

Chapter 1

Linked Landscapes and Conformational Conversions: How Proteins Fold and Misfold

Gareth J. Morgan and Sheena E. Radford

Abstract Polypeptide chains must make a vast, complex array of interactions in order to fold to their correct structures. The mechanisms by which they do so, and the consequences when this process fails, are the subject of intense study. A key conceptual development has been the idea that proteins fold and misfold on linked, funnelled energy landscapes, where rarely populated states (such as folding intermediates) can provide access to misfolded and sometimes aggregation-prone ensembles. Advances in experimental methodologies and computer simulations are driving an increased understanding of the forces involved in folding and aggregation. This knowledge is beginning to be used to stabilise proteins for new biological functions as well as to develop treatments for diseases of misfolding. In this introductory chapter, we focus on the array of experimental methods that are being applied to the key questions of how the folding, misfolding and aggregation of proteins are linked, both in vitro and in the more complex environment of the cell.

1.1 Introduction

A protein emerging from the ribosome must make many specific contacts and interactions if it is to fold to its correct structure. That a protein can fold to a pre-defined, low-energy state, despite the myriad of competing interactions it faces, is a spectacular triumph of evolution, and it is perhaps unsurprising that this process sometimes goes awry. Misfolding and aggregation are the basis of many diseases, and these processes have been studied in great detail with a view to developing therapeutics that target specific aspects of the disease process. In this chapter, we examine key concepts and questions in the field of protein folding

S.E. Radford (✉)
University of Leeds, Mount Preston Street, Leeds LS2 9JT, UK
e-mail: s.e.radford@leeds.ac.uk

and misfolding, and review the experimental methods used to investigate these processes.

How does a polypeptide chain fold into a specific structure? Different proteins fold by different structural mechanisms [1], but the underlying theme that has emerged from theoretical and experimental studies is that the folding energy landscape is funnel-shaped, so that initial favourable interactions, some of which may be present even in the initial unfolded state [2], make further folding more favourable (Fig. 1.1) [3–5]. Protein structures are mainly governed by promiscuous non-covalent interactions, which can form between any suitable pair of residues, so there are many potential ways for non-native contacts to form. The ability of proteins to form alternative, non-native interactions, termed “frustration” [6], can be visualised as roughness on the folding energy landscape, leading to the population of transient intermediates and kinetic traps. Non-native interactions in the early stages of folding can retard or disrupt the protein’s search for its native state, but may also direct the beginning of the folding process [2, 7–9]. The native states of proteins are thought to be “minimally frustrated”, which provides their stability [10]. However, proteins are only marginally stable: indeed, many undergo conformational changes, and some even major unfolding events, as part of their function [11, 12]. So even in their native states, proteins show conformational diversity relating to binding, catalysis or regulation [13, 14]. An emerging theme is that protein structure and stability have been modulated by the requirement for function over the course of evolution [15, 16]. This compromised stability may allow conformational changes, but is also at the root of the numerous diseases of protein misfolding [17]. Misfolded forms of proteins may be inactive, toxic, or may form aggregates such as amyloid fibrils, which are ordered aggregates found in many disorders including Alzheimer’s disease and type 2 diabetes [17]. Although amyloid fibrils or other aggregates may be toxic to cells, recent work has focused on the toxicity of smaller, soluble oligomeric species, termed protofibrils, that are thought to be the underlying cause of several amyloid disorders [18].

The continuum of conformational states that proteins can populate at equilibrium, from unfolded species to partially folded, folded, misfolded and aggregated forms, defines an energy landscape on which single protein molecules can fold and misfold, and ensembles of molecules can interact to form oligomers and aggregates [19]. These two processes may be linked by unfolded species or rarely populated intermediate states that can progress into the folding or misfolding regions of the energy landscape. Figure 1.1 shows an illustration of such a linked landscape.

Five key areas of research in the areas of protein folding and misfolding are described in this chapter:

- The nature of unfolded states under non-denaturing conditions.
- The intermediates that link the folding and aggregation landscapes.
- The structural and dynamic properties of toxic oligomers.
- The structure of mature amyloid fibrils.
- How studies of folding and misfolding *in vitro* relate to the processes that occur in the cell.

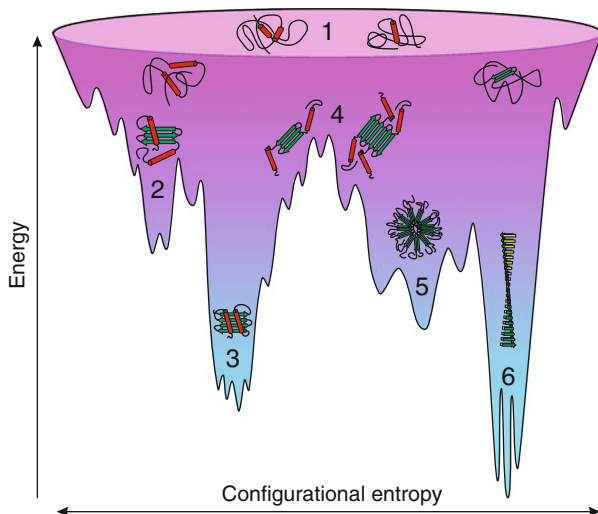


Fig. 1.1 A protein's folding and misfolding energy landscape. The energy of the protein chain is represented by the depth of the wells, and the configurational entropy of the states is represented by their width. The ruggedness of the landscape corresponds to intermediates and other metastable states. (1) Unfolded states showing varied conformations and residual structures; (2) folding intermediates; (3) native state, showing functional conformational changes; (4) misfolded intermediates and small oligomers; (5) toxic oligomers and protofibrils; (6) amyloid fibrils, which may have several distinct morphologies

