The plastic landscape of repeat proteins

Diego U. Ferreiro and Elizabeth A. Komives*

Department of Chemistry and Biochemistry and the Center for Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0378

Nearly 20% of the proteins encoded by the human genome contain multiple repeated units of 30–40 amino acids often occurring in tandem arrays referred to as repeat domains (1). In this issue of PNAS, Werbeck and Itzhaki (2) describe the equilibrium folding mechanism of the largest repeat domain ever studied, D34, a 426-residue fragment encompassing the last 12 ankyrin repeats of Ankyrin R. Although some repeats behave as "beads-on-a-string" and fold independently, it is now becoming clear that many do not (3). Often, repeating elements only fold in the context of similar repeats and form extended superhelical structures, stabilized only by interactions within repeats and adjacent neighbors. In contrast to three-dimensional globular proteins, there are no direct interactions between residues distant in sequence space (4). This near one-dimensionality has enormous consequences for the description of the energy landscapes, thermodynamics, and kinetics of repeat-containing proteins.

Ankyrin-repeat proteins are found in all three phyla and are present in some 6% of eukaryotic protein sequences (1, 5). Sequence and structural alignments reveal that conserved amino acids form a folded scaffold with nonconserved amino acids located preferentially on the surface (6, 7). Several groups have successfully designed stable ankyrin repeats and thereby have established that the conserved residues play the primary role in folding stability (6–8). Mutation of conserved residues nearly always results in reduction of stability, whereas mutation of nonconserved surface residues usually has little effect (9–11). In line with this manifestation of "minimal frustration," it has been shown that the marginally stable ankyrin-repeat domain of IκBα is stabilized by mutating residues to conform to the consensus (12).

The high cooperativity of the folding of globular proteins resembles a phase transition and arises in part from interactions between residues distant in sequence space. Because ankyrin-repeat domains lack such distant interactions, why do they sometimes fold in a highly cooperative manner? Many experiments can be interpreted to imply that ankyrin-repeat proteins have only two equilibrium states: completely folded or unfolded (10, 11, 13, 14). This cooperativity strongly depends on interactions of the repeats with their nearest neighbors (15, 16). Both experiments and simulations converge in showing that this observed cooperativity can be understood if the domains fold up by a mechanism in which the formation of the interface between elements is highly favorable. A funneled energy landscape, like that for globular domains, is consistent with this observation (15, 17, 18). Indeed, simulations on natural ankyrin-repeat domains of different lengths reveal that as the number of repeats increases, the cooperativity tends to break down, presumably because the increasing entropy advantage of introducing a broken interface anywhere between repeats is weighted against a fixed energy cost (18).

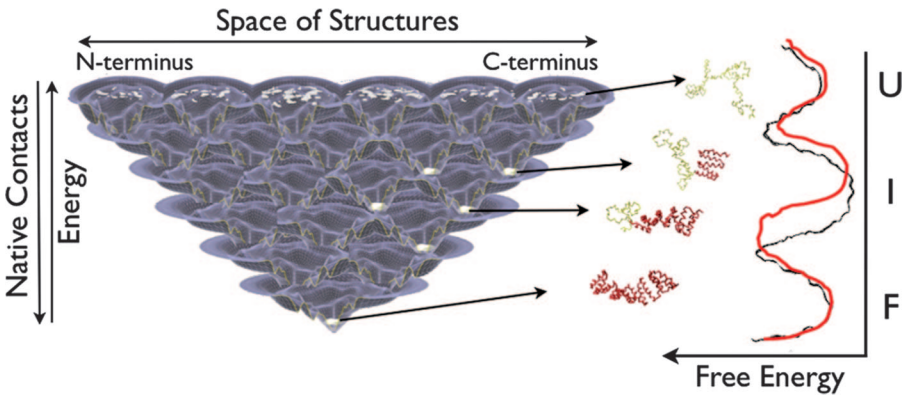

Fig. 1. A plastic landscape for repeat-protein domains. Each repeat has similar structural elements that interact with nearest neighbors, depicted by many folding funnels that merge upon interaction and comprise one overall funneled landscape. High-energy unfolded configurations are near the top, and the native state is at the bottom. In between, an intermediate ensemble of partially folded states can become populated at equilibrium (represented as clouds in the energy landscape). This landscape attempts to portray the relative populations of D34 under mild denaturing conditions; for simplicity, only six funnels are shown, each representing 2 of the 12 repeats. Because of landscape asymmetries, configurations in which at least some C-terminal repeats are folded are more populated. Mutations in the C-terminal part change the interactions in or between the modules, causing the populations to shift toward configurations with fewer of the C-terminal repeats folded (red free-energy profile on the right; U, unfolded; I, intermediate; F, folded). The intermediate ensemble is plastic because it can be molded by modest sequence modifications.

