References

1 Proc. Cold Spring Harbor Symposium on Quantitative Biology (1973) 37
2 Geeves, M.A. and Holmes, K.C. (1999) Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* 68, 687–728
3 Gordon, A.M. *et al.* (2000) Regulation of contraction in striated muscle. *Physiol. Rev.* 80, 853–924
4 Goldman, R., Pollard, T.D. and Rosenbaum, J. (eds) (1976) *Cell Motility*, Cold Spring Harbor Laboratory Press
5 Corrie, J.E.T. *et al.* (1999) Dynamic measurement of myosin light-chain-domain tilt and twist in muscle contraction. *Nature* 400, 425–430
6 Mehta, A.D. *et al.* (1999) Biomechanics, one molecule at a time. *J. Biol. Chem.* 274, 14517–14520
7 Kreis, T. and Vale, R. (eds) (1999) *Guidebook to the Cytoskeletal and Motor Proteins* (2nd edn), Oxford University Press
8 Mermall, V. *et al.* (1998) Unconventional myosins in cell movement, membrane traffic and signal transduction. *Science* 279, 527–533
9 Schafer, D.A. and Schroer, T.A. (1999) Actin-related proteins. *Annu. Rev. Cell Dev. Biol.* 15, 341–365
10 Littlefield, R. and Fowler, V.M. (1998) Defining actin filament length in striated muscle: rulers and caps or dynamic instability? *Annu. Rev. Cell Biol.* 14, 487–525
11 Trinick, J. and Tskhovrebova, L. (1999) Titin: a molecular control freak. *Trends Cell Biol.* 9, 377–380
12 Pollard, T.D. *et al.* (1994) Structure of actin binding proteins: insights about function at atomic resolution. *Annu. Rev. Cell Biol.* 10, 207–249
13 McLaughlin, P.J. and Weeds, A.G. (1995) Actin-binding protein complexes at atomic resolution. *Annu. Rev. Biophys. Biomol. Struct.* 24, 643–675
14 Vale, R.D. and Milligan, R.A. (2000) The way things move: looking under the hood of molecular motor proteins. *Science* 288, 88–95
15 De La Cruz, E.M. *et al.* (1999) The kinetic mechanism of myosin V. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13726–13731
16 Rief, M. *et al.* (2000) Myosin-V stepping kinetics: a molecular model for processivity. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9482–9486
17 Wells, A.L. *et al.* (1999) Myosin-VI is an actin-based motor that moves backwards. *Nature* 401, 505–508
18 Pollard, T.D. *et al.* (2000) Biophysics of actin filament dynamics in nonmuscle cells. *Annu. Rev. Biophys. Biomol. Struct.* 29, 545–576
19 Dramsi, S. and Cossart, P. (1998) Intracellular pathogens and the actin cytoskeleton. *Annu. Rev. Cell Dev. Biol.* 14, 137–166
20 Towbin, J.A. (1998) The role of cytoskeletal proteins in cardiomyopathies. *Curr. Opin. Cell Biol.* 10, 131–139
21 Brown, S.C. and Lucy, J. (1997) *Dystrophin: Gene, Protein and Cell Biology*, Cambridge University Press
22 Higgs, H.N. and Pollard, T.D. (1999) Regulation of actin polymerization by Arp2/3 complex and WASp/Scar proteins. *J. Biol. Chem.* 274, 32531–32534
23 Kaibuchi, K. *et al.* (1999) Regulation of the cytoskeleton and cell adhesion by the Rho family of GTPases in mammalian cells. *Annu. Rev. Biochem.* 68, 459–486
24 Somlyo, A.P. and Somlyo, A.V. (2000) Signal transduction by G-proteins, rho-kinase and protein phosphatase to smooth muscle and non-muscle myosin-II. *J. Physiol.* 522, 177–185
25 Lymn, R.W. and Taylor, E.W. (1971) Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry* 10, 4617–4623
26 Kabsch, W. *et al.* (1990). Atomic structure of the actin–DNase I complex. *Nature* 347, 37–44
27 Rayment, I. *et al.* (1993) Three dimensional structure of myosin subfragment-1: a molecular motor. *Science* 261, 50–58

Protein folding: progress made and promises ahead

Sheena E. Radford

Over the past 25 years, enormous breakthroughs have been made in understanding protein folding mechanisms. We have now reached an exciting stage, with consensuses beginning to emerge that combine both theoretical and experimental approaches. In addition, new fields have emerged and burgeoned, including *in vivo* folding and the study of protein misfolding diseases. In today's post-genomic world, understanding protein folding has never been more important and the topic has wide-ranging impact in fields from structural biology to materials science.

THE VAST MAJORITY of proteins must fold into a unique three-dimensional (3D) structure to be biologically active, yet proteins are synthesized as linear chains of amino acids on ribosomes. As a consequence, each newly synthesized chain has to navigate its way to a unique, active conformation following its synthesis. Understanding how this process is achieved has intrigued scientists for decades. A solution to the protein folding problem is of enormous intellectual importance, in that it will provide the 'missing link' in the flow of information between a gene sequence and the 3D structure of a protein. An appreciation of protein folding will also have far-reaching implications; for example, in the fine tuning of structure-prediction algorithms and the rational design of novel sequences not provided by evolution, as well as in fields ranging from medicine to nanotechnology.

This is strikingly portrayed by the graph in Fig. 1, which shows the growth in the number of manuscripts published on protein folding over the past 40 years. Clearly, the field of folding has moved on from a fascinating but rather difficult topic that was in its infancy some 25 years ago, to a burgeoning field that has impact across science and technology. A comprehensive review of this now enormous field would be impossible for an article of this kind. In this article, therefore, I discuss recent developments in our understanding of protein folding mechanisms. This was the focus of a review by Barry Robson in the first issue of *TiBS*, which was the mainstay of the field at the time[1]. In addition, I describe the insights that have emerged from the relatively new fields of chaperone-assisted folding and protein misfolding diseases.

