Available online at www.sciencedirect.com
ScienceDirect
Archives of Biochemistry and Biophysics 469 (2008) 83–99
ABB
www.elsevier.com/locate/yabbi

Review

Repeat-protein folding: New insights into origins of cooperativity, stability, and topology

Ellen Kloss, Naomi Courtemanche, Doug Barrick *

T.C. Jenkins Department of Biophysics, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA

Received 25 July 2007
Available online 15 September 2007

Abstract

Although our understanding of globular protein folding continues to advance, the irregular tertiary structures and high cooperativity of globular proteins complicates energetic dissection. Recently, proteins with regular, repetitive tertiary structures have been identified that sidestep limitations imposed by globular protein architecture. Here we review recent studies of repeat-protein folding. These studies uniquely advance our understanding of both the energetics and kinetics of protein folding. Equilibrium studies provide detailed maps of local stabilities, access to energy landscapes, insights into cooperativity, determination of nearest-neighbor interaction parameters using statistical thermodynamics, relationships between consensus sequences and repeat-protein stability. Kinetic studies provide insight into the influence of short-range topology on folding rates, the degree to which folding proceeds by parallel (versus localized) pathways, and the factors that select among multiple potential pathways. The recent application of force spectroscopy to repeat-protein unfolding is providing a unique route to test and extend many of these findings.

Keywords: Repeat protein; Ankyrin repeat; Protein folding; Energy landscape; Atomic force microscopy

In the last 25 years, advances in experimental studies and in computation and theory have greatly improved our understanding of protein folding. On the experimental side, advances have come from new and improved techniques, including site-directed mutagenesis, hydrogen exchange (HX)[1] methods, improvements to rapid mixing devices, and development of single-molecule fluorescence and force spectroscopy. Experimental advances have also come in the form of generalizations and insights from expanding databases of thermodynamic and kinetic constants for protein folding [1–5].

On the computational side, advances in our understanding of protein folding have come from huge increases in computer speed and storage capacity, in innovative methods to study energetics of folding such as ensemble-based approaches and replica exchange methods [6,7], and distributed computing methods [8]. As with experiment, computational studies of folding, and in particular, fold prediction, have greatly benefited from expanding databases of protein structure [9,10]. On the theoretical side, the application of ideas from condensed-matter physics and the resulting "landscape" picture of protein folding have provided a new perspective for describing folding [11–14].

Recent observations and related questions in folding of globular proteins

The advances highlighted above have provided some important insights into how proteins fold. One important insight from experimental studies is that for many proteins the rate of folding is inversely correlated to the density of sequence-distant contacts (referred to as "contact order") in the native state [3,15]. This correlation suggests that

* Corresponding author. Fax: +1 4105164118.
E-mail address: barrick@jhu.edu (D. Barrick).

[1] *Abbreviations used:* HX, hydrogen exchange; CD, circular dichroism; N, native; D, denatured; AFM, atomic force microscopy; HMMs, hidden markov models.

0003-9861/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.abb.2007.08.034

the overall fold or "topology" of the native state is established in the rate-limiting step in folding [16]. At a more detailed resolution, rate-equilibrium relationships of single-residue protein variants (and so-called "Φ-values") indicate that detailed side-chain packing interactions vary in their extent of formation in the rate-limiting, or transition state ensemble, often remaining unformed until the native state is formed (for a review of Φ-values in globular proteins, see the article by Royer in this issue).

It is interesting to compare these two measures of the rate-limiting steps in folding with a long-standing and general observation regarding equilibrium protein folding, namely that of cooperativity. For many (but not all) globular proteins, folding appears to proceed in a thermodynamically "two-state" reaction in which all structural features of the native state are formed in a single, concerted fashion. This type of equilibrium cooperativity seems in accordance with the suggestion from the contact order correlation that the limiting step in folding is the *concerted* organization of the bulk of the native backbone fold. However, these two observations are less in accord with the variation of the extent of packing in the transition state ensemble suggested from Φ-value analysis. Does this packing variation indicate a specific pathway in which some tertiary interactions form late, despite their residing within a native backbone topology? Or is the backbone less native-like than the topology correlation suggests? In this regard, what are the limits of the topology-rate correlation?

The above questions raise the issue of pathways and the interactions responsible for specifying them. If there are specific pathways in which some stable structural elements form late, are these folding routes controlled kinetically, rather than reflecting the overall stability of all parts of the protein? Or does the dominant kinetic pathway genuinely reflect a low energy pathway through conformational space in a global sense? And if there is large variation in local stability, what is the glue that holds different regions of the protein together to give rise to equilibrium cooperativity?

One of the contributions of the energy landscape view of folding is a recognition that "pathways" may be much more complex than a single sequentially ordered series of well-defined intermediates. In one limiting case, that of a "folding funnel", there is no structurally defined pathway. Rather, each structural degree of freedom (torsion angles) might "sense" the direction to the native state locally through a thermally meaningful monotonic decrease in energy, regardless of the configuration of the bulk of the chain, towards a minimum centered on the native state. This type of folding, which is referred to as "downhill" to reflect the uniform decrease in energy toward the native state, would be expected to result in very fast (potentially non-exponential) folding for even the largest conformational searches.

In a less severe barrier-limited landscape, following a partial decrease in the available conformations (perhaps by local structural propensity) folding is limited by an unbiased search (through an "entropy bottleneck") to the native state. Although the observed topology correlation would argue against a purely native-centric downhill mechanism, the often diffuse nature of the transition states of many proteins determined by Φ-value analysis [5,17,18], along with a recent interpretation of the folding of a small polypeptide [19,20],[2] are consistent with a downhill mechanism. Clearly, what is needed is a direct and detailed experimental determination of the features of the energy landscapes of protein folding.

Challenges in understanding folding using globular proteins

In motivating the study of folding using repeat proteins, it is worth identifying some of the structural features of globular proteins that limit our understanding of folding. One confounding feature of globular protein structures is that distant segments of the polypeptide chain are often in close contact (the structural basis for high contact order in globular proteins). As described below, the interconnected, "long-range" topologies of globular proteins lead to a "dissection" problem. Another confounding structural feature of globular proteins is their complex, irregular, heterogeneous architectures. Although globular proteins are constructed of a limited and fairly regular set of hydrogen bonded secondary structure elements (α-helices, β-strands, and turns), the arrangements of adjacent secondary structure elements (supersecondary structure) and the higher order tertiary structures of globular proteins are very irregular and highly variable from one protein to the next (Fig. 1a). As described below, this heterogeneity leads to a "comparison" problem.

The dissection problem, which results from the long-range topologies of globular proteins, makes folding energy difficult to dissect because close contacts among distant chain segments are likely to promote cooperativity in folding (local loss of structure will destabilize regions that remain folded). Although concerted folding of globular proteins simplifies analysis of both equilibrium and kinetic data, it prevents dissection of energetics (and also rates of conversion) to different structural elements, which is required to construct energy landscapes for real proteins. Long-range topology is also likely to confound experimental dissection of globular proteins into fragments: in the absence of sequence-distant contacts, protein fragments typically unfold and often aggregate, again preventing determination of local energetics [23].

The comparison problem, which results from the heterogeneous supersecondary and tertiary structures of globular proteins, presents a second challenge in extracting general principles in protein folding: even if local energy distribution within a globular protein can be determined, the underlying origins local energy differences is unclear. For example, in a protein with a pair of helices and a small

[2] Although see [21,22] for an alternative view.