The expected breakdown of strict cooperativity is beautifully confirmed in the daring set of experiments Werbeck and Itzhaki (2) have carried out on the 12-ankyrin-repeat domain, D34. Most fortunately, a single tryptophan residue in between the sixth and seventh repeats provides a well located probe that reveals the presence of a hyperfluorescent intermediate state at equilibrium under mild denaturant concentrations. The intermediates' signal probably arises from the placement of a broken interface near the probe and was not detected by far UV-CD, a global probe of α-helical structure often used to follow the folding of ankyrin repeats.

Werbeck and Itzhaki (2) elegantly generated a set of mutants all along the repeating array. In their constructs, a conserved valine is substituted by an alanine perturbing both interrepeat and intrarepeat interactions (8). Each of these mutations destabilizes locally, allowing site-specific probing of the folding intermediates. Mutation of the valines in the N-terminal repeats reduces the stability, but does not greatly affect the cooperativity of the folded to intermediate (F $\rightarrow$ I) transition so that the F $\rightarrow$ I transition can now be distinguished from the intermediate to unfolded (I $\rightarrow$ U) transition by CD. On the other hand, mutations in the C-terminal repeats dramatically increase the cooperativity of the F $\rightarrow$ I transition and correspondingly decrease the cooperativity of the I $\rightarrow$ U transition. Each

Author contributions: D.U.F. and E.A.K. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 7863.

*To whom correspondence should be addressed. E-mail: ekomives@ucsd.edu.

of these mutations results in population of a different ensemble of intermediate species. The closer the site of destabilization is to the C terminal, the greater the number of repeats that unfold in the F → I transition. The observed "plasticity" by which single mutations can "mold" different intermediate species is highly unusual and has not been seen before.

The energy landscape theory of protein folding argues that the three-dimensionally connected globular proteins must fold along a funneled landscape (19, 20). One-dimensionality would weaken this necessity (21). Nevertheless, folding simulations based on perfectly funneled model landscapes revealed how finer details of the energetic contributions contribute to folding and recapitulate the experimental results (18, 22). This conclusion can also be directly reached from experiments. Mello and Barrick (15) have derived a funneled energy landscape from equilibrium unfolding data on the Notch seven-ankyrin-repeat domain and a simple energy function based on the folding of each repeat and the interaction with its neighbors. The results of Werbeck and Itzhaki (2) also suggest that the folding energy landscape of D34 is funneled but composed of approximately two six-repeat subdomains. The different stabilities of these subdomains hint at unevenness in the energy landscape (Fig. 1). Mutation of the valines in the C-terminal six repeats destabilizes both the interrepeat and intrarepeat interactions and results in a broad and plastic ensemble of species of similar energies that cannot be distinguished from one another. Thus, the mutations cause further landscape unevenness similar to that which was shown to give rise to multistate folding in the Notch ankyrin-repeat domain (16). It is important to note that this unevenness is of a different sort than the "roughness" discussed in landscape theory, which is typically understood as frustration caused by non-native interactions (19, 20). Mutation of the D34 valines most probably does not cause the formation of non-native contacts but destabilizes key interactions in or between elements, thus populating different configurations of the ensemble that have only a subset of native interactions. This is a hallmark of a funnel. It will be extremely interesting to simulate the folding of this domain to understand the microscopic nature of the intermediate ensembles in both the wild-type and mutant forms of D34.

The folding energy landscape of D34 is funneled.

Energy landscape theory also makes predictions about the kinetics of protein folding. Experimental data as well as theory converge and indicate that, owing to their near one-dimensionality, small energetic perturbations can strongly affect the folding kinetics of repeat proteins. Lowe and Itzhaki (10) have also recently tackled the folding kinetics of the four-ankyrin-repeat domain of Myotrophin and have shown that it folds through two parallel routes nucleating at either terminal repeat pair. They further designed mutations that change the relative population of these routes by changing the relative free energy of the transition-state ensembles, the kinetic bottlenecks separating the folded and unfolded basins. In contrast, Notch appears to fold preferentially via a single route, as does the four-ankyrin-repeat domain of p16-INK4 (23, 24). Also, single-molecule mechanical unfolding of long ankyrin-repeat proteins shows that they can unfold in a stepwise manner (25). These observations illustrate the underlying richness of the structural ensembles populated on repeat protein folding landscapes.

