

An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms

Alice I Bartlett & Sheena E Radford

In recent years, improvements in experimental techniques and enhancements in computing power have revolutionized our understanding of the mechanisms of protein folding. By combining insights gained from theory, experiment and simulation we are moving toward an atomistic view of folding landscapes. Future challenges involve exploiting the knowledge gained and methods developed to enable us to elucidate a molecular description of folding dynamics in the complex environment of the cell.

Most proteins are required to adopt a specific three-dimensional structure to be biologically active. How they achieve this has been the subject of immense scientific interest spanning the decades since the first structures of proteins were elucidated¹. The conformational space accessible to the polypeptide chain is astronomically large, yet proteins fold on a biologically relevant timescale, with some obtaining their native structure *in vitro* in just microseconds². How rapid folding is achieved has been rationalized by a number of concepts: the presence of nonrandom interactions in the initial denatured state that limit the conformational space available at the start of the folding reaction³; folding via intermediates that mark the way to the native structure⁴; and the realization that proteins fold on funnelled energy landscapes⁵ that describe folding as the inevitable consequence of the requirement to lower the free energy (increase stability) as more native contacts form⁶ (Fig. 1). In this landscape view of folding, the denatured state of the protein populates a large ensemble of structures. The polypeptide chain may then fold by numerous pathways, potentially adopting multiple partially folded ensembles *en route* to the native state⁶.

For a protein that folds via a two-state transition (with a mechanism in which only the denatured and native states are populated) the energy landscape is relatively smooth. Such a landscape lacks deep valleys and high barriers and effectively funnels the polypeptide chain to its native state (Fig. 1a). Such an ideal folding scenario is rare⁴, and many proteins fold on rough, rugged landscapes. Using new methods that can detect sparsely populated and/or transient non-native species (Table 1), even small, simple proteins have now been shown to fold through one or more partially folded states^{4,5}. In general, folding energy landscapes are rugged entities that are suboptimal for folding (Fig. 1b) through which the polypeptide chain has to navigate to the native state⁵. Landscape ruggedness arises as the consequence of the simple fact that native protein structures are stabilized by thousands of

mutually supportive, weak interactions that cannot all be satisfied simultaneously during folding. As a result, energetic minimization of individual interactions can be conflicting, leading to 'frustration' in the energy landscape^{5,7}. This ruggedness may be attributed to opposing evolutionary pressures on protein sequences to enable them to fold reliably but also to avoid aggregation and to carry out specific biological functions^{5,7}.

Landscape theory predicts an additional folding scenario in which the native structure is attained without encountering any substantial energy barriers, so-called 'downhill folding'⁶. In principle, at least, proteins that fold in a downhill manner open the door to characterization of the folding landscape in immense detail via the myriad of non-native conformations that are accessible experimentally for such a folding mechanism. Barrierless folding is difficult to demonstrate unequivocally by experiment, as proteins that fold in this manner are expected to obtain their native structure with rates close to the folding 'speed limit'². In addition, the experimental hallmarks of this type of folding are difficult to define⁸⁻¹⁰. Nonetheless, downhill folding has been suggested (with much debate^{11,12}) for a number of model proteins¹³⁻¹⁵. Single-molecule experiments have been proposed as a means to differentiate downhill and two-state folding^{10,16}, although this is not straightforward even with such a powerful approach¹⁷. A more detailed discussion of fast protein folding and downhill folding scenarios can be found in ref. 18.

Characterization of all the non-native species (unfolded states, transition states and partially folded intermediates) encountered by proteins that fold in a barrier-limited manner is essential if we are to realize our quest to understand how proteins fold in all-atom detail. Substantial advances toward this goal have been realized for a handful of small proteins¹⁹⁻²⁵. This has been enabled by the development of experimental approaches with faster timescales of measurement²⁶ and enhanced sensitivity (Table 1 and references therein), together with improvements in computing power and new theoretical tools^{6,27}. Today, the arsenal of biophysical methods available to the experimentalist allows transitions from picosecond to second (or longer) timescales to be monitored and species populated to as little as

Astbury Centre for Structural Molecular Biology and Institute of Molecular and Cellular Biology, University of Leeds, Leeds, UK. Correspondence should be addressed to S.E.R. (s.e.radford@leeds.ac.uk).