We provide a brief introduction to each of these areas of research, highlighting the methodological developments that have provided recent insights. A recurring theme is the characterisation of sparsely populated, heterogeneous or dynamic conformations that are intractable to conventional high-resolution structural analysis by nuclear magnetic resonance (NMR) or X-ray crystallography. Information on these states has been gleaned from developments in methods to analyse rare conformations at equilibrium or in transition, in tandem with increasingly powerful simulations that provide detailed interpretations of these results [20,21]. A complete description of an energy landscape requires both structural and dynamic information on all the states populated. Different techniques complement each other by providing information at different resolutions, from atomic structures to overall shape and size, and on different timescales, from bond vibrations and side chain rotations to domain movements and oligomerisation. Methods used to study the folding and misfolding of proteins are shown in Fig. 1.2. These techniques can provide structural and dynamic information at various levels of resolution, and can be combined with rapid perturbation (such as continuous- or stopped-flow mixing, temperature jump and pressure jump methods) or quenching (pH jump or flash freezing) in order to study kinetic processes. As structural ensembles become more diverse, it is more difficult to extract high-resolution data; methods that report on the dynamic behaviour of proteins in solution, and thus define the size and

		Native state	Amyloid fibrils	Oligomers	Folding intermediates	Unfolded ensembles
Structural detail	X-ray crystallography	•	•			
	Solid state NMR	•	•	•	•	
	Electron microscopy		•	•		
	EPR	•	•	•	•	•
	Hydrogen exchange	•	•	•	•	•
	Single molecule FRET	•			•	•
	Intrinsic fluorescence	•		•	•	
	Infrared	•	•	•	•	•
	Raman	•	•	•	•	•
	Fluorescence anisotropy	•	•	•	•	•
Protein dynamics	Mass spectrometry	•	•	•	•	•
	Solution NMR	•			•	•
	Dye binding	•	•	•	•	•
	X-ray/Neutron/light scattering	•	•	•	•	•
	Ultracentrifugation	•	•	•	•	•
	Fluorescence correlation	•	•	•	•	•

Fig. 1.2 Methods used to investigate different states in protein folding and misfolding. Colours correspond to the position of states on the landscape in Fig. 1.1. Information about both local structures and the global dynamics of individual protein molecules and multimeric assemblies can be gained via different techniques. *NMR* nuclear magnetic resonance; *EPR* electron paramagnetic resonance; *FRET* Förster resonance energy transfer

shape of oligomeric species, therefore complement techniques able to reveal higher resolution information. An important development is the use of single molecule techniques to analyse rare events against the background of the population of molecules [22]. We consider only soluble proteins (although significant progress has been made in the field of membrane protein folding in recent years, the experimental challenges involved remain considerable [23]). More detailed reviews of folding and misfolding, and the techniques used to study them, can be found elsewhere [24, 25].

1.2 The Unfolded Ensemble Under Native Conditions

Proteins sample many states at equilibrium, including partially folded, disordered and unfolded species. High-energy species are populated to a low extent, however, and unfolded states of globular proteins have mostly been accessed by altering the solution conditions so as to preferentially stabilise unfolded states by temperature, denaturant or pH [26, 27]. How these denatured states relate to the unfolded states that the protein visits under conditions that favour folding (at physiologically relevant pH, temperature and solution conditions) is also beginning to be understood [27]. A complementary approach is to mutate or chemically

modify proteins so that the native state is significantly destabilised under ambient conditions [28, 29]. An emerging theme is that unfolded proteins are not simply random coils, but are often compact, containing residual or transient structures that may represent the starting points for folding [21]. The importance of unfolded states has been highlighted by the increasing realisation that many proteins are unstructured in their active states. These “natively unfolded” or “intrinsically disordered” proteins have many functions and are especially common in binding and signalling systems [30]. Many natively unfolded proteins can fold into a specific structure upon binding to their ligands, and the mechanisms by which this occurs may be informative for understanding folding in general. In addition, many amyloid-forming or aggregation-prone proteins are disordered in their monomeric solution states [17]. These include the short peptides such as amyloid beta (A β) – involved in Alzheimer’s disease – and the type 2 diabetes-associated islet amyloid polypeptide (IAPP). Larger, intracellular proteins such as α -synuclein and the N-terminal region of huntingtin, which are involved in Parkinson’s and Huntington’s diseases, respectively, are also unstructured *in vitro*. For these proteins, the absence of a defined globular structure is presumably important for their function, but it also allows them to access more easily aggregation-prone conformations that cause the pathologies associated with these proteins.

Solution NMR spectroscopy can provide detailed information at atomic resolution on unstructured proteins [21, 31, 32]. Although classical structural parameters such as long-range nuclear Overhauser effect (NOE) distance restraints cannot be observed for most unfolded proteins, chemical shifts, nuclear relaxation properties and residual dipolar couplings can provide significant structural information. These methods have been used to observe residual or transient structure in unfolded proteins that fluctuates on a picosecond to millisecond timescale, at the level of individual residues [33]. The transient nature of the species identified at equilibrium provides information on the diversity of states within the unfolded ensemble. Even in denaturant many proteins retain residual structure, observed through relaxation behaviour, proximity to paramagnetic moieties or residual dipolar couplings [26, 34, 35]. In the absence of denaturant, two strategies have been used to study unfolded states: either sensitive detection of rare states from within the native ensemble, or disruption of the native state by mutation. Single molecule fluorescence techniques have been particularly successful at observing rarely populated unfolded molecules [22]. This method can be used to monitor the amplitudes and timescales of distance fluctuations in unfolded states via Förster resonance energy transfer (FRET), promising very detailed descriptions of conformational fluctuations, although only a limited amount of information can be gained from a single experiment [22, 36, 37]. Studying severely destabilised variants of proteins, from which key core contacts have been removed by mutation, is a complementary method [28, 29]. Since these variants populate unfolded ensembles to a much higher extent, more mainstream structural techniques, in particular the NMR experiments described above, can be applied to gain a rich array of information [28, 29]. However, as with all mutational studies, the results need to be interpreted with caution, in case the mutations disrupt the folding landscape to too great an extent.