S.E. Radford is at the School of Biochemistry and Molecular Biology, University of Leeds, Leeds, UK LS2 9JT.
Email: s.e.radford@leeds.ac.uk

The view in 1976

In his review of protein folding 25 years ago, Robson stated that 'the mechanism by which a protein adopts its biologically active conformation after biosynthesis is becoming a little clearer'[1]. Even at this early stage in the field (15 years or so after Anfinsen first studied protein refolding of ribonuclease A and only 3 years after he received the Nobel prize for his work[2]), the foundations of modern views of folding had been laid. Progress was remarkable given the scarcity of proteins

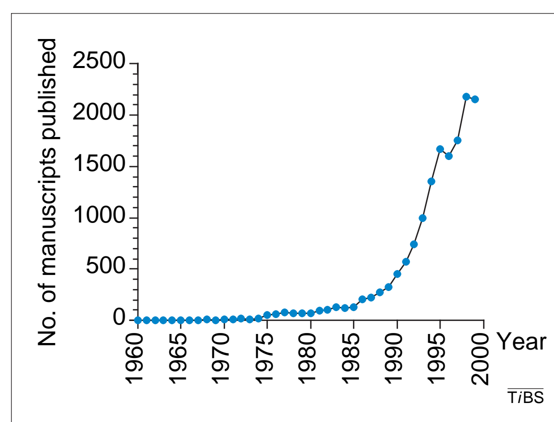

Figure 1
The number of manuscripts published each year (1960–1999) with the words 'protein' and 'folding' in either the title or the abstract. The data were taken from the PubMed database.

0968–0004/00/$ – See front matter © 2000, Elsevier Science Ltd. All rights reserved. PII: S0968-0004(00)01707-2

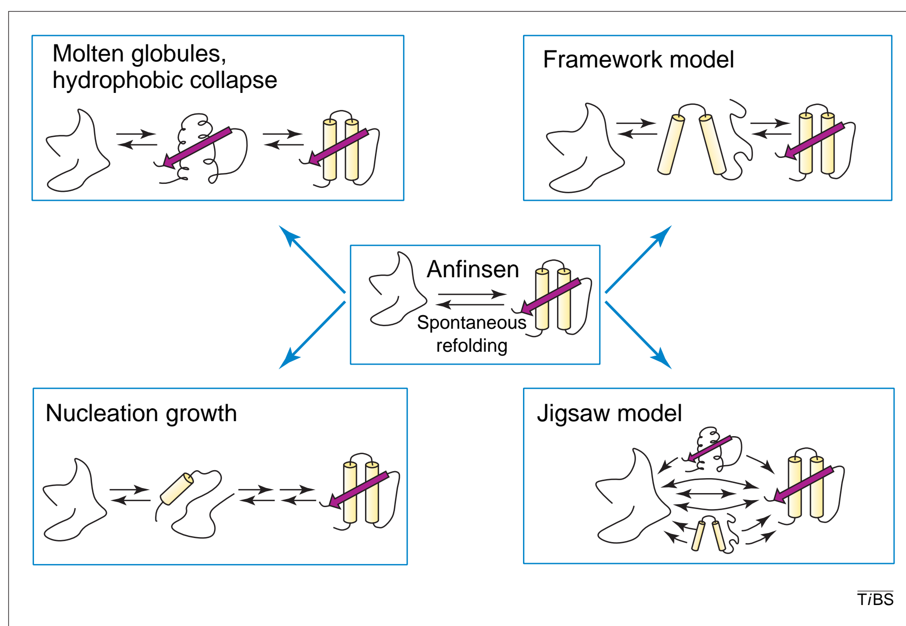

Figure 2
Early models for mechanisms of folding. Anfinsen's original experiments demonstrated that proteins fold spontaneously and reversibly into their native conformation. The nucleation growth model proposed that residues adjacent in sequence form a nucleus from which the native structure then develops in a sequential manner. By contrast, the framework model suggested that local elements of secondary structure form first and that these then dock into the native tertiary structure of the protein, possibly by a diffusion–collision mechanism. In the hydrophobic collapse model, a protein buries its hydrophobic side chains from solvent water early during folding, forming a collapsed intermediate or molten globule species, from which the native state develops by searching within this conformationally restricted state. Finally, the jigsaw model suggests that each protein molecule could fold by a different path[18]. Although it is now clear that there is not a single sequential folding route, as was implicit in some of these early models, features of these models are relevant today in the context of energy landscape models of folding.

available for study and the limitations of experimental and theoretical methods at that time. Nevertheless, it was clear that protein folding is a reversible and spontaneous event (at least for most proteins under appropriate conditions). In addition, kinetic experiments had already been performed showing that folding occurs rapidly, usually within a few seconds or so.

By contrast, simple calculations suggested that even a small protein of 100 amino acids would take over a billion years to find its native state if folding were to occur by a random search of all possible conformations[3]. These observations led to the view that there must be a specific pathway for folding, and a crucial aim was to understand how a protein finds the 'right' pathway and avoids the others. In addition, it was appreciated that metastable, partially folded states of proteins exist and, if they could be detected and studied, these species might provide important clues about the stepping stones to the native state. Using the methods available, models for folding began to emerge (Fig. 2). These early results shaped the field of folding for the next decade or so and highlighted the power of combining theoretical and experimental approaches to studies of folding that is still important today[4,5].

New views of folding

Major advances have been made in elucidating the folding mechanisms of proteins since the original models described above. These have been derived from a wealth of new and exciting experimental approaches (Table 1), combined with the development of powerful theoretical methods. The key to discerning the nature of folding mechanisms is to combine the results from different techniques so that different aspects of folding can be probed and the results combined into a common picture of the folding process[6]. Experimental methods commonly used today include fluorescence and circular dichroism, which provide information about the extent of secondary and tertiary structure formation, respectively. When combined with stopped-flow methods, they can access folding events on the millisecond timescale.

Information about folding at the level of individual residues can be obtained using hydrogen exchange[7] and protein engineering[8]. Hydrogen exchange, when combined with detection by multidimensional NMR, can be used to determine the location and stability of individual hydrogen bonds at different stages of folding (although the identity of the hydrogen bond acceptor can only be inferred). Protein engineering, by contrast, provides site-specific information about the role of individual side chains in stabilizing populated intermediates and transient high-energy transition states. In this approach, an amino acid side chain is removed from the protein of interest (e.g. by mutating Val to Ala) and the effect of the mutation on the stability of the native protein (determined by equilibrium denaturation) and the intermediate or transition state (determined using kinetics) are measured and compared[8]. The ratio of these stabilities is known as a Φ value. By determining many Φ values for residues spread throughout the native protein, the structure of intermediates (if they are populated) and the rate-limiting transition state can then be inferred.

