Jackson† and Lynne Regan*‡

Although they are widely distributed across kingdoms and are involved in a myriad of essential processes, until recently, repeat proteins have received little attention in comparison to globular proteins. As the name indicates, repeat proteins contain strings of tandem repeats of a basic structural element. In this respect, their construction is quite different from that of globular proteins, in which sequentially distant elements coalesce to form the protein. The different families of repeat proteins use their diverse scaffolds to present highly specific binding surfaces through which protein–protein interactions are mediated. Recent studies seek to understand the stability, folding and design of this important class of proteins.

Addresses
*Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA
†Cambridge University Chemical Laboratory, Cambridge University, Lensfield Road, Cambridge CB2 1EW, UK
‡Department of Chemistry, Yale University, New Haven, CT 06520, USA
e-mail: lynne.regan@yale.edu

Current Opinion in Structural Biology 2003, **13**:482–489

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/S0959-440X(03)00105-2

Abbreviations

ank	ankyrin
CD	circular dichroism
HPR	hexapeptide repeat
LRR	leucine-rich repeat
TPR	tetratricopeptide repeat

Introduction

During the past few years, several exciting studies have emerged that begin to characterize a hitherto understudied but ubiquitous class of proteins. These nonglobular folds, called repeat or solenoid proteins, are formed from repeated motifs of between 20 and 40 amino acids that stack together to produce extended and striking superhelical structures [1,2]. The repetitive and elongated nature of repeat proteins causes them to differ radically in their construction from normal globular proteins; repeat proteins are dominated by short-range and regularised interactions (Figure 1a,c,d), whereas globular proteins exhibit complex topologies that frequently have numerous long-range interactions (Figure 1b). In this review, we discuss the recent progress made in the characterisation and design of repeat proteins, and explore the similarities and differences between repeat and globular proteins.

Repeat proteins: structure and function

There are approximately 20 classes of repeat protein for which the three-dimensional structure of at least one representative has been solved [3,4•]. In each class, a repeated 20–40 amino acid motif defines a variety of secondary structural units. These range from simple motifs that consist of two linked secondary structure components to more complex motifs that include many more. Some examples are:

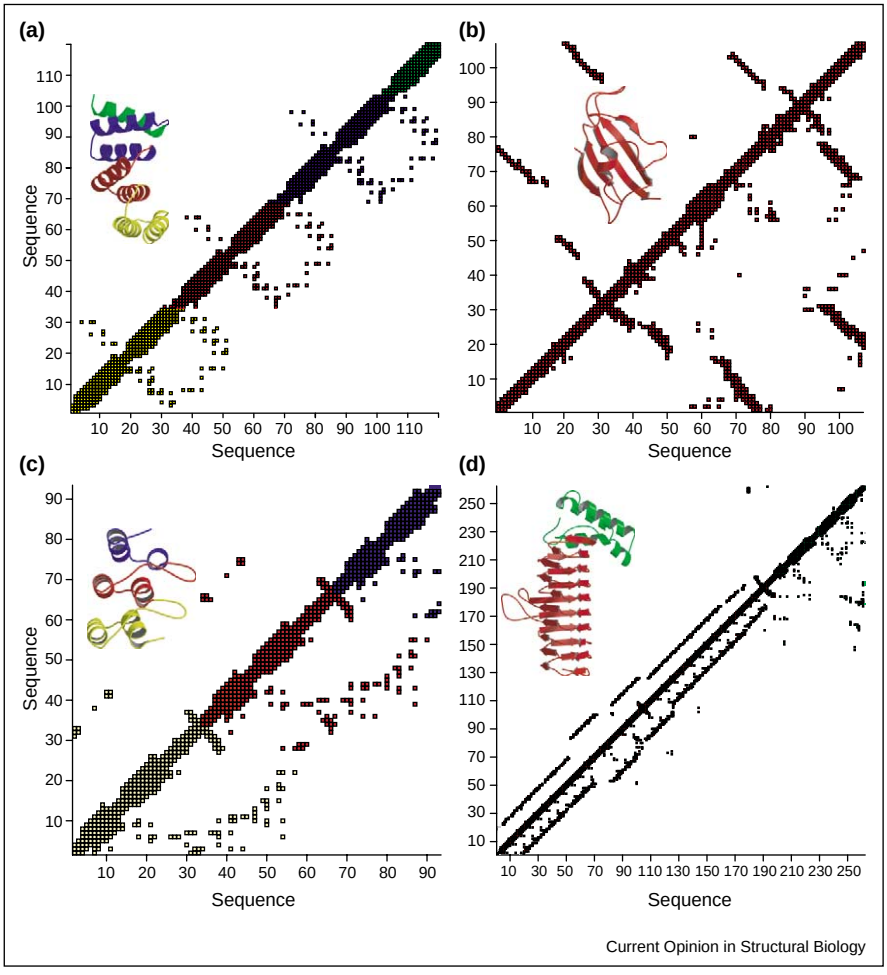
1. α helix – link – α helix; for example, tetratricopeptide (TPR) and HEAT repeats.
2. α helix – link – β strand; for example, leucine-rich repeats (LRRs).
3. α helix – β hairpin or loop – α helix; for example, ankyrin (ank) repeats.
4. $\{\beta \text{ strand} - \text{link} - \beta \text{ strand}\}_3$; for example, hexapeptide repeats (HPRs).
5. $\{\beta \text{ strand} - \text{link} - \beta \text{ strand}\}_4$; for example, WD40 (40-residue repeat with Trp [W] – Asp [D] motif).