Published online 3 June 2009; doi:10.1038/nsmb.1592

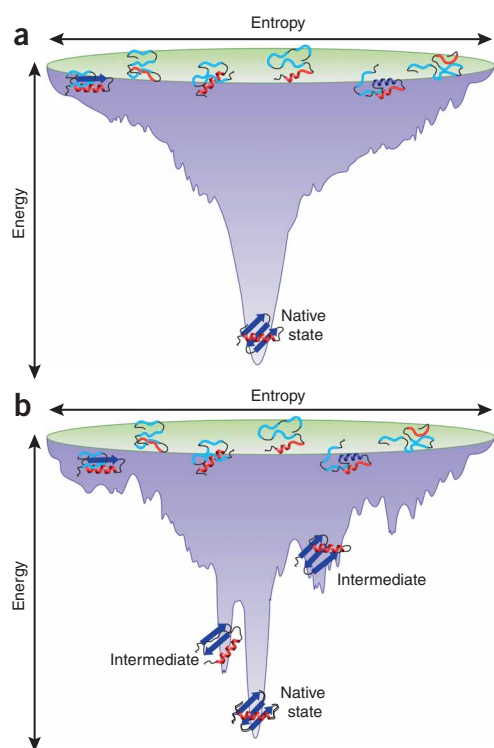


Figure 1 Schematic representation of folding funnels. Example of a smooth energy landscape, through which the polypeptide chain is effectively funneled to the native structure (a), and a more rugged landscape, through which the polypeptide chain has to navigate, possibly via one or more populated intermediates, to the native state (b). In both examples, the denatured state occupies a broad ensemble of structures containing elements of both native and non-native interactions.

0.5% to be identified and structurally assessed²⁸. In principle, single-molecule techniques offer the potential to map folding events one molecule at a time. Using this approach, rare species can be detected and characterized that may be hidden by the averaging inherent within ensemble experiments¹⁷. This approach also enables the measurement of intramolecular diffusion coefficients in denatured and partially folded states, providing detailed insights into the nature of the polypeptide chain at different stages of folding^{29,30}. Perhaps most importantly, these experiments can link models based on chemical kinetics commonly used in protein folding with the more physical description of folding in terms of quantitative free-energy surfaces¹⁷.

A detailed review of the insights that have revolutionized our understanding of protein-folding mechanisms and their impact on biology is beyond the scope of this short article. Here we focus on three areas that have seen major advances in recent years: (i) the structural diversity and properties of non-native states, (ii) current knowledge about folding pathways and the extent to which protein sequences are optimized for folding efficiency, and (iii) new approaches that are beginning to allow us to take the knowledge gained from *in vitro* studies toward a molecular description of folding in the cell. In each area we highlight a selection of recent studies showing how different experimental approaches have been used to elucidate new details of protein-folding mechanisms.

The structural properties and diversity of non-native species

A major challenge in the structural, kinetic and thermodynamic characterization of folding landscapes is the transient and heterogeneous

nature of non-native species. Defining the structural properties of non-native states, and determining how they interconvert so as to align them in the context of a folding pathway, has remained a central issue since this field began⁴. Today, the use of experimental methods with enhanced time resolution and sensitivity, in combination with molecular dynamics simulations, are beginning to reveal all-atom models of non-native ensembles^{20–24,31}. Although there remains further room for optimization of this approach^{32,33}, it has been particularly useful in allowing visualization and interrogation of ensembles of structures that represent the experimental data (rather than a unique solution to the experimental observables). These models can then inspire new experiments to test and refine the structural ensembles produced^{22,34}.

The denatured protein ensemble is of particular interest in the context of the landscape view of folding, because this is the state from which folding initiates. As the denatured protein ensemble is rarely populated at equilibrium, obtaining structural information about this species is a challenging task. This has been achieved using naturally unstable proteins in which the native and denatured states are in equilibrium under ambient conditions³⁵ or by creating proteins by mutation¹⁹ or chemical modification³⁶ that are denatured under conditions that typically favor folding in their wild-type counterparts. Alternative strategies involve denaturing the protein under acidic conditions or adding chaotropes, with the caveat that the structural properties of these ensembles may differ from those under less harsh conditions^{37,38}.