A major developing area in folding is the establishment of techniques that can monitor the process on submillisecond timescales, so that important early events can be monitored[9] (Table 1). Such methods include ultra-rapid mixing, temperature jump and pressure jump[10,11]. When combined with detection methods such as fluorescence or circular dichroism, these can access folding events occurring on very fast (nanosecond to microsecond) timescales. In parallel to these advances, a number of theoretical methods have been developed to simulate protein folding. These include molecular dynamics that can be used to monitor the unfolding of proteins at atomic resolution (usually at very high temperatures to speed up the process) and the folding of small proteins for up to milliseconds using the computer power that is now available[12]. By contrast, lattice simulations rely on very simple models for proteins based on polymer beads but have the advantage that the conformational space can be searched exhaustively[4]. Together, these approaches have provided new insights into the folding and unfolding processes of several proteins and agree well with experimental results[8,9].

So what have we learned about folding using these methods? Some of the

most exciting developments have come from combining the results from studies of a large number of proteins[13,14] and by comparing folding across protein families[15–17], so that trends and patterns in folding mechanisms emerge and generic features of folding can be revealed. Some of the major changes in our perceptions of folding since 1976 are summarized briefly below.

The new view of folding

Perhaps the most important breakthrough in our perception of folding today is that there is not a single, specific folding pathway, as was suggested in some early models[18]. Instead, a multidimensional energy landscape or folding funnel better describes the folding process (Fig. 3). Thus, there are potentially a plethora of routes to the native state and which pathways are populated will depend on the details of the system being studied (e.g. the amino acid sequence, the topology and the experimental conditions). Different routes might be populated and/or different intermediates and transition states observed as a consequence of relatively small alterations of a common free-energy profile. Although this new view of folding presents the experimentalist with the notoriously difficult problem of monitoring rapid transitions between ensembles of molecules that are simultaneously populated, it is encouraging to note that some pathways might be much more populated than others[19]. These are the routes that the experimentalist watching 10^{18} molecules at a time (in micromolar solutions) will observe.

Folding with and without intermediates

A good starting point when studying folding is to choose simple, small proteins as model systems. The folding of a wealth of small proteins or isolated domains with less than 100 amino acids have now been studied and these have provided some of the greatest new insights into folding mechanisms[13]. The first of these proteins to be studied was chymotrypsin inhibitor 2 (CI2)[20]. This simple 64-residue protein was found to fold without populating intermediates in a so-called two-state transition. Similar behaviour has now been observed for many small, single-domain proteins[13], although exceptions do exist[15]. By contrast, larger proteins (more than 100 residues in size) usually fold in three-state transitions involving populated intermediates. These intermediates might contain structured regions corresponding

Table 1. Experimental techniques used to measure folding[a]

Technique	Timescale[b]	Structural parameter probed	Refs
Fluorescence	ns–s		10,11
(i) Intrinsic		Environment of Trp and Tyr	
(ii) ANS binding		Exposure of hydrophobic surface area	
(iii) Substrate binding		Formation of the active site	
(iv) FRET		Inter-residue distances	
(v) Anisotropy		Correlation time	
Circular dichroism	ns–s		51,52
(i) Far UV		Secondary-structure formation	
(ii) Near UV		Tertiary-structure formation	
Protein engineering	Depends on the method of detection	Role of individual residues in stabilizing intermediates and transition states	8
Small-angle X-ray scattering	⩾ms	Dimension and shape of polypeptide	53
Absorbance (near UV)	ns–s	Environment of aromatic residues or co-factors	10,11
FTIR	ns–s	Secondary-structure formation	54
NMR			55,56
(i) Real time	ms–s	Environment of individual residues	
(ii) Dynamic NMR	250 μs	Lineshape analysis provides folding–unfolding rates close to equilibrium	
Hydrogen exchange (HX)			7,57
(i) Native state	min–months	Global stability and metastable states	
(ii) Pulsed HX NMR	ms–s	Hydrogen-bond formation in specific residues	
(iii) Pulsed HX ESI MS	ms–s	Folding populations	
Force spectroscopy (AFM or ‘optical tweezers’)	s	Unfolding forces and unfolding-rate constants of single molecules	58

[a]The table was adapted from Ref. 9.
[b]The timescale available depends on the method used to trigger folding. The fastest reactions can be monitored using optical methods and temperature perturbations (ns); ultra-rapid mixing methods permit measurements following the dilution from denaturant on the μsec timescale, whereas stopped flow methods provide information on the ms timescale.
Abbreviations: AFM, atomic force microscopy; ANS, 1-anilino napthalene sulphonic acid; ESI MS, electrospray ionization mass spectrometry; FRET, fluorescence resonance energy transfer; FTIR, fourier transform infra-red.