When these structural units stack together to form the complete, functional protein, two main categories of interaction are paramount in defining the final characteristic elongated tertiary structure: the packing within each repeat motif; and stacking interactions between adjacent repeat motifs (Figure 1). It appears that tandem arrays of multiple repeats are required for the formation of stable folded structures, because no single repeat motif has yet been shown to form a stable folded unit. In theory, there should be no limit to the number of repeats possible within a domain [5]; for example, 29 ank repeats have been identified in NOMPC (an ion channel of 1619 amino acids) [6]. However, the intricate relationship between stability, repeat motif number and function is still unclear.

The profusion and thus evolutionary success of repeat proteins (six families are present in the top 20 families in the PFAM database [7]) may be explained through:

1. The relative ease of evolution — large, stable proteins can be produced from simple duplication events, as opposed to the more complex creation of secondary structure *de novo* [5].
2. A capacity to acquire diverse functions in many cellular processes — for example, the TPR motif is linked to functions as varied as cell cycle regulation, transcriptional control, protein transport, neurogenesis and assisting protein folding [8].

Figure 1



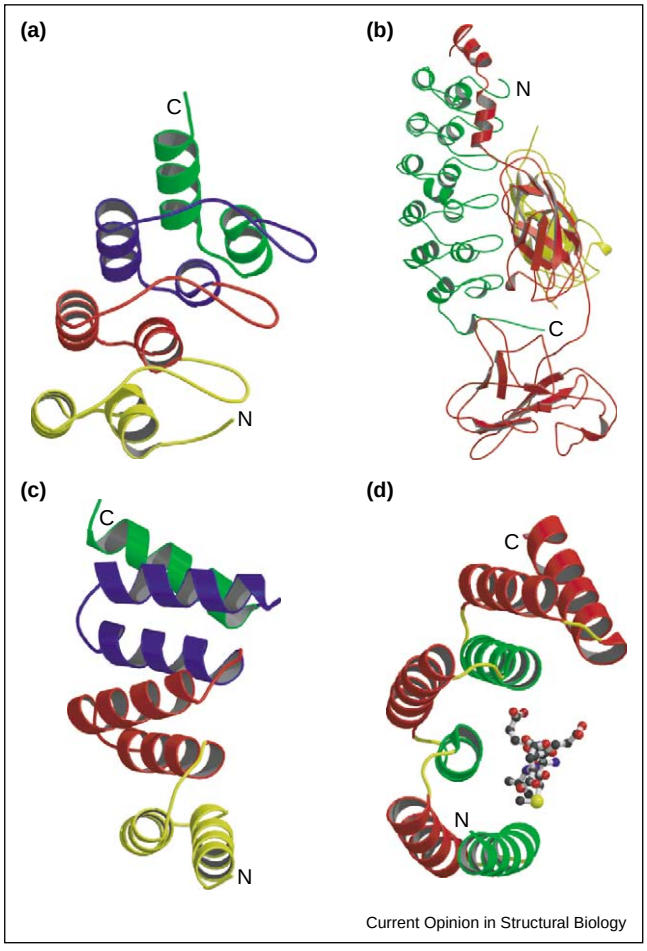
A comparison of the contact maps of **(a)** CTPR3 — a designed repeat protein containing 3.5 consensus TPR motifs, **(b)** FKBP12 – a globular protein of 107 amino acids, **(c)** 3ANK – a designed repeat protein containing 3 consensus ank motifs and **(d)** *N*-acetylglucosamine acyltransferase – a protein that contains approximately 9 HPRs. The maps were produced using the program MOLMOL [32]. A ribbon representation of each protein is shown in the top left-hand corner (created using MOLSCRIPT [33] and rendered using RASTER 3D [34]). The x and y axes both show the residue numbers for the complete protein sequence. A square is placed at each position where a residue is within 5 Å of another residue. The diagonal represents self and local sequence contacts. The off-diagonal points represent longer-range interactions. Those above the diagonal represent only backbone contacts and those below represent all types of contact. Certain plots are colour coded. In (a,c), yellow denotes the first repeat, red for the second, blue for the third and green for the fourth. In (d), red denotes the HPR domain and green denotes the globular domain. As one can see, the repeat proteins in (a,c,d) are dominated by regular short-range interactions, whereas the globular protein (b) has numerous more complex long-range interactions.