At high denaturant concentration, proteins behave like random coils or simple polymers, whose hydrodynamic radius is dependent on their length and degree of solvation. However, as denaturant is removed, polypeptides populate an array of more compact structures, the nature of which is more sequence dependent. These structural features may contain native-like or non-native contacts and are often only formed transiently. Interpretation of data on unfolded states is therefore best achieved by considering ensembles of structures that represent the data, rather than attempting to imagine a single dominant conformation. These residual structures have been observed in several proteins [32]. The presence of residual structure in unfolded ensembles determines the starting point for folding, and may direct the protein to fold towards its native state; alternatively, they may impede the folding process [2]. Although conformational conversions between unfolded conformations are assumed to be rapid by funnel-shaped landscape models [3], a recent intriguing study based on exhaustive simulation data suggests that unfolded states are kinetically isolated and may actually interconvert most readily via the native state [7].

1.3 Folding and Misfolding Intermediates

Transiently populated, partially folded species are ubiquitous on folding energy landscapes, and describing their properties is crucial in order to understand the folding and misfolding of proteins. Transition states and intermediates on folding pathways serve as gateways and junctions for different routes through the energy landscape, defining the likely conformational states that can be visited en route to the native state. Stable or metastable intermediates have been observed in an increasing number of small proteins, including some previously thought to fold via a two-state process, suggesting that landscape ruggedness is a ubiquitous feature of the folding of even small, simple proteins [38].

Detailed structural analysis of intermediates is difficult because of their transient nature and conformational dynamics, and this problem is exacerbated in the case of transition states, which by definition have a vanishingly short life span. Intermediates and transition states can be detected by kinetic methods. Much of the structural knowledge of these states comes from the Φ -value analyses of small proteins, whereby the effects of mutations on folding kinetics are interpreted in terms of changes in stability, and hence, structure of transient states [39]. Recently, more adventurous Φ -value analyses have been used to investigate the structural determinants of mechanical stability [40], to examine the folding of repeat proteins [41] and to monitor solvation and core packing [42]. Although Φ -values are relatively crude measures of structure, they can be used to restrain or test molecular dynamics simulations, which can reveal models of transient states in atomistic detail [20]. Φ -values obtained from continuous-flow and stopped-flow kinetic measurements of the folding of the bacterial immunity protein Im7 allowed calculations of the structures of Im7's on-pathway intermediate and two

transition state ensembles [8]. Meanwhile, *ab initio* calculations have been used to simulate the entire folding pathway of the 40-residue protein NTL9 over a millisecond timescale, and the folding trajectories compared with experimental data [43]. These computational methods allow previously invisible or ambiguous interactions to be identified, which in turn suggest further experimental work, and can help identify patterns between different proteins. Simulations of many folding trajectories uniquely allow important structural intermediates and metastable states to be identified that are invisible to experimental approaches. These include very early collapsed species, which direct the initial stages of folding, and high-energy intermediates that are kinetically invisible [7].

Solution NMR techniques can reveal atomic resolution information about intermediates that are visited at equilibrium. Hydrogen exchange kinetics report on the hydrogen bonding and solvent accessibility of exchange-labile hydrogen atoms within a protein, and are particularly sensitive to the presence of secondary structure elements [44]. Hierarchical unfolding of subdomains (known as “foldons”) of cytochrome *c* has been described in great detail [45]. In some cases, it is possible to correlate dynamics measured by hydrogen exchange with specific folding intermediates, as described for Im7 [46]. Alternatively, folding and hydrogen exchange kinetics may report on different parts of the energy landscape, as is the case for the ribosomal protein S6 [47]. Another method that has allowed partially folded species to be observed is relaxation dispersion NMR, which allows the presence of “invisible” species populated to <1% at equilibrium to be detected via their effect on the relaxation of NMR signals [48]. Chemical exchange on the timescale of an NMR experiment (typically milliseconds) causes line broadening in the spectra, and precise quantitation of this effect allows information about the structure and dynamics of the non-native species to be extracted. This effect was first used to discover a hidden intermediate in the folding of an SH3 domain and the chemical shift changes between the native and intermediate states used to restrain a molecular dynamics simulation to generate an all-atom model of the intermediate ensemble [49]. Finally, a recent approach using flash-freezing and solid-state NMR (ssNMR) offers the opportunity to gain high-resolution information on freeze-trapped intermediate states of even fast-folding proteins [50].

An alternative approach to observing intermediates is to use single molecule techniques to detect species that are obscured by ground states at equilibrium. Single molecule FRET has been used to measure distance distributions of many individual proteins to build up a description of the protein’s folding energy landscape [22]. At least in theory, these techniques allow transitions of individual molecules to be characterised. However, the extremely rapid timescales of these events present a formidable technical challenge. Recent progress by Eaton and co-workers has determined an upper time limit for these transitions of 200 μ s for the protein GB1 immobilised on a surface, which is 10,000 times faster than the average folding time [51].

The soluble, globular form of a protein may be dramatically different from its conformation within an amyloid fibril. Understanding the structural transitions between these two states, therefore, requires that the intermediates and transition

states on the pathway to the misfolded state are characterised. Folding intermediates may be a crucial part of the process, as they may be aggregation-prone themselves, or have the conformational flexibility needed to access the direct precursors for amyloid formation [19, 25]. Understanding a protein's folding mechanism can therefore provide a gateway to studying the transient states that link the folding and aggregation regions of the protein's conformational energy landscape (Fig. 1.1). Such intermediates were first observed in mutants of lysozyme [52] and have now been characterised for several proteins [25, 53, 54]. A well-studied example of such a state is a folding intermediate of β_2 -microglobulin (β_2m) containing a non-native *trans* proline bond at position 32. Population of this species correlates with the rate of amyloid elongation, suggesting that this conformation is a gateway to further misfolding or oligomerisation [55]. Amyloid formation by β_2m under physiological conditions can be initiated by Cu (II) binding to a species that has the same *trans* proline residue [56]. This conformation may be a useful target for drugs designed to prevent dialysis-related amyloidosis, which is caused by deposition of β_2m . This knowledge-based drug development is exemplified by attempts to prevent amyloidosis of the human protein transthyretin (TTR), which is associated with familial amyloid polyneuropathy [53]. The folding and misfolding landscape of TTR is well characterised, enabling drugs to be designed that stabilise the native tetramer of TTR, thus preventing amyloidosis. These compounds are among the first drugs for amyloid disease in clinical trials [53].

