Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Ligand binding by repeat proteins: natural and designed

Tijana Z Grove[1], Aitziber L Cortajarena[1] and Lynne Regan[1,2]

Repeat proteins contain tandem arrays of small structural motifs. As a consequence of this architecture, they adopt non-globular, extended structures that present large, highly specific surfaces for ligand binding. Here we discuss recent advances toward understanding the functional role of this unique modular architecture. We showcase specific examples of natural repeat proteins interacting with diverse ligands and also present examples of designed repeat protein–ligand interactions.

Addresses
[1] Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA
[2] Department of Chemistry, Yale University, New Haven, CT 06520, USA

Corresponding author: Grove, Tijana Z (tijana.grove@yale.edu), Cortajarena, Aitziber L (aitziber.lopezcortajarena@yale.edu) and Regan, Lynne (lynne.regan@yale.edu)

Current Opinion in Structural Biology 2008, **18**:507–515

This review comes from a themed issue on
Engineering and design
Edited by Dek Woolfson and Lynne Regan

Available online 23rd July 2008

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2008.05.008

Introduction

Repeat proteins consist of tandem arrays of a small structural motif. Here we focus on six different families (Table 1), highlighting examples that illustrate general themes of repeat protein function. In the past two years there have been significant advances both in our understanding of the functioning of natural repeat proteins and in our ability to engineer novel repeat proteins. Here we discuss features of repeat protein function that are intimately dependent upon their unique structure. Repeat proteins are, by their very nature, extended structures, and thus have a larger surface area to volume ratio than typical globular proteins. They are therefore particularly well suited to mediate protein–protein interactions and to organize multiple proteins into functional complexes. Moreover, the modular structure allows for different sets of repeats to be used to bind to different ligands.

For several classes of repeat proteins, including Ank, TPR, and LRR (Table 1), analyses of the amino acid variability at different positions within a single repeat have revealed an interesting feature—the residues that comprise the ligand binding site are significantly more variable than other positions on the protein surface (Figure 1a) [1–3]. Further analysis of these hypervariable positions, in examples where structures are available, reveals that they are in direct contact with ligand. This finding is consistent with the idea that the repeat protein provides a constant framework that can be 'decorated' with functional residues. In fact, sequence 'hypervariation' alone can be used to predict the ligand binding sites. This observation is exactly analogous to that of Wu and Kabat, who first identified the complementarity determining regions (CDR) of antibodies on the basis of sequence hypervariability [4].

An additional mechanism by which to enhance binding diversity is seen in the variable lymphocyte receptors (VLR) of the adaptive immune system of jawless vertebrates [2••]. VLRs are composed of LRRs, with a constant N-terminal and C-terminal LRRs flanking an array of one to eight variable central LRRs. The central LRRs are responsible for antigen binding. Germ-line rearrangements insert different numbers of central LRRs as a unique means by which to increase the diversity of ligand recognition (Figure 1b).

No conformational changes upon ligand binding

A common theme in repeat protein structure is that both the individual repeats, and their positions relative to each other, are the same regardless of the protein in which they occur. Moreover, upon ligand binding there is typically little, if any, conformational change. The structures of different repeat proteins, with and without ligand bound, have been compared with RMSD between the ligand-bound and ligand-free structures as little as 1.0 Å. An example for TPR proteins is shown in Figure 1c [5]. In other cases, the same protein with and without ligand bound has been compared, with RMSD differences as little as 0.4 Å and 0.43 Å for LRR and WD40 examples, respectively [6,7].

The structure of β-catenin, a protein containing 12 armadillo repeats, in ternary complex with BCL9 and Tcf4 clearly shows that the armadillo domain is virtually identical to the previously reported structure of this domain alone (Figure 4a) [8]. Slit proteins 1–3 each have an N-terminal domain composed of four tandem LRR domains (D 1–4), each of which contains multiple LRR. The second LRR domain of the Slit protein always interacts with the first Ig domain of the Robo protein (Figure 1d). This complex participates in development of bilateral symmetry in both insects and vertebrates. A comparison of the structure of the Slit2 D2-Robo1 Ig1 complex with that of the uncomplexed domains shows that no structural