The underlying link between repeat proteins in all such processes is a shared ability to mediate protein–protein interactions. The repeated units form a scaffold that utilizes their extended surfaces to expose a specific interacting interface. However, as one would expect, different repeats provide very different interfaces that bind a variety of partners. These range from small-peptide-binding pockets, such as the pentapeptide-binding pocket in Hop [9] (Figure 2), to extended binding surfaces in which multiple interaction sites are used to bind globular proteins [10] (Figure 2). The ability of repeat protein families to bind multiple binding partners on essentially similar scaffolds, combined with their simple modular nature, makes them ideal candidates for protein design. Furthermore, their novel nonglobular structures raise interesting questions concerning the factors that contribute to their stability and folding pathways. In this review, we will mainly focus on the ank and TPR α-helical repeat proteins, which represent the best-characterized systems.

Ank and TPR α-helical repeat proteins

The ank and TPR motifs are composed of repeating units of 33 and 34 amino acids, respectively. Both repeated motifs encode a pair of α helices. In each case, although

Figure 2



Four illustrations showing the structures of **(a)** a designed repeat protein called 4ANK, which contains 4 consensus ank repeats [28••]; **(b)** a complex showing the 6 ank repeats (green) of NF-κB–IκBα interacting with NF-κB (yellow) [10]; **(c)** a designed repeat protein called CTPR3, which contains 3.5 consensus TPR motifs [27••]; and **(d)** a complex showing the 3 TPRs of Hop TPR2A interacting with the C-terminal pentapeptide from Hsp90 [9]. All structures are shown as ribbon representations, except the pentapeptide in (d), which is rendered in ball and stick. Each figure is colour coded to highlight differing structural features. In (a,c), each repeat is coloured differently — yellow denotes the first repeat, red for the second, blue for the third and green for the fourth. In (b), the interacting ank repeats are coloured green and the other proteins in the complex are coloured red and yellow. In (d), the α helices of Hop that interact with the substrate are coloured green. All pictures were created using MOLSCRIPT [33] and rendered using RASTER 3D [34].

the repeating sequences are highly degenerate, with no position invariant, there is a consistent pattern of key residues that are essential to the structural integrity of each fold. These residues define the fold and contribute to overall stability. In addition, a second subset of residues can be defined that vary in response to the specific protein–protein interaction. Despite such similarities, there are several important differences between the two types of repeats.

Ank motif

The two antiparallel α helices within the ank motif are linked by a β hairpin/loop that projects outward from the helices at angle of about 90°. The β hairpin is approximately perpendicular to the α helices, producing an L-shaped conformation. In nearly all ank-mediated protein–protein interactions, the binding surface is primarily composed of the β-hairpin/loop region [1]. However, certain complexes also use the surfaces of the inner α helices. Tandem ank repeats stack in parallel, causing the helices in one repeat to stack directly over those in adjacent positions. The stacking interactions also produce a slight curve and left-handed twist in the overall structure. Figure 2a shows the ank fold and Figure 2b shows the ank repeats of NF-κB–IκBα in complex with its substrate.

TPR motif

In contrast to the ank motif, the two helices (A and B) of the TPR motif are connected by a tight turn. Tandem TPRs also stack in parallel, but form a very different structure to the ank proteins. A highly conserved pattern of small and large hydrophobic residues defines the TPR fold by causing the A and B helices within one TPR unit to pack at an angle of approximately 160° and force successive repeated units to stack in a fashion reminiscent of steps in a spiral staircase. In general, the stacking of the repeats forms a right-handed superhelix (with a repeat of approximately eight TPR motifs) with little overall curvature; however, when the structure of PEX5 (seven TPRs) with its peptide substrate was solved, it formed two three-repeat modules that enclose the substrate like a clam [11]. The major interaction surface of TPR proteins seems to be formed from a groove exposed by the superhelical twist, which is composed of residues on the A helix of each repeat. Figure 2c shows the TPR fold and Figure 2d shows three TPR motifs of Hop in complex with its substrate, the C-terminal peptide of Hsp90.

Comparing globular and repeat proteins

The nonglobular nature of linear repeat proteins has stimulated several productive studies on the factors contributing to their stability and folding. The main body of this work has centred on protein domains containing ank repeats (although one parallel β-helix fold motif, similar to the HPR shown in Figure 1d, has also been characterized). The stability and folding of the following proteins have been studied: $p16^{INK4}$, four ank repeats [12••,13••, 14,15]; myotrophin, four ank repeats [16]; $p19^{INK4d}$, five ank repeats [17]; ank repeats from the notch receptor of *Drosophila melanogaster*, seven ank repeats [18•–20•]; and pectate lyase C, approximately eight three β-strand repeats [21–24].

Equilibrium unfolding

The first major question to be answered was whether or not the modular nature of repeat proteins is reflected in their unfolding behaviour under equilibrium conditions.