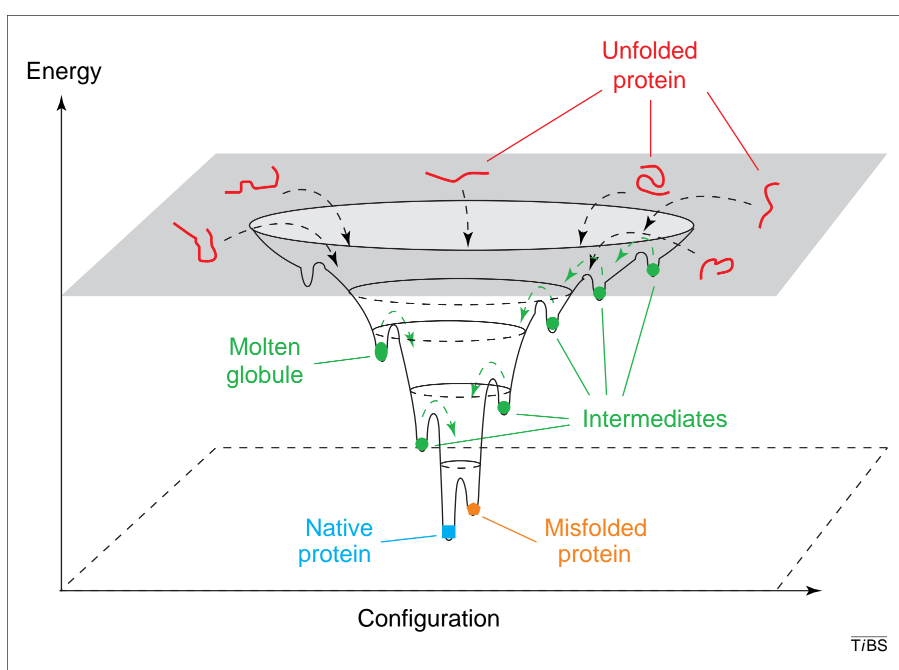

Figure 3
Schematic diagram of a folding energy landscape. Denatured molecules at the top of the funnel might fold to the native state by a myriad of different routes, some of which involve transient intermediates (local energy minima) whereas others involve significant kinetic traps (misfolded states). For proteins that fold without populating intermediates, the surface of the funnel would be smooth[48]. Reproduced, with permission, from Ref. 49.

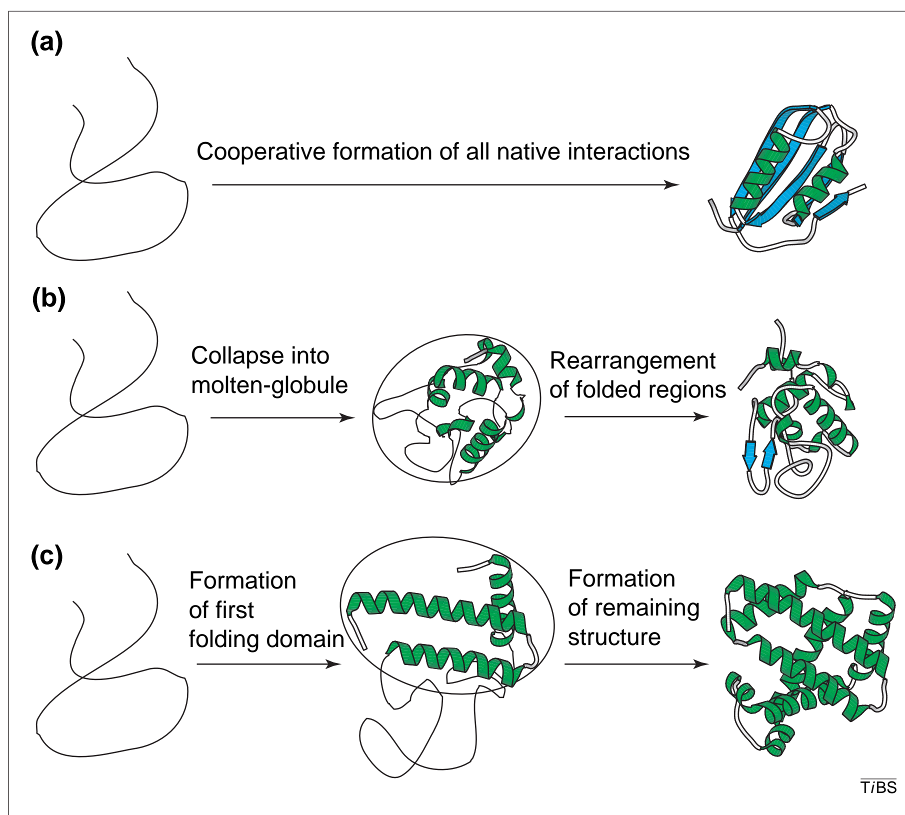

Figure 4
Small proteins fold in two-state transitions **(a)** involving the co-operative formation of all native interactions. However, proteins with more than 100 residues usually populate intermediates during folding. Some of these proteins fold through a hydrophobically collapsed (molten-globule) state **(b)**. This involves collapse into a molten globule followed by the rearrangement of the folded regions. Other proteins follow a multidomain folding pathway **(c)** that involves the formation of the first folding domain and only then the formation of the remaining structure. Reproduced with permission from Ref. 50.

to domains or subdomains in the native state (Fig. 4). For these proteins, the rate-limiting step involves the reorganization of structured elements in the collapsed state or the folding of unstructured regions in multidomain proteins.

Two-state folding suggests that stable (populated) intermediates are not a prerequisite for folding, but high-energy species (i.e. intermediates that do not accumulate to a significant extent) cannot be ruled out. Under some conditions, therefore, intermediates can be visualized in small proteins, particularly under stabilizing conditions. The role of intermediates in the folding of proteins is a controversial issue. In some cases, they appear to be important stepping stones for productive folding (i.e. they are on-pathway), whereas, in other cases, they might arise by nonspecific collapse of the polypeptide chain or accumulate because they are trapped by non-native interactions (i.e. they are off-pathway)[9]. Although off-pathway species do not teach us about the direct route between the denatured and native states, they are nonetheless important in that they form part of the energy landscape for folding and might be the substrates for molecular chaperones or precursors to aggregation (see below).

Barriers to folding and the rate-limiting transition state

A key issue in defining the mechanism of any chemical reaction is to characterize species along the reaction coordinate, including the unfolded state and intermediates, as well as the 'transition state' that defines the nature of the free-energy barrier between the reactants and products. Detailed analyses of the conformational properties of rate-limiting transition states have now been performed for several proteins using Φ analysis[8]. For small proteins such as CI2, a single rate-limiting transition state ensemble is observed that is an expanded version of the native state, containing an extended, delocalized nucleus that is stabilized by numerous partially formed interactions. Further collapse and consolidation of the structure then occurs onto the scaffold provided by the nucleus, in a mechanism termed nucleation–condensation[8].

In other small proteins, the rate-limiting transition state is more 'polarized' than that found in CI2, with some regions being almost fully formed[16,17]. Such behaviour is also found in larger proteins, which is consistent with their folding by domains[8]. In some proteins that share the same native fold, different routes to the native state through different transition states are observed[21], although the rate-limiting transition state in most proteins is usually conserved in family members (see below). This highlights the degeneracy in folding pathways (there is more than one way to navigate the funnel) and the importance of specific inter-residue interactions in shaping the energy landscape for folding. Different proteins thus appear to use different mechanisms to search the energy landscape for folding, emphasizing the need to study an array of proteins before making generalizations about folding mechanisms.