1.4 Protofibrils, Oligomers and Toxicity

For a protein to aggregate, molecules in an aggregation-competent conformation must encounter one another and form oligomeric species that eventually grow into large assemblies such as amyloid. Although amyloid fibrils have long been associated with disease, it is not clear how the mature fibrils contribute to the pathogenesis of amyloid disorders [17]. A turning point was the observation that progression of Alzheimer's disease correlates with the concentration of soluble, oligomeric forms of the 42-residue isoform of the amyloid beta peptide ($A\beta_{42}$), rather than levels of amyloid plaques [18]. Soluble oligomers, also known as protofibrils, have since been implicated as toxic species in several amyloid diseases, although the precise roles of fibrils and oligomers in the progression of disease remain unclear. A wide array of oligomeric species is formed by different proteins, which populate a heterogeneous array of oligomeric species depending on the experimental conditions. Structural information about these oligomers is limited by their heterogeneity and dynamic behaviour, and characterising these species is an important future challenge for the field. An important observation is that toxic, soluble oligomers formed from several different proteins are recognised by a conformation-specific antibody, called A11, suggesting a common underlying structure and, possibly, mechanism of toxicity [57]. Electron paramagnetic resonance (EPR) measures the mobility of nitroxide spin labels and their proximity to one another. The behaviour of a spin label

attached to A β ₄₀ has been shown to vary between fibrils and two different oligomeric states, supporting the theory that A11-reactive oligomers have a distinctly different structure to both other oligomeric species and mature fibrils [58].

New experimental methods are allowing transiently populated oligomeric species en route to amyloid formation to be characterised. An important development is the use of ion mobility mass spectrometry to resolve and measure the mass and conformational properties of pre-amyloid oligomers. This technique relies on separation of ions in the gas phase, based on their mass, cross-sectional area and charge, and allows mass spectra to be deconvoluted. Diverse populations of oligomers have been observed during the fibrillogenesis of A β [59] and β _{2m} [60], and distinct species of monomeric IAPP have been identified by similar methods [61]. Importantly, the time resolution of these experiments allows the dynamics of these species to be studied: mixing samples of ¹⁴N- and ¹⁵N-labelled protein at different stages of oligomer formation and monitoring the rate of their interconversion [60]. Atomic resolution structural data on oligomeric intermediates is difficult to obtain, but an ssNMR investigation into the structure of a neurotoxic, spherical intermediate formed by A β ₄₀ showed that the peptide had a conformation very similar to that found within mature fibrils [62].

Alongside experimental observations of the process of amyloid formation, simulations and modelling are increasing our understanding of these processes. A range of computational techniques have been applied to the study of protein aggregation, from mathematical approaches that model the entire population of species [63, 64], through coarse-grained simulations of many peptides interacting, to atomistic simulations of the precise nature of the aggregation process [65]. As in the case of folding intermediates, the combination of experimental and computational data is greatly increasing the depth of understanding of these systems. Computational methods can report on the entire population of misfolding molecules, without the ambiguity inherent in the low-resolution ensemble methods by which these processes are followed. The mechanisms involved in the initiation of amyloid fibril formation and their subsequent elongation remain poorly characterised at a molecular level. Future investigations using the techniques described above, as well as further method development, should continue to increase our knowledge of this important research area.

1.5 Amyloid Structure

Through an electron microscope, amyloid fibrils formed from a wide variety of structurally diverse proteins look remarkably similar: they are long, unbranched, straight structures with a periodic twist and are usually around 10-nm wide [17]. Combined with their common cross- β architecture, revealed by X-ray fibre diffraction [66], and the observation that almost any protein can form amyloid under suitable conditions [67], the structural similarity in amyloid fibril architecture suggests that amyloid may be a universal alternative conformation for all

polypeptides [17]. However, closer examination of the structure of fibrils formed by different proteins by ssNMR and cryo-electron microscopy (cryo-EM) has revealed a diverse array of architectures and organisation of the β -strands, even among fibrils formed from the same polypeptide sequence [68–70], leaving open the question of how similar amyloid fibrils really are at the atomic level. Since they are large, non-crystalline aggregates, amyloid fibrils are intractable to either conventional X-ray crystallography or high-resolution studies by solution NMR. However, recent developments in these techniques, as well as in ssNMR and cryo-EM, have now provided detailed structural models of several amyloid fibrils [71–75]. Further insights can be obtained from other spectroscopic methods, particularly EPR [76] and infrared (IR) spectroscopies. IR is particularly useful in the study of amyloid fibrils and their precursors, since it is sensitive to changes in β -sheet structure [77]. Many amyloid fibrils have a distinctive strong peak in the amide I band around $1,620\text{ cm}^{-1}$, which can be used to follow the formation of ordered β -sheet structure [78]. Isotopic labelling with ^2H , ^{13}C or ^{18}O alters the resonance frequencies of bonds and can be used to increase resolution within samples. Developments in 2D IR [77] and in the specific isotopic labelling of proteins [79] hold significant promise in this field, as described elsewhere in this volume.