To date, similar to the behaviour of most single-domain globular proteins, repeat proteins display a cooperative or 'two-state' equilibrium unfolding transition for both heat- and chemical-induced denaturation [14,16,17,20•,21]. In contrast to globular proteins and consistent with their modular nature, studies have shown that it is possible to remove varying numbers of repeats from a repeat protein and still produce stable, folded fragments [12••,19•]. In some cases, the addition or removal of the N/C-terminal repeats can have a more drastic effect on stability/helicity than the removal of internal repeats [12••,19•,20•]. This could be due to the introduction of capping interactions that close the hydrophobic core from solvent. Furthermore, when Bradley and Barrick [18•] engineered a series of Ala→Gly substitutions at conserved alanine residues within each repeat of the Notch domain (seven anks), one mutant seemed to uncouple the cooperative two-state folding. It was postulated that the mutation, in repeat 6, could destabilize the terminal repeats, thereby uncoupling them from the folding of the rest of the protein.

Response to mutation

Several studies have engineered single and multiple site mutations into ank-containing proteins [13••,14,15,18•]. In general, the response of the proteins to mutation is equivalent to that seen in globular proteins. Of particular interest is the study by Tang *et al.* [14], who investigated the effect of oncogenic mutations on the tumour suppressor p16. They found that any oncogenic mutation, coupled with p16's already low stability, caused both aggregation and unfolding. In related work, reminiscent of p53 rescue [25], Cammett *et al.* [15] showed that the deleterious effect of the oncogenic mutations could be reversed by designing secondary stabilizing mutations elsewhere in the protein.

Folding pathways

Because repeat proteins are constructed as a modular array, one might envisage that each repeat could form separate folding units (foldons), which then coalesce to form the final structure [26]. Certainly, studies on p19, pectate lyase C and p16 have shown the population of intermediates during folding [14,17,23]. In the equilibrium unfolding of p19, probed by 2D ^{1}H-^{15}N HSQC NMR spectroscopy, an intermediate species was observed, even though there seemed to be little evidence of such a state from kinetic studies. When pectate lyase C was studied, using fluorescence and near- and far-UV CD as probes, the temperature dependence of the folding rate constant suggested that an on-pathway intermediate, in rapid equilibrium with the unfolded protein, may be present [23]. In kinetic studies of p16, using fluorescence as a probe, at least one intermediate state was observed to accumulate during folding. Subsequently, Tang *et al.* [13••] rigorously characterised the folding pathway of p16 by performing Φ-value analysis of the major transition state for folding. This was the first such analysis of the folding of a repeat protein and showed that each repeat unfolded sequentially: the two N-terminal repeats unfold first, followed by the unfolding of the two C-terminal repeats. It is interesting to note that, in another study, the smallest fully folded fragment of p16 that could be produced corresponded to the two C-terminal repeats [12••]. Thus, in this case, it appears that the most stable repeat folds early.

Designing repeats

The attraction of repeat proteins to design can be easily explained. It stems from the ability of each repeat family to mediate a host of protein–protein interactions from a scaffold built from a simple repeated module. Furthermore, using multiple sequence alignment and current structural characterization, we can start to delineate those residues within the repeated module that are responsible for fold preservation or substrate interaction (Figure 3). Thus, we should be able to either modify existing repeat functions or create repeat proteins into which we can design novel binding properties. Excitingly, in the past year, there has been one successful design of novel TPR proteins [27••] and two successful designs of novel ank repeat proteins [28••,29••]. These designs, coupled with detailed solution characterisation and high-resolution X-ray crystal structures (Figures 1, 2 and 4), described below, have expanded our current knowledge of repeat proteins, given insight into the relationship between structure and sequence, and posed several interesting questions.