It is interesting to note that the intermediates' plasticity in the D34 domain of Ankyrin R occurs in the C-terminal subdomain, the region of the protein that binds the "C-terminal extension loop." It is possible that binding and folding events are somehow coupled, as has been shown for IκBα ankyrin-repeat domain, in which the least stable subdomain folding is coupled to NF-κB binding (26). Ankyrin R is thought to bind multiple targets, not all of which have yet been identified. Binding at one subdomain provides a mechanism to modulate the folding landscape of the other, which provides the means to differentially transmit allosteric effects between distant sites along the ankyrin-repeat array (27). In this way, the binding of multiple partners to Ankyrin R may influence the binding properties of other partners by modifying the plastic landscape. Such a cooperativity may be much more common in molecular biology than the usual independent domains diagrams suggest.

D.U.F. is a Jane Coffin Childs Postdoctoral Fellow. E.A.K. was supported by National Institutes of Health Grant P01 GM071862.

1. Bjorklund AK, Ekman D, Elofsson A (2006) *PLoS Comput Biol* 2:0959–0970.
2. Werbeck ND, Itzhaki LS (2007) *Proc Natl Acad Sci USA* 104:7863–7868.
3. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1999) *J Mol Biol* 293:151–160.
4. Main ER, Jackson SE, Regan L (2003) *Curr Opin Struct Biol* 13:482–489.
5. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) *Protein Sci* 13:1435–1448.
6. Mosavi LK, Minor DL, Jr, Peng ZY (2002) *Proc Natl Acad Sci USA* 99:16029–16034.
7. Forrer P, Stumpp MT, Binz HK, Pluckthun A (2003) *FEBS Lett* 539:2–6.
8. Forrer P, Binz HK, Stumpp MT, Pluckthun A (2004) *ChemBioChem* 5:183–189.
9. Zweifel ME, Leahy DJ, Hughson FM, Barrick D (2003) *Protein Sci* 12:2622–2632.
10. Lowe AR, Itzhaki LS (2007) *J Mol Biol* 365:1245–1255.
11. Tang KS, Guralnick BJ, Wang WK, Fersht AR, Itzhaki LS (1999) *J Mol Biol* 285:1869–1886.
12. Ferreiro DU, Cervantes CF, Truhlar SME, Cho SS, Wolynes PG, Komives EA (2006) *J Mol Biol* 365:1201–1216.
13. Zweifel ME, Barrick D (2001) *Biochemistry* 40:14357–14367.
14. Mosavi LK, Williams S, Peng Z-y (2002) *J Mol Biol* 320:165–170.
15. Mello CC, Barrick D (2004) *Proc Natl Acad Sci USA* 101:14102–14107.
16. Street TO, Bradley CM, Barrick D (2007) *Proc Natl Acad Sci USA* 104:4907–4912.
17. Wolynes PG (1997) *Proc Natl Acad Sci USA* 94:6170–6174.
18. Ferreiro DU, Cho SS, Komives EA, Wolynes PG (2005) *J Mol Biol* 354:679–692.
19. Bryngelson JD, Wolynes PG (1987) *Proc Natl Acad Sci USA* 84:7524–7528.
20. Onuchic JN, Wolynes PG (2004) *Curr Opin Struct Biol* 14:70–75.
21. Luthey-Schulten Z, Ramirez BE, Wolynes PG (1995) *J Phys Chem* 99:2177–2185.
22. Wolynes PG (2005) *Q Rev Biophys* 38:405–410.
23. Bradley CM, Barrick D (2006) *Structure (London)* 14:1303–1312.
24. Tang KS, Fersht AR, Itzhaki LS (2003) *Structure (London)* 11:67–73.
25. Li L, Wetzel S, Pluckthun A, Fernandez JM (2006) *Biophys J* 90:L30–32.
26. Truhlar SME, Torpey JW, Komives EA (2006) *Proc Natl Acad Sci USA* 103:18951–18956.
27. Tripp KW, Barrick D (2007) *J Mol Biol* 1187–1200.