Cryo-EM methods allow the direct visualisation of individual amyloid fibrils. Reconstructions of amyloid fibrils from different proteins have revealed substantial variation from the simple ladder cross- β models once envisioned [68]. Although EM reconstructions of amyloid fibrils do not yet approach atomic resolution, mainly due to the inherent heterogeneity of the samples, the details of the fibril architectures revealed are beautiful, fascinating and perplexing in equal measure. For example, the diversity of amyloid architectures observed in fibrils of $\beta_2\text{m}$ [68] and A β [69, 70], even within a single sample, suggests that fibril conformations are sensitively balanced and affected by many factors. Hydrogen exchange kinetics can be used to monitor solvent exposure and hydrogen bonding in amyloid fibrils [80]. Although the fibrils themselves are not amenable to direct analysis, the exchange reaction can be quenched and the fibrils dissociated using aprotic organic solvents. This maintains the exchange patterns of individual residues, which can then be analysed by mass spectrometry or solution NMR. These two detection methods are complementary: MS providing information about populations of molecules, while NMR can supply residue-specific information. The structured cores of several amyloid fibrils, as well as pre-fibrillar oligomers [81], have been mapped in this way, and the exchange kinetics have been interpreted as representing dynamic recycling of molecules from fibril ends [82]. EPR measurements can also provide information about the local structure and dynamics of spin labels introduced into amyloid fibrils. The structured regions of amyloid fibrils formed from $\beta_2\text{m}$ were mapped by measuring spin label mobility and thiol accessibility, and the presence of spin–spin interactions suggested a parallel, in-register structure for these fibrils [83], in common with other fibrils created from unrelated protein sequences [73, 74, 76].

An important development in recent years has been the determination of the first atomic resolution structures of amyloid fibrils. X-ray crystal structures of

several short peptides in an amyloid-like conformation have been solved [71, 72]. Microcrystals with needle-like shapes were selected from solutions, which also encouraged the growth of amyloid fibrils. Their structures revealed precise conformations of different peptides that are consistent with subtle variations on the cross- β architecture, including the arrangement of β -strands within and between the β -sheets of the fibrils. Six out of eight possible topologies and multiple inter-sheet interfaces have been observed in crystals and these data have been used to develop models of other amyloid fibrils [72]. Although crystals of only small peptides have been successfully grown using this technique, the information has provided an important validation of previous models of amyloid fibrils based on the cross- β architecture.

Solid-state NMR can, in theory, resolve molecular structures of amyloid fibrils in great detail. However, spectral crowding means that the interpretation of spectra is difficult and laborious, generally requiring many samples, isotopically labelled at specific sites. Detailed models of A β ₄₀ [74] and IAPP [73] have been published, which show stacked, in-register β -strands in a strand-turn-strand motif with tightly packed cores. The fungal prion HET-s, the largest protein for which the structure of an amyloid fibril has been determined, has a more complex solenoid structure, where each molecule forms two loops, contributing two β -strands to each of four β -sheets that run parallel to the fibril axis [75]. Although these structures represent a major step in our understanding of amyloid structure, it is not yet clear how the conformation of polypeptides within an amyloid fibril determines their overall gross morphology, mechanical properties or any interactions the fibrils may make in vivo. Correlating the wide array of structural information with biological effects remains a major challenge for the future.

1.6 From the Test Tube to the Cell

Much of our knowledge of protein folding and misfolding has come from studying pure, isolated proteins in vitro. However, the environment of a cell in which a newly synthesised protein must fold is far more complex, and translating knowledge of a protein's folding mechanism in vitro into the context of cell remains a challenge [84]. Experimental approaches and analytical tools are being developed to monitor protein folding and misfolding in vivo. The cellular environment influences all parts of a protein's folding and misfolding landscape, from interactions of the nascent chain with the ribosome on which it is synthesised, through binding to chaperones and functional partners, to the eventual destruction of a protein by regulated proteolysis. Figure 1.3 describes some of the states that a protein chain can form within cells and displays the dynamic interplay of different factors in maintaining cellular homeostasis.

The first interaction that a protein makes in the cell is with the ribosome on which it is synthesised. Solution state NMR of ribosome-bound nascent chains has been used to demonstrate formation of tertiary structure in domains still tethered to the ribosome [85] and has shown that the nascent chain of an SH3 domain does

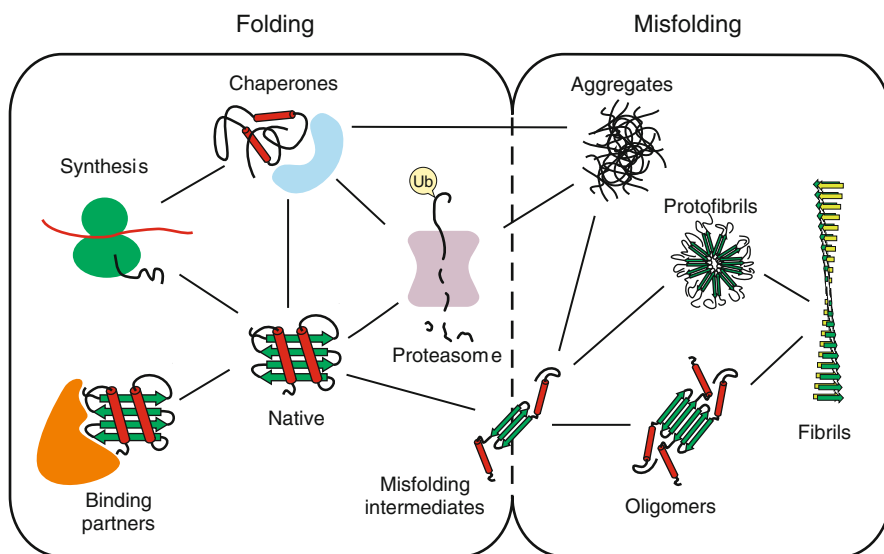


Fig. 1.3 Proteins interact with many partners in the cell, which may influence their folding and misfolding. After synthesis on ribosomes, proteins may interact with chaperones or binding partners and may be targeted for degradation by the proteasome by ubiquitination (indicated by “Ub”). Chaperones and disaggregases may prevent the formation of misfolded species, or perhaps sequester toxic oligomers into inert amyloid fibrils