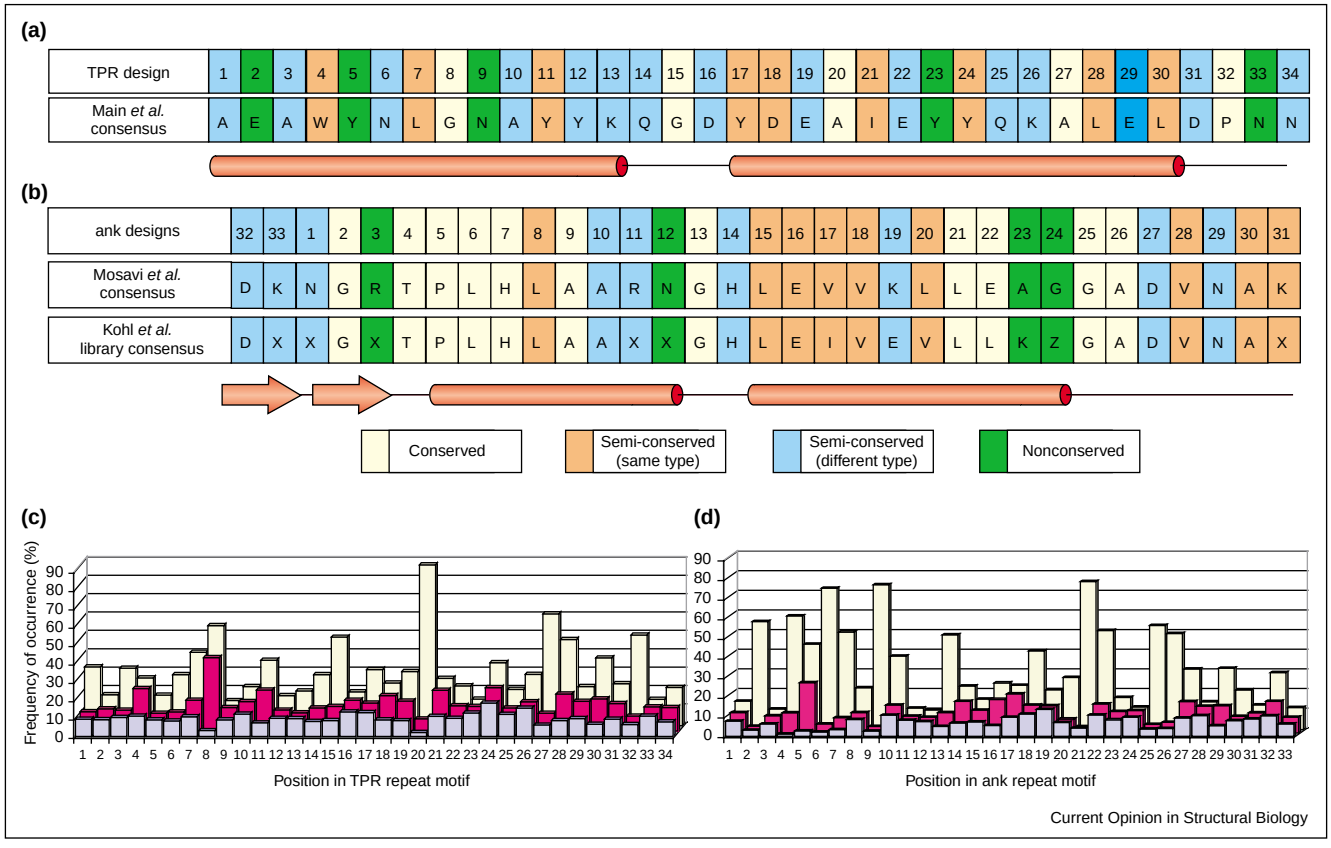
Reaching a consensus — basic design

To successfully design a repeat protein, one must first identify those residues within each repeat that define the fold. One way to achieve this is to use the consensus amino acid at each position — usually the most frequently occurring (Figure 3). Repeat proteins, with their extensive protein sequence databases, are ideally suited to such analysis, as large data sets are required for accurate results. Moreover, the lower the variance of an amino acid or type of amino acid at a given position, the more likely it is to be key to fold conservation. In the successful TPR and ank designs, a single idealised repeat was obtained using a consensus design strategy. However, in each case, a slightly different statistical analysis was performed to obtain the consensus repeat. These are summarized in Table 1 and shown in Figure 3. It is interesting to note that these designs do not explicitly take account of co-variance. However, if such analysis is performed [28••], nearly all of the co-variant residues, particularly in the hydrophobic core, are present. This presumably reflects the essential sidechain complementarity required to specify the fold.

From single to multiple tandem repeats. When are stable, folded proteins produced?

To obtain the final designed repeat proteins, the single consensus repeat produced in each study was duplicated to generate tandem arrays. In the study by Main *et al.*

Figure 3



The final sequences of the idealized **(a)** TPR and **(b)** ank repeats for each design. Each representation shows the conservation level of the selected sequence and a schematic of the secondary structural elements they encode. Positions that are more than 50% conserved are classed as 'conserved' and coloured yellow, positions where 2–4 similar amino acids dominate are classed as 'semi-conserved (same type)' and coloured orange, positions where 2–4 different amino acids dominate are classed as 'semi-conserved (different type)' and coloured blue, and those positions that have no preference for an amino acid are classed as 'nonconserved' and are coloured green. In the Kohl *et al*. [29••] consensus, an X denotes any amino acid except cysteine, glycine or proline, and a Z denotes histidine, asparagine or tyrosine amino acids. A statistical analysis showing the three most frequent amino acids at each position of **(c)** the TPR and **(d)** the ank repeat motifs. The most frequent is coloured yellow, second most frequent is coloured pink and third most frequent is coloured blue.