The importance of topology

Much excitement has been stirred recently by the observation that the folding rates and mechanisms of small proteins are not determined by the specific details of the amino acid sequence but by a more generic feature of the polypeptide chain – its topology (although changes in side chains will determine the stability of individual species and hence tailor the shape of the landscape[22]). This observation stems from the finding that the rate of folding of small proteins correlates amazingly well with a simple parameter, the contact order (which reflects the sequence separation of contacting residues in the native state)[14]. This correlation is important in that it suggests that simple models based on the structure of the native state should be able to describe the folding process. It also explains why lattice simulations can capture the essence of folding despite their being based on such simplified models and the observation that the structure of the rate limiting transition state ensemble is usually conserved in proteins within a structural family, even though they might have unrelated sequences[22].

Several models have recently been developed that are remarkably successful in predicting both the folding rates and the folding mechanism of real proteins[22]. Their success suggests that non-native interactions are relatively unimportant in determining folding, at least for small proteins that fold in two-state transitions, in accord with the

experimental data described above. The challenges ahead will be to improve the accuracy of these models so that the folding process can be described in detail and the folding of more complex proteins predicted. In the long term, such models might even be able to predict protein structures *ab initio*.

Protein folding in the cell

The field of protein folding has changed dramatically over the past 25 years and now encompasses topics that extend far beyond biophysical studies of the spontaneous folding process *in vitro*. Since the late 1980s, it has become clear, for example, that many proteins require assistance to fold in the cell and that this is provided by helper proteins, collectively know as molecular chaperones[23]. This is not to say that Anfinsen's original observation that folding is a spontaneous event is erroneous but that the normally spontaneous folding process needs assistance in the cell.

Why is folding *in vivo* so difficult? First, folding and biosynthesis are intimately coupled and the newly emerging polypeptide chains can fold before synthesis is complete [the rate of folding of a typical protein ($t_{1/2}$ < 1 sec) is much greater than the rate of synthesis (4–20 amino acids sec^{-1})][24]. As soon as enough amino acids protrude from the ribosome, they will fold to the lowest available energy minimum for that length of chain. For short sequences, this might involve secondary structure formation but, for longer sequences, the lack of complete sequence information will inevitably lead to incorrect tertiary structure formation. In some cases, an entire domain could fold co-translationally[24]. For a multidomain protein, however, the absence of a partner domain could expose a dangerous hydrophobic surface, especially in the highly concentrated environment of the polysome.

The difficulty of folding in the cell is not only linked to the biosynthetic process. A second danger is caused by the highly crowded environment of the cellular milieu [~300 (g protein) l^{-1}]. At these concentrations, the thermodynamic activity (effective concentration) of partially folded states is increased 10–100-fold[25]. As a consequence, the intermolecular binding constants between partially folded states are increased, leading to an increased probability of aggregation during folding[26].

The saviours of folding in the cell are the molecular chaperones. These proteins assist the folding process *in vivo* by taking part in the biosynthetic process and by capturing misfolded states in the cytosol post-translationally and allowing them renewed opportunities to fold. They are also important in protecting proteins from aggregation as a consequence of heat stress and play a role in a host of important cellular events, including signal transduction, targeting and protein degradation[27].

The first molecular chaperone identified was nucleoplasmin, which is important in mediating the assembly of nucleosomes[28]. The term has now been expanded and generalized, first by Ellis and colleagues based on their observation that a molecular chaperone is required for the biogenesis of the plant protein, ribulose bisphosphate carboxylase[23]. Today, the chaperones include ~20 protein families of different molecular weights, structures, cellular locations and precise biological roles. Since their discovery, more than 4000 papers have been published on this subject, including more than 1000 manuscripts in 1999[27]. The dramatic expansion of this field over the past 13 years not only reflects the ubiquitous nature of these proteins but also emphasizes their importance in the life and health of cells.

Over the past decade, a wealth of impressive structural and functional information about molecular chaperones has been derived from elegant and ingenious experiments both *in vitro* and *in vivo*[29,30]. The two best understood chaperones are currently the hsp70 and hsp60 families. The hsp70s have a wide range of biological activities, one of which is to bind to short hydrophobic sequences (around seven residues long) as they emerge from the ribosome. The proteins thus prevent the aggregation of newly synthesized chains during protein biosynthesis (Fig. 5). The hsp60 chaperones, by contrast, play a role later in folding, once synthesis is complete. These so-called chaperonins are built of two doughnut-shaped rings stacked back to back, the central cavity forming the important folding cage. In the case of the *Escherichia coli* chaperonin GroEL, the central folding cavity is capped by the binding of a second chaperone called GroES (Fig. 5). The hsp70 and hsp60 chaperones co-operate *in vivo*, forming a chaperone relay that sequesters the newly synthesized polypeptide until it has folded to a stage in which aggregation is no longer a threat (although the relay is thought to be more tightly coupled in eukaryotes than in prokaryotes[24,29]).

At least ~50% of proteins bind one or more hsp70 molecular chaperones during biosynthesis, and ~10% of newly synthesized proteins in *E. coli* and 15% of these proteins in eukaryotes bind to the hsp60 chaperonins for folding (Fig. 5)[24,29]. Thus, chaperonins are required for the folding of a restricted and relatively well-defined set of proteins. In *E. coli*, about half of its substrate proteins have now been identified and, interestingly, these are involved in a range of important cellular events including transcription, translation and metabolism[24]. Even more fascinating is the observation that these proteins often contain several domains with αβ folds. Large proteins such as these with a complex topology would be expected to fold slowly and to populate intermediates. It is thus not surprising that they present the cell with a particularly difficult folding task.

What do folding mechanisms derived *in vitro* tell us about the chaperone-assisted process? Most proteins that require chaperones to fold efficiently are poor folding models for biophysical experiments and, conversely, proteins suitable for *in vitro* folding studies do not usually require chaperone assistance. As a result, a direct comparison of the mechanism of spontaneous and assisted folding is not easy. Nevertheless, detailed biophysical studies of the folding of several small proteins have been performed in both the presence and the absence of GroEL[31]. Although these proteins do not represent the important 10% of proteins that require GroEL to fold, they do provide an opportunity to determine the effect of the chaperonin on the energy landscape for folding, at least for this group of proteins. These studies have revealed that GroEL can bind many types of folding intermediates, ranging from the unfolded ensemble to early and late intermediate species, and even native-like conformations. The chaperone thus has a highly non-specific mode of polypeptide binding, recognizing exposed hydrophobic surface area, which is typical of non-native states.