not fold until it has fully emerged from the peptide exit tunnel [86]. Many proteins cannot fold efficiently alone and require the assistance of chaperones to reach their native state [87]. The chemical modification of proteins by the covalent addition of, among others, phosphate, sugars or even polypeptides such as ubiquitin plays critical roles in their functions, and the consequences of these post-translational modifications on folding have recently begun to be investigated in molecular detail. Glutathionation of superoxide dismutase, for example, promotes dimer dissociation and thereby increases amyloid propensity [88]. Direct observations of folding *in vivo* are complicated by the difficulty of isolating signals from the protein of interest from those arising from other cellular components. Specific fluorescent labelling using dyes has allowed direct observation of folding and unfolding events within cells [89], while proteins’ movements and interactions within cells can be readily tracked using genetic fusions with fluorescent proteins [90]. Another challenge for these studies is the difficulty of perturbing the protein’s environment without damaging the cell. A novel method using rapid laser temperature jump *in vivo* is one solution to this problem [91]. Solution NMR of labelled proteins within living cells holds great promise in the study of proteins in their native environment, including their dynamics and interactions with chaperones [92]. Initial results from these techniques show that the broad conclusions from *in vitro* studies remain valid within cells, but subtle differences may modulate the folding landscape in important ways.

Cells have a complex network of systems whose role is to maintain the integrity of their proteins. The holistic function, termed “protein homeostasis” or “proteostasis”, of these networks is beginning to be investigated [93]. Misfolding diseases are increasingly being seen as caused by a failure of the cell’s proteostasis machinery. However, the mechanisms of toxicity, or even the identity of toxic species, involved with misfolding and amyloid formation are not clear. As protein conformational diversity is revealed in more detail, and as new chaperones continue to be identified, the challenge of integrating these data into a working model of how proteins fold and misfold within cells becomes more complex [84]. A systems biology approach to the folding and export of aggregation-prone proteins from the endoplasmic reticulum has been used to rationalise the effects of destabilising mutations on different model proteins [94], and these integrative approaches offer a way to make sense of the ever-expanding array of information gained from an increasingly diverse range of experiments.

1.7 Conclusions

Our knowledge of how proteins fold continues to expand. The folding energy landscapes of many soluble proteins have been investigated in great detail through a combination of the techniques described above, and structural and kinetic information is now available for states which had previously resisted atomic characterisation. Although great progress has been made, many areas remain to be fully investigated. Open questions include the nature of unfolded states in different proteins, the structural transitions that link folding and misfolding and the mechanisms of toxicity in diseases of misfolding. The remainder of this volume describes the contribution to this important and burgeoning field that is being made using infrared methods that extend and complement the biophysical approaches described here.

References

1. A. Zarrine-Afsar, S.M. Larson, A.R. Davidson, *Curr. Opin. Struct. Biol.* **15**, 42 (2005)
2. M.A.C. Reed, C. Jelinska, K. Syson, M.J. Cliff, A. Splevins, T. Alizadeh, A.M. Hounslow, R.A. Staniforth, A.R. Clarke, C. Jeremy Craven, J.P. Waltho, *J. Mol. Biol.* **357**, 365 (2006)
3. J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, *Proteins* **21**, 167 (1995)
4. K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl, *Annu. Rev. Biophys.* **37**, 289 (2008)
5. M. Karplus, *Fold. Des.* **2**, S69 (1997)
6. D. Bryngelson, P.G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524 (1987)
7. G.R. Bowman, V.S. Pande, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10890 (2010)
8. C.T. Friel, D.A. Smith, M. Vendruscolo, J. Gsponer, S.E. Radford, *Nat. Struct. Mol. Biol.* **16**, 318 (2009)
9. A. Zarrine-Afsar, S. Wallin, A.M. Neculai, P. Neudecker, P.L. Howell, A.R. Davidson, H.S. Chan, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9999 (2008)