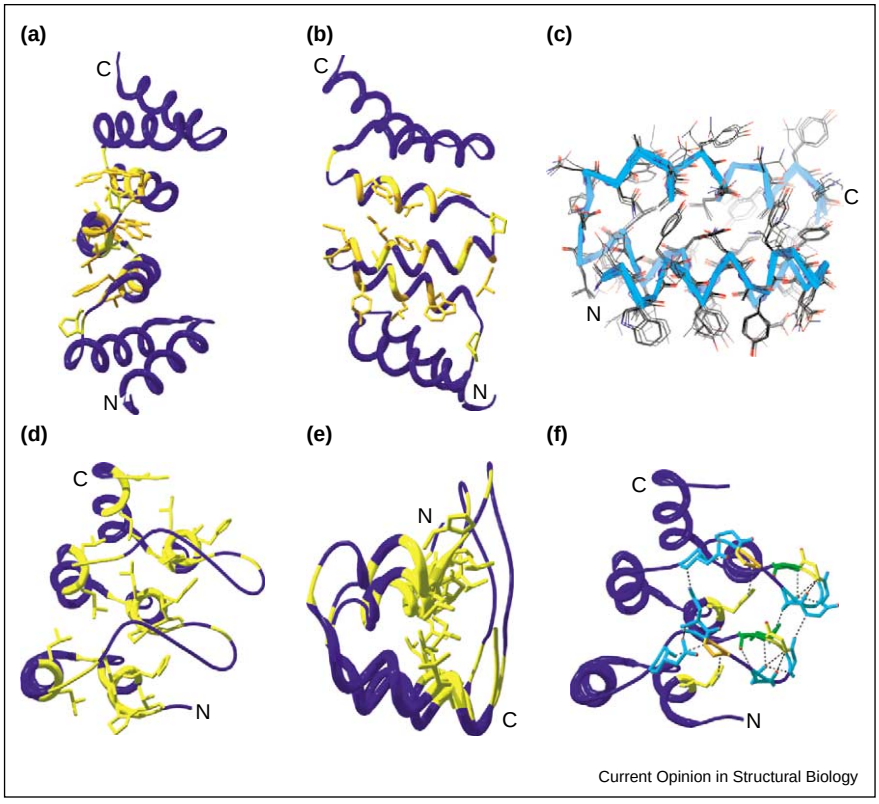
[27••], this corresponded to 1.5, 2.5 and 3.5 TPR motifs. Mosavi *et al*. [28••] produced 1, 2, 3 and 4 ank motifs, whereas Kohl *et al*. [29••] produced libraries of 4–6 ank repeats, in which the N- and C-terminal repeats were designed to cap the protein, and the middle repeats contained a backbone consensus with certain regions randomised [29••]. In each case, to determine whether the structures were properly folded, the designed proteins were characterized using fluorescence, far- and near-UV CD and NMR. High-resolution X-ray crystallographic structures were also obtained for the 2.5 and 3.5 TPR motif proteins [27••], 3 and 4 ank motif proteins [28••] and one library member containing 5 ank repeats [29••].

For the TPR and Mosavi's ank study, a critical number of repeats had to be reached before a correctly folded, monomeric repeat protein was produced. In Nature, both TPR- and ank-containing domains rarely contain less than three tandem motifs. Therefore, by building both TPR and ank repeat motifs from single units, these studies were able to probe whether the minimum repeat number reflects a necessary structural or functional requirement. For the TPR design, only 1.5 repeats were required to generate a correctly folded structure, whereas 3 repeats were needed for the ank design. This difference would seem to stem from differences in the packing of the repeating units. The TPR fold has a more distinct modularity, composed of 1.5 repeats, which make more contacts within the module than the rest of the protein (Figure 1). Although the ank repeat does not possess such clear modularity, if the contact maps of 3ANK (3 repeats) are studied, one can imagine a module encompassing 2 ank repeats (Figure 1). This is consistent with the result that the 2 ank repeat is stable but not monomeric.

Consensus designs — more stable than nature

Similar to their natural homologues, equilibrium unfolding by heat or chemical denaturation of the consensus

Figure 4



Ribbon representations of the X-ray crystal structures of **(a–c)** the designed TPR protein CTPR3 [27••] and **(d,e)** the designed ank protein 3ANK [28••]. (a,b,d,e) The structural roles of the most conserved 'signature' residues in each motif are illustrated. These sidechains are represented as sticks and coloured by level of conservation — residues more than 50% conserved are yellow and semi-conserved residues of the same type are orange. In the case of the CTPR3 protein, only the conserved residues from the middle A-B-A′ helices of the second and third TPR motifs are shown. (c) The identical nature of the designed repeats is revealed by five overlaid structures of the A-B-A′ motif of CTPR2 (two motifs) and CTPR3 (three motifs). The sidechains and Cα backbone are represented by black lines and a blue tube, respectively. **(f)** The hydrogen-bonding network in the β-hairpin/loop region of 3ANK. Each residue is coloured by level of conservation – residues more than 50% conserved are yellow, semi-conserved residues of the same type are orange, semi-conserved residues of a different type are blue and nonconserved residues are green. All figures produced using Swiss-PDBViewer [35] and rendered with the program POV-Ray [36].

proteins was found to be two state, rather than multistate. Furthermore, as the number of repeats increased, so did the stability. Interestingly, when the consensus proteins were matched to comparable natural repeat-containing proteins, they were found to have greatly increased thermal and chemical stabilities. This increase is probably due to a combination of factors, such as the removal of unfavourable interactions (possibly present as a consequence of functional requirements), the absence of irregular insertions or deletions in a repeat, and optimised intra-repeat and inter-repeat packing (because the repeating units are identical). The increased stability over natural counterparts mirrors the higher stabilities observed when consensus designs have been used to generate enzymes [30] and antibodies [31]. It also illustrates that, although evolution selects for function, it rarely optimises stability more than is necessary for a given milieu.