Studies of the denatured ensemble of the helical protein Im7, formed in 6 M urea, using chemical shift analysis and NOE measurements, revealed that this species lacks regular elements of secondary structure. Despite this, the polypeptide chain contains clusters of interacting hydrophobic side chains in those regions that ultimately form helices in the native state, potentially priming the protein for subsequent folding events³⁷. The existence of structure in the denatured state of Im7 in the presence of chaotrope builds on an increasing body of data that indicates conformational restriction in denatured chains^{35,36,38}. Folding of the helical λ repressor protein also commences from a highly nonrandom state³⁶. When denatured under ambient conditions by oxidation of methionine residues, this protein has been shown to possess nascent α -helical structure in the N-terminal region, whereas the C-terminal region remains nonhelical yet conformationally restrained³⁶. All-atom images of the denatured ensembles of the all-helical acyl coenzyme A binding protein (ACBP) and the all β -sheet drkN SH3 domain have been obtained using NMR paramagnetic relaxation enhancement to provide restraints for simulations^{35,38}. These experiments revealed denatured ensembles containing species, ranging from expanded to highly compact, that are stabilized by both native and non-native interactions. Together, these studies indicate that the denatured states of proteins are highly heterogeneous, containing polypeptide chains that vary widely in their individual conformational properties.

By placing donor and acceptor chromophores at different positions, Schuler and co-workers have exploited the power of single-molecule Förster resonance energy transfer (FRET) and fluorescence correlation spectroscopy (FCS) to determine distance distributions in the unfolded ensemble of the cold-shock protein CspTm and the rate of intramolecular diffusion of the polypeptide chain in the denatured state at different denaturant concentrations²⁹. These results suggest that the polypeptide behaves as a Gaussian chain even at low denaturant concentrations where the protein is collapsed and contains $\sim 20\%$ of its native β -sheet structure³⁰. This study reconciles the seemingly contradictory results from small-angle X-ray scattering

Table 1 Experimental techniques that have been applied to the study of protein folding

Technique	Timescale	Information content	Comments	Refs.
Intrinsic tryptophan fluorescence	\geq ns ^a	Environment of tryptophan (through measurement of intensity and λ_{max})	Tryptophan can be introduced (or removed to create a single-tryptophan protein) by protein engineering	97
Far UV CD	\geq μ s ^a	Secondary-structure content	Synchrotron CD may allow a more accurate interpretation of structure content. Can be complicated by aromatic contributions to the spectrum	98
Near UV CD	\geq μ s ^a	Packing of aromatic residues	Only fixed interactions give a near UV CD signal	98
Raman spectroscopy	\geq μ s ^a	Solvent accessibility, conformation of aromatic residues	Information content depends on the frequency used. Not widely applied to folding studies	99
Infrared spectroscopy	\geq ns ^a	Secondary-structure content	Combined with solvent-exchange, information about hydrogen-exchange protection can be obtained	100
ANS (1-anilino-8-naphthalene sulfonic acid) binding	\geq μ s ^a	Exposure of aromatic surface area	Care needs to be taken to ensure that ANS itself does not perturb folding	97
FRET	\geq ps ^a	Molecular ruler, dependent on the distance between two fluorophores (r^{-6} dependence assuming free rotation of the dyes)	Information about rapid fluctuations is possible. Careful design needed to incorporate dyes without perturbing folding	17
FCS	\geq ps	Diffusion time (and hence size and shape)	A powerful method capable of resolving co-populated conformers and their rates of interconversion over ps–ms timescales	101
Anisotropy	\geq μ s ^a	Correlation time measurements provide information about shape and size of molecule	Can provide useful complementary information to FRET distance distributions	97
Small-angle X-ray scattering	\geq μ s ^a	Radius of gyration	With modeling, information about three-dimensional structure can be obtained	102
Absorbance	\geq ns ^a	Environment of chromophore	Peptide bond, aromatic residue or extrinsic moiety may be used	26
Real-time NMR	> min	Structural information via chemical shifts and measurement of NOEs	Powerful method for analysis of denatured states and intermediates in slowly folding proteins	103
Native-state hydrogen exchange	h	Global stability, detection of metastable states	Rare species in equilibrium with the native state that are difficult to detect using other methods can be revealed	104
Pulsed H/D exchange by NMR	\geq ms	Hydrogen exchange protection of folding intermediates on a per-residue basis	Multiple exponential hydrogen-exchange behaviour indicates parallel pathways	104
Pulsed H/D exchange by ESI-MS	\geq ms	Hydrogen exchange protection of folding populations	Quantification of the population of species within heterogeneous ensembles with different hydrogen-exchange properties	105
NMR relaxation methods	\sim ms	Nonrandom structure in denatured states and conformational exchange between different species	If exchange between species occurs at a suitable rate (ms), structural, kinetic and thermodynamic information about rare species can be obtained	28,103
Protein engineering	Depends on probe used	Role of an individual residue in determining the rate of folding and stability of a species of interest	Double-mutant cycles provide pairwise information. Φ -values provide indirect structural information via free-energy changes. Ψ -values use metal chelation to bihistidine motifs to identify specific side chain contacts	42