Table 1

Repeat protein families and their characteristics

Repeat protein family	Number of amino acids in a repeat	Structural motif of a repeat	Range of numbers of repeats for natural proteins* most common number
HEAT	37–47	Two α-helices (A & B)	3–36
TPR	34	Helix-turn-helix (A & B)	3–16, 3*
Armadillo	42	Three α-helices (H1, H2, H3)	6–15, 12*
Ank	30	Helix-helix-loop (or β-hairpin) (H1, H2)	4–24, 6*
LRR	20–29	β-strand-loop-helix	Up to 28
WD40	40–60	Four-stranded (a–d) antiparallel β-sheet	3–16, 7*–8*

Names for protein families originate from: HEAT, Huntingtin; Elongation factor 3; A subunit of PP2A; lipid kinase TOR; TPR, tetratricopeptide repeat; Armadillo, the appearance of embryos that are mutant for the *Drosophila* segment polarity gene *armadillo*; Ank, ankyrin-like repeat; LRR, leucine rich repeat; WD40 (also known as WD or beta-transducin repeats) amino acid motifs, often terminating in a Trp-Asp (WD) dipeptide.

changes occur upon binding. Interestingly, in this example, complex formation involves an interface that consists of two distinct regions—the first region encompassing repeats 1 and 2 is predominantly electrostatic in nature, whereas the second, encompassing repeats 4, 5, and 6, is predominantly hydrophobic [6]. One can speculate about the modularity of recognition being taken even further in this case.

Karyopherinβ2 (Kapβ2), a protein involved in transporting other proteins into the nucleus, contains about 20 tandem HEAT repeats divided into three major segments: HEAT repeats 1–8 form a Ran binding site [9], whereas HEAT repeats 9–13 and 14–18 are substrate binding sites (Figure 1e) [10••]. The structures of the HEAT repeats, in either complex with different NLS (nuclear localization signal) sequences [11,12], or without ligand bound, are essentially identical [10••]. The interesting conformational change that does occur in this protein upon substrate binding is a rotation of the sets of HEAT repeats relative to each other, as rigid elements, about a flexible hinge between HEAT repeats 13 and 14 (Figure 1f) [10••]. Similar behavior is seen in the importin-β family [10••].

Repeat proteins bind extended ligands

Employing multiple repeats to form an extended surface area of interaction with an extended ligand is an efficient route to tight binding. A common theme is that the repeat protein interacts with an extended peptide (Figure 2b) or element of secondary structure in the target protein, allowing maximal surface area of contact per amino acid. For example, the eight LRRs of the protein internalin interact with a long, extended, β-strand of E-cadherin (Figure 2a) [13••]. Similarly, 12 consecutive HEAT repeats of karyopherin β2 make extensive contacts with the extended NLS peptides (Figure 1e) [10••]. A particularly dramatic example is seen in the interaction of human importin α5 with the C-terminal domain of influenza virus polymerase PB2, where the last 20 residues of PB2 actually unfold into an extended conformation for interaction with the 10 armadillo repeats of importin (Figure 2c) [14••].

Repeat proteins as multi-protein complex organizers

The extended, modular nature of repeat proteins also allows for different regions of the repeat protein to be used to interact with different ligands, thus bringing them together into a functional complex. Such repeat protein mediated multi-protein organization can occur in different ways.

In Hsp organizing protein (HOP), for example, there are two discrete sets of TPR modules, one of which binds to Hsp70 and the other to Hsp90, bringing the chaperones together into a functional complex (Figure 3a) [15].

HEAT repeats are used to assemble multi-protein complexes in proteins that function in nucleocytoplasmatic transport in a very different fashion. The 20–40 HEAT repeats in karyopherin form a superhelix, the external convex surface is involved in nucleoporin binding whereas the inner, concave face interacts with the NLS of cargo proteins (Figure 3b, left) [10••,11,12]. The concave face also presents a binding site for the regulatory protein Ran-GTP (repeats 1–8) (Figure 3b, right) [9].

Another example of using different sets of HEAT repeats for binding different proteins is seen in protein phosphatase 2A (PP2A). PP2As are heterotrimeric proteins in which a 'scaffolding' A subunit (with 15 HEAT repeats) binds to both the regulatory 'B' subunit and the catalytic 'C' subunit with different sets of HEAT repeats (Figure 3c, left) [16••]. Both AC and ABC versions of the complex exist, demonstrating the independence of binding to different sets of repeats within the HEAT domain. Interestingly, it has recently been shown that SV40 small T antigen perturbs the functioning of PP2A