10. D.U. Ferreira, J.A. Hegler, E.A. Komives, P.G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19819 (2007)
11. S. Hammes-Schiffer, S.J. Benkovic, *Annu. Rev. Biochem.* **75**, 519 (2006)
12. E.M. Puchner, A. Alexandrovich, A.L. Kho, U. Hensen, L.V. Schafer, B. Brandmeier, F. Grater, H. Grubmuller, H.E. Gaub, M. Gautel, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13385 (2008)
13. O.F. Lange, N.A. Lakomek, C. Fares, G.F. Schroder, K.F. Walter, S. Becker, J. Meiler, H. Grubmuller, C. Griesinger, B.L. de Groot, *Science* **320**, 71 (2008)
14. J.S. Fraser, M.W. Clarkson, S.C. Degnan, R. Erion, D. Kern, T. Alber, *Nature* **462**, 669 (2009)
15. M.A. DePristo, D.M. Weinreich, D.L. Hartl, *Nat. Rev. Genet.* **6**, 678 (2005)
16. L. Foit, G.J. Morgan, M.J. Kern, L.R. Steimer, A.A. von Hacht, J. Titchmarsh, S.L. Warriner, S.E. Radford, J.C. Bardwell, *Mol. Cell.* **36**, 861 (2009)
17. F. Chiti, C.M. Dobson, *Annu. Rev. Biochem.* **75**, 333 (2006)
18. J. Hardy, D.J. Selkoe, *Science* **297**, 353 (2002)
19. T.R. Jahn, S.E. Radford, *FEBS J.* **272**, 5962 (2005)
20. M. Vendruscolo, *Curr. Opin. Struct. Biol.* **17**, 15 (2007)
21. H.J. Dyson, P.E. Wright, *Chem. Rev.* **104**, 3607 (2004)
22. B. Schuler, W.A. Eaton, *Curr. Opin. Struct. Biol.* **18**, 16 (2008)
23. P.J. Booth, J. Clarke, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3947 (2010)
24. A.I. Bartlett, S.E. Radford, *Nat. Struct. Mol. Biol.* **16**, 582 (2009)
25. F. Chiti, C.M. Dobson, *Nat. Chem. Biol.* **5**, 15 (2009)
26. C.S. Le Duff, S.B. Whittaker, S.E. Radford, G.R. Moore, *J. Mol. Biol.* **364**, 824 (2006)
27. Y. Li, F. Picart, D.P. Raleigh, *J. Mol. Biol.* **349**, 839 (2005)
28. B. Anil, Y. Li, J.H. Cho, D.P. Raleigh, *Biochemistry* **45**, 10110 (2006)
29. T.L. Religa, J.S. Markson, U. Mayor, S.M. Freund, A.R. Fersht, *Nature* **437**, 1053 (2005)
30. H.J. Dyson, P.E. Wright, *Nat. Rev. Mol. Cell. Biol.* **6**, 197 (2005)
31. D. Eliezer, *Curr. Opin. Struct. Biol.* **19**, 23 (2009)
32. T. Mittag, J.D. Forman-Kay, *Curr. Opin. Struct. Biol.* **17**, 3 (2007)
33. A.K. Mittermaier, L.E. Kay, *Trends Biochem. Sci.* **34**, 601 (2009)
34. M.J. Cliff, C.J. Craven, J.P. Marston, A.M. Hounslow, A.R. Clarke, J.P. Waltho, *J. Mol. Biol.* **385**, 266 (2009)
35. R. Mohana-Borges, N.K. Goto, G.J. Kroon, H.J. Dyson, P.E. Wright, *J. Mol. Biol.* **340**, 1131 (2004)
36. S.D. Pugh, C. Gell, D.A. Smith, S.E. Radford, D.J. Brockwell, *J. Mol. Biol.* **398**, 132 (2010)
37. D. Nettels, A. Hoffmann, B. Schuler, *J. Phys. Chem. B* **12**, 6137 (2008)
38. D.J. Brockwell, S.E. Radford, *Curr. Opin. Struct. Biol.* **17**, 30 (2007)
39. A. Matouschek, J.T. Kellis Jr., L. Serrano, A.R. Fersht, *Nature* **340**, 122 (1989)
40. D.P. Sadler, E. Petrik, Y. Taniguchi, J.R. Pullen, M. Kawakami, S.E. Radford, D.J. Brockwell, *J. Mol. Biol.* **393**, 237 (2009)
41. N. Courtemanche, D. Barrick, *Structure* **16**, 705 (2008)
42. A.I. Bartlett, S.E. Radford, *J. Mol. Biol.* **396**, 1329 (2010)
43. V.A. Voelz, G.R. Bowman, K. Beauchamp, V.S. Pande, *J. Am. Chem. Soc.* **132**, 1526 (2010)
44. S.W. Englander, *Annu. Rev. Biophys.* **29**, 213 (2000)
45. M.M. Krishna, H. Maity, J.N. Rumbley, Y. Lin, S.W. Englander, *J. Mol. Biol.* **359**, 1410 (2006)
46. S.A. Gorski, C.S. Le Duff, A.P. Capaldi, A.P. Kalverda, G.S. Beddard, G.R. Moore, S.E. Radford, *J. Mol. Biol.* **337**, 183 (2004)
47. E. Haglund, J. Lind, T. Oman, A. Ohman, L. Maler, M. Oliveberg, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21619 (2009)
48. P. Neudecker, P. Lundstrom, L.E. Kay, *Biophys. J.* **96**, 2045 (2009)
49. D.M. Korzhnev, X. Salvatella, M. Vendruscolo, A.A. Di Nardo, A.R. Davidson, C.M. Dobson, L.E. Kay, *Nature* **430**, 586 (2004)
50. K.N. Hu, W.M. Yau, R. Tycko, *J. Am. Chem. Soc.* **132**, 24 (2010)
51. H.S. Chung, J.M. Louis, W.A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11837 (2009)
52. D.R. Booth, M. Sunde, V. Bellotti, C.V. Robinson, W.L. Hutchinson, P.E. Fraser, P.N. Hawkins, C.M. Dobson, S.E. Radford, C.C. Blake, M.B. Pepys, *Nature* **385**, 787 (1997)