For a review of theories and simulation methods of folding, see refs. 6,27.

^aThe timescale depends on the method used to initiate folding: temperature jump (ns), pressure jump (μ s), ultra-rapid mixing (μ s), stopped flow (ms) or manual mixing (s).

experiments that demonstrate that the overall dimensions of unfolded proteins are consistent with random coil models³⁹ and increasing evidence of residual structure in unfolded ensembles from NMR studies^{35–38}. Even within such a structured denatured state, the global reconfiguration time is rapid (\sim 50 ns)²⁹. A study of loop formation in unfolded polypeptides using triplet-triplet energy transfer also observed large-scale motions involving chain diffusion occurring on a timescale of 10–100 ns, whereas faster kinetics were observed on the 50–500 ps timescale corresponding to local fluctuations⁴⁰.

For the characterization of more highly structured non-native species, such as partially folded intermediates and transition states, the use of protein engineering (Φ -value analysis) is well established^{41,42}. In recent years Φ -values have been used as restraints for molecular dynamics simulations to generate atomic-level structural models of these ensembles^{20–23,31}. Recent analysis of the early and late

transition state ensembles of two homologous PDZ domains using this approach demonstrated that the early transition states of the two domains are less similar in structure than the subsequent rate-limiting transition state ensembles. This is consistent with the landscape view that conformational space is less restricted earlier in folding²³. The late transition state of both proteins adopts a narrow ensemble of structures with native-like topology. This demonstrates that conformational sampling is highly restricted by this stage of folding, as has been found previously in several other proteins^{21,31}.

Although the interpretation of Φ -values (energetic parameters) in structural terms requires caution^{43–45}, for populated intermediates independent analysis of Φ -values, chemical shifts and hydrogen exchange protection factors allows assessment of the quality of the ensembles that result from molecular dynamics simulations using different observable parameters as restraints²⁰. The characterization

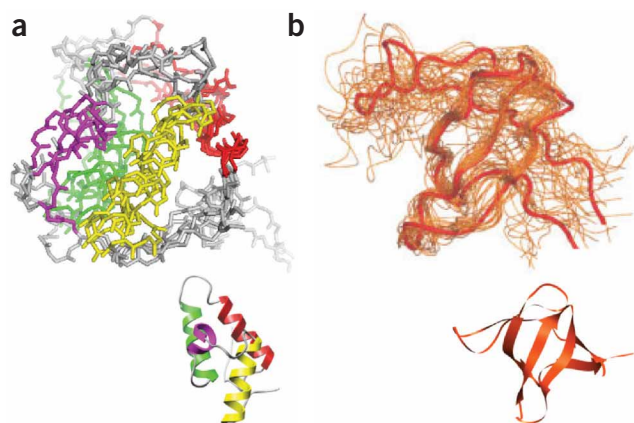


Figure 2 Models of the conformational properties of the ensembles representing the folding intermediates of the bacterial immunity protein Im7 (a; ref. 21) and the rare folding intermediate of the G48V variant of the Fyn SH3 domain²⁴ (b). The native structure of each protein is shown below (PDB 1AYI for Im7 (ref. 94); PDB 1SHF for Fyn SH3 domain⁹⁵). Comparison with the native structure demonstrates that the native topology is well defined in the folding intermediates of both these small proteins. Images of the intermediate ensembles reproduced from *Nat. Struct. Mol. Biol.* (ref. 21) and *Nature* (ref. 24).