Structural characteristics

As stated, the structures that were solved were 2.5 and 3.5 TPR motifs [27••], 3 and 4 ank repeats [28••], and 5 ank repeats [29••]. One striking outcome of the consensus design and repeat fold is the identical interactions and structure of each repeat within the consensus repeat protein (rmsd Cα usually <0.5 Å between repeats; Figure 4c, e). The identical nature of the repeats allows one to delineate the roles of the highly conserved structural residues (Figures 3 and 4). The designed TPR repeat proteins are dominated by hydrophobic interactions, whereas in the ank structures there are extensive hydrogen-bonding networks, stabilizing the β-hairpin/loop regions, that are combined with a cluster of hydrophobic interactions between the α helices. Interestingly, in the TPR structures, the 'signature' large hydrophobic residues force the α helices apart into their characteristic elongated structure. In contrast, the ank repeat's

Table 1

Summary of the design characteristics of consensus repeats.

Design and no. of repeats	Selection of amino acids in single repeat motif (Figure 3)	Other design features
TPR [27••] 1.5 to 3.5 repeats	*Those with highest global propensity (P_g). P_g is the ratio of the percentage occurrence of an amino acid at a given position to its percentage occurrence in the whole protein database.	A helix nucleating sequence was added to first TPR motif. Solvating helix (the A helix of the TPR motif with four hydrophobic residues mutated to hydrophilic residues) was added to terminal TPR motif to shield hydrophobic core.
Ank1 [28••] 1 to 4 repeats	Positions that had one amino acid present >50% — automatically selected. Positions where 2–4 amino acids of the same type dominate — most frequent chosen. Positions with no preference or where the amino acids were not of the same type — assigned ''in an effort to satisfy the amino acid distribution data and location in the proposed ank repeat secondary structure''. The pairwise co-variation between amino acids was calculated. However, the results did not affect the design as the most highly co-variant pairs were already present.	N-terminal repeat starts at position 1. C-terminal repeat terminated before the β-hairpin/loop region (position 26).
Ank2 [29••] 4 to 6 repeats	†A framework of structurally important residues was defined (by a mixture of consensus and structure alignment) and the rest were randomised in a library format.	†The consensus repeats were inserted between designed N- and C-terminal capping ank repeats. These were used to shield the hydrophobic core from solvent. Six randomly chosen library members were then selected for characterisation.

*The exception being cysteine, which was mutated to alanine. †More information is described in patent application PCT/EP01/10454.

'signature' residues are used to lock together the α helices either side of the β-hairpin/loop region (Figure 4).

These robust designs seem perfect for use as scaffolds on which novel binding specificities can be engineered. In fact, Kohl *et al.* [29••] designed in these features by creating randomised libraries in the loop region of their ank proteins. In this paper, they also report that, by using ribosome display, they have been able to select several library members to bind to a number of globular proteins (specific binding in the low nanomolar range) [29••].

Conclusions

In this review, we have described the recent exciting advances in the design and biophysical understanding of the relatively simple repeat protein folds. It seems that, despite their unique nonglobular structures, they follow many of the same principles that are observed for globular proteins; for example, they cooperatively (un)fold and react in a similar fashion to mutation. However, unlike globular proteins, repeat proteins can be trimmed into smaller fragments that remain folded. The trimming is directly related to the modular nature of the folds and can be roughly gauged by the analysis of residue contact maps for each folded structure. It will be interesting to see if other repeat proteins fold in a similar manner to the sequential unfolding of p16.

The advent of several consensus designs (ank and TPR motifs) has led to promising results, as well as raising some pertinent questions. The designs have shown that hyperstable, monomeric and correctly folded proteins can be constructed from an idealised motif. These proteins increase in stability as more units of the repeat are added. Furthermore, the regularised protein structures produced not only highlight the important interactions made by the highly conserved or 'signature' residues, but also have been shown to be stable interacting scaffolds onto which new specificities can be built. They provide excellent starting points from which to redesign superhelical structure, to study stability changes and to elucidate folding pathways.

As a final note, after such design success, one wonders how difficult it is to produce designed repeat proteins. What are the minimal requirements? For example, would it be possible to reduce the idealised repeat to an alanine framework with only the highly conserved motif-forming residues present? Could other more complex repeat protein topologies, for example, the WD40 repeat, be so easily designed?

Acknowledgements

We thank Andrew Brown and Laura Itzhaki for their insightful comments on the manuscript. We are grateful for the contributions to our understanding of TPR structure and function from our colleagues Luca D'Andrea, Melanie Cocco and Yong Xiong. Ewan Main is funded by the Wellcome Trust and is a Research Fellow at Girton College, Cambridge.