Sequestration of aggregation-prone non-native states from solution within the central folding cage is undoubtedly an important facet of chaperonin activity[24]. Several studies have suggested that GroEL also assists folding by an unfoldase activity, in which misfolded states are unfolded by binding to the GroEL central cavity and, as a consequence, are given renewed attempts to

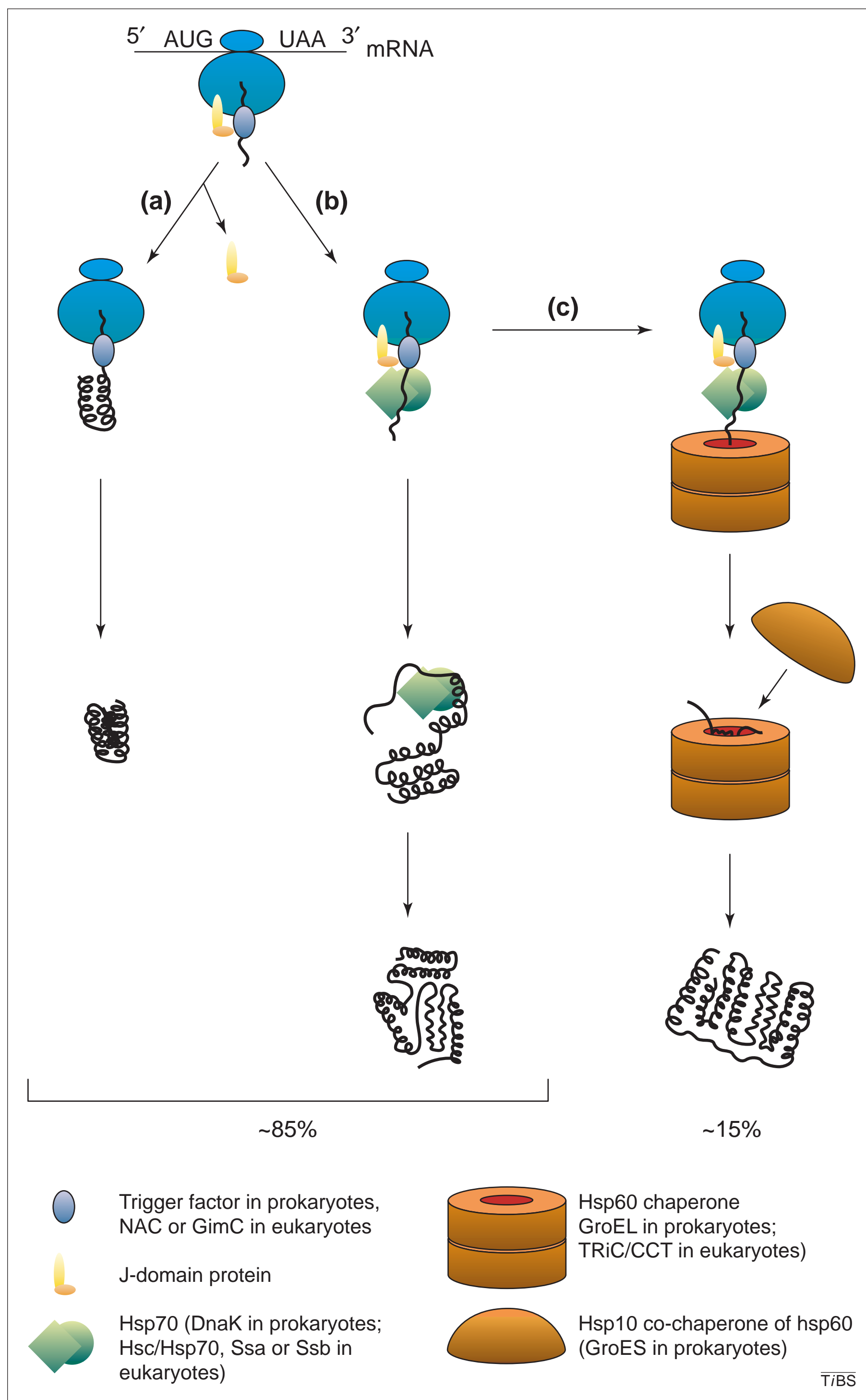

Figure 5
Schematic diagram of *de novo* folding in the cytosol of eukaryotes and prokaryotes. **(a)** Folding that does not depend on hsp70 or hsp60. **(b)** Folding assisted by hsp70 or trigger factor (TF). **(c)** Folding assisted by both hsp70 or TF and an hsp60 chaperonin. The diagram shows common features between chaperone-assisted folding in prokaryotes and eukaryotes, although there are differences. For example, TF is found in prokaryotes whereas nascent-chain-associated protein (NAC) and GimC are found in eukaryotes. In both prokaryotes and eukaryotes, a significant proportion of proteins fold without the assistance of either hsp70 or the chaperonins (a). In other proteins, folding of the polypeptide chain might involve proteins of the hsp70 class and/or TF (b). In addition, interaction with hsp60 chaperonins is required to fold complex proteins (c). Known substrates of hsp60 include actin and tubulin in eukaryotes, and an array of proteins involved in transcription, translation and metabolic enzymes in prokaryotes. Co-translational domain folding and the association of the nascent chain with chaperonins is favoured in eukaryotes. Adapted from Refs 24,29.

fold[31]. For other, larger, proteins, it has been shown that folding occurs by the same overall mechanism both assisted by GroEL and free in solution, but that the chaperonin facilitates the docking of preformed folding domains[32]. In essence, the chaperonins smooth the energy landscape for folding, removing or reducing the size of potentially dangerous traps. Such behaviour has been observed in simulations of chaperonin-assisted folding, in accord with these experimental results[33,34].