of folding intermediates that are stably populated is particularly important because, when on-pathway, they represent stepping stones *en route* to the native state⁴. Such species have also been implicated in misfolding diseases⁴⁶, and their structural characterization offers prospects for therapeutic intervention. By the careful manipulation of experimental conditions to modulate the population of intermediate species and their rates of interconversion, it is possible to characterize these species using the range of approaches listed in **Table 1**. Even 'hidden' intermediates that are kinetically invisible (because they form after the rate-limiting transition state) can be detected and structurally characterized using native-state hydrogen exchange⁴⁷. Rare intermediates can also be detected and structurally analyzed using relaxation dispersion NMR²⁸ (**Table 1**). Structural ensembles representing the folding intermediate of the Fyn SH3 domain have been calculated using chemical shifts determined from relaxation dispersion NMR experiments as restraints²⁴ (**Fig. 2**). A similar strategy was used to determine an ensemble of intermediate structures for Im7 using a combination of Φ -values, hydrogen exchange protection factors and chemical shifts as restraints^{20,21} (**Fig. 2**). The finding that both of these small, single-domain proteins fold via intermediates underlines the generic importance of partially folded species in protein-folding reactions. The conformational properties of these species show that, for these proteins, the native topology is well defined by this point in folding. Perhaps more surprisingly, for

Figure 3 Overlapping nucleation motifs in the ribosomal protein S6. (a) Above, structure of wild-type S6 (S6wt; PDB 1RIS⁹⁶) colored to show the possibility for the protein to fold via different folding nuclei: $\alpha 1$ (red) and $\alpha 2$ (blue). Both nuclei share the central $\beta 1$ strand (purple). Below, schematic of the secondary structure of S6, demonstrating overlap of the two folding nuclei. (b) Schematics demonstrating how local loop entropy influences which of the two nuclei dominates folding and, hence, the structural folding mechanism of the protein. P13-14 and P81-82 are circular permutants in which the N and C termini of the wild-type protein are linked and new termini created between positions 13 and 14, and 81 and 82, respectively. Figure redrawn from ref. 63.

both proteins substantial numbers of non-native (as well as native) contacts are formed in the intermediate ensembles^{20,21,25}, indicative of frustration in folding landscapes of even these small, simple proteins.

Direct observation of (un)folding trajectories in real time using single-molecule fluorescence techniques offers further opportunities to monitor folding reactions and to reveal rare events or species hidden by the averaging of ensemble experiments¹⁷. Using immobilization techniques on surfaces or encapsulation within liposomes to increase observation times, the first trajectories of folding reactions of individual proteins in real time are emerging^{48,49}. Although single-molecule fluorescence studies such as these should be able to expose multiple species on the reaction coordinate, significant challenges lie ahead in developing experiments to allow the properties of rapidly interconverting species to be discerned. Mechanical manipulation using optical tweezers or the atomic force microscope has already revealed the presence of intermediates when individual proteins are unfolded under force^{50,51}.

Evolution of folding pathways

The landscape view presents a powerful picture of protein folding, in that it allows a clear portrayal of the heterogeneity of species on the folding surface. It also highlights the importance of native contacts in funneling the folding chain toward the native state⁶. As a consequence, the native topology determines the sequence of folding events, rationalizing why the structural mechanism of folding is conserved in protein families⁵² (even if the kinetic mechanism (for example, two- or three-state) varies⁵³). It also explains the observed correlation between folding rate and the complexity of the native fold (contact order)⁵⁴.

Important questions result from viewing folding as a multidimensional search process. These include how many routes to the native state are taken by a folding polypeptide chain and the sensitivity of the pathways taken to the experimental conditions and protein sequence. Some proteins seem to fold via a single route through the energy landscape, as famously portrayed by chymotrypsin inhibitor 2 (ref. 55). For other proteins, the route map is more diverse^{56,57}. For multidomain proteins, the possibility of folding via parallel routes is an obvious, and real⁵⁶, possibility. Other long-established causes of parallel routes and alternative conformations involve *cis-trans* proline isomerization or disulfide oxidation⁵⁸⁻⁶⁰. In more recent work, Oliveberg and co-workers suggested that the number of pathways accessible to a polypeptide chain may be linked to the number of