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. Groves MR, Barford D: **Topological characteristics of helical repeat proteins**. *Curr Opin Struct Biol* 1999, **9**:383-389.

2. 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.

3. Classification and properties of solenoid proteins on World Wide Web URL: http://cmm.info.nih.gov/kajava/solenoidtable.html

4. • Kajava AV: **What curves alpha-solenoids? Evidence for an alpha-helical toroid structure of Rpn1 and Rpn2 proteins of the 26 S proteasome**. *J Biol Chem* 2002, **277**:49791-49798.
An interesting study of the factors that contribute to the twist and curvature of repeat proteins.

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

6. Walker RG, Willingham AT, Zuker CS: **A *Drosophila* mechanosensory transduction channel**. *Science* 2000, **287**:2229-2234.

7. PFAM — top twenty families on World Wide Web URL: http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00400

8. Blatch GL, Lassle M: **The tetratricopeptide repeat: a structural motif mediating protein-protein interactions**. *Bioessays* 1999, **21**:932-939.

9. 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.

10. Jacobs MD, Harrison SC: **Structure of an IkappaBalpha/NF-kappaB complex**. *Cell* 1998, **6**:729-731.

11. Gatto GJ Jr, Geisbrecht BV, Gould SJ, Berg JM: **Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5**. *Nat Struct Biol* 2000, **7**:1091-1095.

12. •• Zhang B, Peng Z: **A minimum folding unit in the ankyrin repeat protein p16(INK4)**. *J Mol Biol* 2000, **299**:1121-1132.
The first study to show that it is possible to remove various numbers of repeats from a repeat protein and still produce stable, folded fragments.

13. •• Tang KS, Fersht AR, Itzhaki LS: **Sequential unfolding of ankyrin repeats in tumor suppressor p16**. *Structure* 2003, **11**:67-73.
The first study to characterize the folding transition state of a repeat protein using Φ-value analysis.

14. 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.

15. Cammett TJ, Luo L, Peng Z: **Design and characterization of a hyperstable p16(INK4a) that restores Cdk4 binding activity when combined with oncogenic mutations**. *J Mol Biol* 2003, **327**:285-297.

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

17. 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.

18. • 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.
A study that characterises the cooperative nature of the equilibrium unfolding of a larger ank repeat protein in response to specific point mutations.

19. • 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.
Back-to-back studies [20•] showing the initial biophysical characterisation of a larger ank-containing protein.

20. • 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.
Back-to-back studies [19•] showing the initial biophysical characterisation of a larger ank-containing protein.

21. Kamen DE, Griko Y, Woody RW: **The stability, structural organization, and denaturation of pectate lyase C, a parallel beta-helix protein**. *Biochemistry* 2000, **39**:15932-15943.

22. Kamen DE, Woody RW: **Identification of proline residues responsible for the slow folding kinetics in pectate lyase C by mutagenesis**. *Biochemistry* 2002, **41**:4724-4732.

23. Kamen DE, Woody RW: **Folding kinetics of the protein pectate lyase C reveal fast-forming intermediates and slow proline isomerization**. *Biochemistry* 2002, **41**:4713-4723.

24. Kamen DE, Woody RW: **A partially folded intermediate conformation is induced in pectate lyase C by the addition of 8-anilino-1-naphthalenesulfonate (ANS)**. *Protein Sci* 2001, **10**:2123-2130.

25. Nikolova PV, Wong KB, DeDecker B, Henckel J, Fersht AR: **Mechanism of rescue of common p53 cancer mutations by second-site suppressor mutations**. *EMBO J* 2000, **19**:370-378.

26. Daggett V, Fersht AR: **Is there a unifying mechanism for protein folding?** *Trends Biochem Sci* 2003, **28**:18-25.

27. •• 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**:1-20.
See annotation to [29••].

28. •• 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.
See annotation to [29••].

29. •• 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.
Three studies [27••–29••] that describe the successful design of repeat proteins from a consensus sequence.

30. Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, van Loon AP, Wyss M: **The consensus concept for thermostability engineering of proteins: further proof of concept**. *Protein Eng* 2002, **15**:403-411.

31. Knappik A, Ge L, Honegger A, Pack P, Fischer M, Wellnhofer G, Hoess A, Wolle J, Pluckthun A, Virnekas B: **Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides**. *J Mol Biol* 2000, **296**:57-86.

32. Koradi R, Billeter M, Wuthrich K: **MOLMOL: a program for display and analysis of macromolecular structures**. *J Mol Graph* 1996, **14**:51-55.

33. Kraulis P: **MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures**. *J Appl Crystallogr* 1991, **24**:946-950.

34. Merritt EA, Murphy MEP: **Raster3D version 2.0. A program for photorealistic molecular graphics**. *Acta Crystallogr* 1994, **D50**:869-873.

35. Guex N, Peitsch MC: **SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling**. *Electrophoresis* 1997, **18**:2714-2723.

36. POV-Ray and the Persistence of Vision Raytracer on World Wide Web URL: http://www.povray.org/