What happens when the chaperones cannot assist folding in the cell, for example, under severe stress conditions or when mutant proteins with aberrant folding or stability are produced? Given the widespread importance of molecular chaperones, it is not surprising that such conditions can lead to a number of deadly diseases. For example, mutations in the protein p53 are responsible for about half of all reported cases of cancer. Using the techniques initially devised to study the biophysics of protein folding, the majority of these sequence changes have been shown to be destabilizing[35]. A second example is found in cystic fibrosis, in which the common ΔF508 mutation in an important chloride-channel protein leads to aberrant protein folding and secretion, such that the channel is not delivered to the cell membrane at sufficiently high concentrations for normal function[36]. A third, very topical, example is the amyloid diseases, a family of about 20 known human disorders to which Alzheimer's and the prion diseases belong[37]. In these conditions, protein misfolding or aberrant processing of normally soluble proteins leads to aggregation and the formation of insoluble plaques that are characteristic of these diseases. In some cases, the wild-type protein is involved but, in others, proteins with point mutations are the causative agents. In both cases, a critical first step in the disease process is thought to involve the partial unfolding of a normally native protein. To date, there are few effective treatments for the diseases. The new methods available to study protein folding described above and the increase in our understanding of protein folding mechanisms and dynamics at the molecular level are now being applied to these diseases. The hope is that biophysical studies of protein folding and aggregation of the proteins involved in the different diseases will lead the way to effective therapies. Excitingly, the first possibilities are now beginning to emerge. For example, by designing

substrate analogues that bind and stabilize the native tetramer of the human protein transthyretin, Kelly *et al.* have produced some of the first effective agents that inhibit amyloidosis of this protein, at least *in vitro*[38]. In addition, the recent identification of the proteins involved in the processing of the Alzheimer's protein *in vivo*, together with a wealth of information learned from biophysical experiments about the mechanism of aggregation of the resulting Aβ peptide *in vitro*, have now provided the armoury needed to develop small drug inhibitors against this important human disease[39,40].

Challenges ahead

It is clear that, over the past 25 years, our concepts of folding have advanced dramatically. Nonetheless, there is still some way to go until a complete solution to the folding problem is found and major advances in the future will rely on the synergy between experimental and computational approaches. The increasing durations that can be simulated and the advancement in rapid experimental methods mean that the timescales of the two approaches should soon overlap. The next challenge will be to devise experiments that can directly probe the energy landscape for folding so that the relevance of funnel models developed by theoreticians can be directly tested for real proteins.

Although ensemble measurements will undoubtedly continue to provide important information, new innovations are sure to come from studies of the folding of single molecules, and the first steps towards this goal are now being made[41,42]. For example, new insights into the mechanical strength of single protein molecules is beginning to emerge through exciting experiments that use the atomic force microscope to unfold a protein mechanically, literally by pulling its termini apart. Not only do these experiments have an important role to play in understanding how proteins have evolved to withstand mechanical stress (e.g. when muscles contract or cells divide) but they also provide a new method of measuring the folding and binding of any protein or biomolecule at the single-molecule level.

Single-molecule fluorescence experiments on molecules diffusing freely in solution can provide new information about the behaviour of individual molecules that are often masked in ensemble measurements. For example, subpopulations of molecules in a heterogeneous sample can be individually tracked and their dynamic behaviour mapped in real time. In the context of protein folding, these experiments might be able to track rare intermediates that cannot be observed in ensemble measurements and, in the long term, might provide an experimental view of the shape of the folding energy landscape. These experiments are currently in their infancy but have already been used to measure the two-state folding of CI2 and the folding and unfolding of a variant of the yeast transcription factor GCN4 immobilized onto a glass surface[42].

Future objectives will be to develop ingenious new experiments so that protein folding and dynamics in the cell can be visualized directly and that more complex systems than the model proteins studied to date can be analysed. The amazing images of the ribosome in atomic detail[43] will undoubtedly stimulate biophysicists to apply their skills to tackle the enormous challenge of understanding protein biosynthesis and co-translational folding at the molecular level, and also remind us that the RNA world is an important area that is currently relatively poorly understood in terms of folding and function. Intriguingly, for example, it has been reported that rRNA might be a chaperone[44].

In his article 25 years ago, Robson predicted that a solution to the folding problem would involve combining experiments with molecular dynamics simulations; that there could be alternative, non-biological conformations of a protein that are more stable than the native state; and that the applications of folding would lie in structure prediction and in 'molecular microminiaturization'. These important objectives are now within reach. However, the field of protein folding today encapsulates much more than this. Important objectives include using the knowledge gained through biophysical studies of folding *in vitro* in new areas, such as making designer proteins or novel nanomaterials, and progress has already been made in these areas. For example, sequences that spontaneously fold into a desired 3D structure have been designed *de novo* and shown experimentally to be correctly folded[45], and peptides have been designed *de novo* that self-assemble under precise experimental conditions into fibrillar material reminiscent of amyloid[46]. These have huge potentials in a wealth of applications ranging from artificial skins to new scaffolds for tissue repair and tissue engineering[47].

The past 25 years have seen remarkable advances in our understanding of folding but it is clear that there are new challenges to come and much excitement ahead in developing its potential in these and other applications.

Acknowledgements

I would like to thank past and current members of my research group and my collaborators and colleagues, especially Chris Dobson and Ulrich Hartl, for many helpful discussions over the years. In a short review such as this, it is impossible to describe all of the important contributions to the folding field made over the past 25 years. I have attempted where possible to cite reviews that can provide the reader with a broad view of the field. I acknowledge with thanks Chris Dobson and Christian Schulz for providing Fig. 3 and Fig. 4, respectively. Financial support was provided by the University of Leeds, the Wellcome Trust and the BBSRC. I am a member of the Astbury Centre for Structural Molecular Biology, which is part of the North of England Structural Biology Centre (NESBiC) and is funded by the BBSRC.