53. S. Connelly, S. Choi, S.M. Johnson, J.W. Kelly, I.A. Wilson, *Curr. Opin. Struct. Biol.* **20**, 54 (2010)
54. R.A. Staniforth, S. Giannini, L.D. Higgins, M.J. Conroy, A.M. Hounslow, R. Jerala, C.J. Craven, J.P. Waltho, *EMBO J.* **20**, 4774 (2001)
55. T.R. Jahn, M.J. Parker, S.W. Homans, S.E. Radford, *Nat. Struct. Mol. Biol.* **13**, 195 (2006)
56. C.M. Eakin, A.J. Berman, A.D. Miranker, *Nat. Struct. Mol. Biol.* **13**, 202 (2006)
57. R. Kaye, E. Head, J.L. Thompson, T.M. McIntire, S.C. Milton, C.W. Cotman, C.G. Glabe, *Science* **300**, 486 (2003)
58. J.W. Wu, L. Breydo, J.M. Isas, J. Lee, Y.G. Kuznetsov, R. Langen, C. Glabe, *J. Biol. Chem.* **285**, 6071 (2010)
59. S.L. Bernstein, N.F. Dupuis, N.D. Lazo, T. Wyttenbach, M.M. Condron, G. Bitan, D.B. Teplow, J.-E. Shea, B.T. Ruotolo, C.V. Robinson, M.T. Bowers, *Nat. Chem.* **1**, 326 (2009)
60. D.P. Smith, S.E. Radford, A.E. Ashcroft, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6794 (2010)
61. N.F. Dupuis, C. Wu, J.E. Shea, M.T. Bowers, *J. Am. Chem. Soc.* **131**, 18283 (2009)
62. S. Chimon, M.A. Shaibat, C.R. Jones, D.C. Calero, B. Aizezi, Y. Ishii, *Nat. Struct. Mol. Biol.* **14**, 1157 (2007)
63. W.F. Xue, S.W. Homans, S.E. Radford, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8926 (2008)
64. T.P.J. Knowles, C.A. Waudby, G.L. Devlin, S.I.A. Cohen, A. Aguzzi, M. Vendruscolo, E.M. Terentjev, M.E. Welland, C.M. Dobson, *Science* **326**, 1533 (2009)
65. B. Ma, R. Nussinov, *Curr. Opin. Chem. Biol.* **10**, 445 (2006)
66. M. Sunde, L.C. Serpell, M. Bartlam, P.E. Fraser, M.B. Pepys, C.C. Blake, *J. Mol. Biol.* **273**, 729 (1997)
67. F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, C.M. Dobson, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3590 (1999)
68. H.E. White, J.L. Hodgkinson, T.R. Jahn, S. Cohen-Krausz, W.S. Gosal, S. Muller, E.V. Orlova, S.E. Radford, H.R. Saibil, *J. Mol. Biol.* **389**, 8 (2009)
69. J. Meinhardt, C. Sachse, P. Hortschansky, N. Grigorieff, M. Fändrich, *J. Mol. Biol.* **386**, 869 (2009)
70. A.K. Paravastu, R.D. Leapman, W.M. Yau, R. Tycko, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18349 (2008)
71. R. Nelson, M.R. Sawaya, M. Balbirnie, A.O. Madsen, C. Riek, R. Grothe, D. Eisenberg, *Nature* **435**, 773 (2005)
72. M.R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S.A. Sievers, M.I. Apostol, M.J. Thompson, M. Balbirnie, J.J. Wiltzius, H.T. McFarlane, A.O. Madsen, C. Riek, D. Eisenberg, *Nature* **447**, 453 (2007)
73. S. Luca, W.M. Yau, R. Leapman, R. Tycko, *Biochemistry* **46**, 13505 (2007)
74. A.T. Petkova, Y. Ishii, J.J. Balbach, O.N. Antzutkin, R.D. Leapman, F. Delaglio, R. Tycko, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16742 (2002)
75. C. Wasmer, A. Lange, H. van Melckebeke, A.B. Siemer, R. Riek, B.H. Meier, *Science* **319**, 1523 (2008)
76. M. Margittai, R. Langen, *Q. Rev. Biophys.* **41**, 265 (2008)
77. Z. Ganim, H.S. Chung, A.W. Smith, L.P. Deflores, K.C. Jones, A. Tokmakoff, *Acc. Chem. Res.* **41**, 432 (2008)
78. G. Zandomenighi, M.R. Krebs, M.G. McCammon, M. Fändrich, *Protein Sci.* **13**, 3314 (2004)
79. S.-H. Shim, R. Gupta, Y.L. Ling, D.B. Strasfeld, D.P. Raleigh, M.T. Zanni, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6614 (2009)
80. M. Hoshino, H. Katou, K. Yamaguchi, Y. Goto, *Biochim. Biophys. Acta* **1768**, 1886 (2007)
81. N. Carulla, M. Zhou, M. Arimon, M. Gairi, E. Giral, C.V. Robinson, C.M. Dobson, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7828 (2009)
82. N. Carulla, G.L. Caddy, D.R. Hall, J. Zurdo, M. Gairi, M. Feliz, E. Giral, C.V. Robinson, C.M. Dobson, *Nature* **436**, 554 (2005)
83. C.L. Ladner, M. Chen, D.P. Smith, G.W. Platt, S.E. Radford, R. Langen, *J. Biol. Chem.* **285**, 17137 (2010)
84. L.M. Gierasch, A. Gershenson, *Nat. Chem. Biol.* **5**, 774 (2009)

85. S.T. Hsu, P. Fucini, L.D. Cabrita, H. Launay, C.M. Dobson, J. Christodoulou, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16516 (2007)
86. C. Eichmann, S. Preissler, R. Riek, E. Deuerling, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9111 (2010)
87. F.U. Hartl, M. Hayer-Hartl, *Nat. Struct. Mol. Biol.* **16**, 574 (2009)
88. K.C. Wilcox, L. Zhou, J.K. Jordon, Y. Huang, Y. Yu, R.L. Redler, X. Chen, M. Caplow, N.V. Dokholyan, *J. Biol. Chem.* **284**, 13940 (2009)
89. Z. Ignatova, L.M. Gierasch, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 523 (2004)
90. J. Lippincott-Schwartz, E. Snapp, A. Kenworthy, *Nat. Rev. Mol. Cell. Biol.* **2**, 444 (2001)
91. S. Ebbinghaus, A. Dhar, J.D. McDonald, M. Gruebele, *Nat. Methods* **7**, 319 (2010)
92. K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hiroaki, M. Shirakawa, *Nature* **458**, 106 (2009)
93. W.E. Balch, R.I. Morimoto, A. Dillin, J.W. Kelly, *Science* **319**, 916 (2008)
94. R.L. Wiseman, E.T. Powers, J.N. Buxbaum, J.W. Kelly, W.E. Balch, *Cell* **131**, 809 (2007)

Protein Folding and Misfolding
Shining Light by Infrared Spectroscopy
Fabian, H.; Naumann, D. (Eds.)
2012, XVI, 244 p., Hardcover
ISBN: 978-3-642-22229-0