References

1 Robson, B. (1976) Protein folding. *Trends Biochem. Sci.* 1, 49–50
2 Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science* 181, 223–230
3 Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys.* 65, 44–45
4 Dinner, A.R. *et al.* (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* 25, 331–306
5 Alm, E. and Baker, D. (1999) Matching theory and experiment in protein folding. *Curr. Opin. Struct. Biol.* 9, 189–196
6 Dobson, C.M. *et al.* (1994) Understanding how proteins fold – the lysozyme story so far. *Trends Biochem. Sci.* 19, 31–37
7 Englander, S.W. (2000) Protein folding intermediates and pathways studies by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* 29, 213–238
8 Fersht, A. (1998) In *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, pp. 540–614, W.H. Freeman
9 Brockwell, D.J. *et al.* (2000) Protein folding mechanisms: new methods and emerging ideas. *Curr. Opin. Struct. Biol.* 10, 16–25
10 Eaton, W.A. *et al.* (2000) Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 29, 327–359
11 Roder, H. and Shastry, M.C.R. (1999) Methods for exploring early events in protein folding. *Curr. Opin. Struct. Biol.* 9, 620–626
12 Duan, Y. and Kollman, P.A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282, 740–744
13 Jackson, S.E. (1998) How do small single-domain proteins fold? *Fold. Des.* 3, R81–R91
14 Plaxco, K.W. *et al.* (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277, 985–994
15 Ferguson, N. *et al.* (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* 286, 1597–1608
16 Grantcharova, V.P. *et al.* (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Biol.* 5, 714–720

17 Martinez, J.C. and Serrano, L. (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* 6, 1010–1016

18 Kim, P.S. and Baldwin, R.L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of folding. *Annu. Rev. Biochem.* 51, 459–489

19 Lazaridis, T. and Karplus, M. (1997) 'New view' of protein folding reconciled with the old through multiple unfolding simulations. *Science* 278, 1928–1931

20 Jackson, S.E. and Fersht, A.R. (1991) Folding of chymotrypsin inhibitor-2, 1. Evidence for a 2-state transition. *Biochemistry* 30, 10428–10435

21 Ternstrom, T. *et al.* (1999) From snapshot to movie: Φ analysis of protein folding transition states taken one step further. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14854–14859

22 Baker, D. (2000) A surprising simplicity to protein folding. *Nature* 405, 39–42

23 Ellis, R.J. (1996) Discovery of the molecular chaperones. *Cell Stress Chaperones* 1, 155–160

24 Agashe, V.R. and Hartl, F-U. (2000) Roles of molecular chaperones in cytoplasmic protein folding. *Semin. Cell Dev. Biol.* 11, 15–25

25 Minton, A.P. (1999) Implications of molecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* 10, 34–39

26 van den Berg, B. *et al.* (1999) Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* 18, 6927–6933

27 Ellis, R.J. (2000) Introduction – 10 years of molecular chaperones. *Semin. Cell Dev. Biol.* 11, 1–5

28 Laskey, R.A. *et al.* (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275, 416–420

29 Feldman, D.E. and Frydman, J. (1999) Protein folding *in vivo*: the importance of molecular chaperones. *Curr. Opin. Struct. Biol.* 10, 26–33

30 Bukau, B. and Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* 92, 351–366

31 Coyle, J.E. *et al.* (1997) Structural and mechanistic consequences of polypeptide binding by GroEL. *Fold. Des.* 2, R93–R104

32 Coyle, J.E. *et al.* (1999) GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. *Nat. Struct. Biol.* 6, 683–690

33 Chan, H.S. and Dill, K.A. (1996) A simple model of chaperonin-mediated protein folding. *Proteins Struct. Funct. Genet.* 24, 345–351

34 Todd, M.J. *et al.* (1996) Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4030–4035

35 Bullock, A.N. *et al.* (1997) Thermodynamic stability of wild-type and mutant p53 core domain. *Proc. Natl. Acad. Sci. U. S. A.* 94, 14338–14342

36 Thomas, P.J. *et al.* (1995) Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* 29, 266–270

37 Dobson, C.M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci.* 24, 329–332

38 Klabunde, T. *et al.* (2000) Rational design of potent human transthyretin amyloid disease inhibitors. *Nat. Struct. Biol.* 7, 312–321

39 Golde, T.E. *et al.* (2000) Biochemical detection of Aβ isoforms: implications for pathogenesis, diagnosis, and treatment of Alzheimer's disease. *Biochim. Biophys. Acta* 1502, 172–187

40 Drouet, B. *et al.* (2000) Molecular basis of Alzheimer's disease. *Cell. Mol. Life Sci.* 57, 705–715

41 Fisher, T.E. *et al.* (1999) The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.* 24, 379–384

42 Weiss, S. (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.* 7, 724–729

43 Ban, N. *et al.* (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–920

44 Ellis, R.J. (1997) Do molecular chaperones have to be proteins? *Biochem. Biophys. Res. Commun.* 238, 687–692

45 Dahiyat, B.I. and Mayo, S.L. (1997) *De novo* protein design: fully automated sequence selection. *Science* 278, 82–87

46 Aggeli, A. *et al.* (1997) Engineering of peptide β-sheet nanotapes. *J. Mater. Chem.* 7, 1135–1145

47 Holmes, T.C. *et al.* (2000) Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6728–6733

48 Dill, K.A. and Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4, 10–19

49 Schultz, C.P. (2000) Illuminating folding intermediates. *Nat. Struct. Biol.* 7, 7–10

50 Hamada, D. *et al.* (2000) Evidence concerning rate-limiting steps in protein folding from the effects of trifluoroethanol. *Nat. Struct. Biol.* 7, 58–61

51 Goldbeck, R.A. *et al.* (1997) Fast natural and magnetic circular dichroism spectroscopy. *Annu. Rev. Phys. Chem.* 48, 453–479

52 Akiyama, S. *et al.* (2000) Stepwise formation of α-helices during cytochrome *c* folding. *Nat. Struct. Biol.* 7, 514–520

53 Pollack, L. *et al.* (1999) Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10115–10117

54 Troullier, A. *et al.* (2000) Transient non-native secondary structures during the refolding of α-lactalbumin detected by infrared spectroscopy. *Nat. Struct. Biol.* 7, 78–86

55 Dobson, C.M. and Hore, P.J. (1998) Kinetic studies of protein folding using NMR spectroscopy. *Nat. Struct. Biol.* 5 (Suppl), 504–507

56 Burton, R.E. *et al.* (1997) The energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nat. Struct. Biol.* 4, 305–310

57 Miranker, A. *et al.* (1996) Investigation of protein folding by mass spectrometry. *FASEB J.* 10, 93–101

58 Fisher, T.E. *et al.* (1999) The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.* 24, 379–384

Pete Jeffs is a freelancer working in Paris, France.