© Birkhäuser Verlag, Basel, 1997

CMLS, Cell. mol. life sci. 53 (1997) 557–567
1420-682X/97/070557-11 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Reviews

Protein folding: concepts and perspectives

J. M. Yon

Laboratoire d'Enzymologie physico-chimique et moléculaire, Unité de Recherche Associée du Centre National de la Recherche Scientifique, Université de Paris-Sud, F-91405 Orsay (France), e-mail: jyk@epcm.u-psud.fr

Abstract. In this review, the main concepts of protein folding, as deduced from both theoretical and experimental in vitro studies, are presented. The thermodynamic aspects from Anfinsen's postulate, Levinthal's paradox to the concept of folding funnel as proposed by Wolynes and coworkers are described. Concerning the folding pathway(s), particular attention is brought to bear on the early steps that initiate the process in the light of the results of the fast and even ultrafast techniques presently being used. The role of structural domains as folding units is discussed. Last, from the recent studies, it can be concluded that the main rules deduced from the in vitro folding studies are valid for the folding of a nascent polypeptide chain in vivo.

Key words. Protein folding; folding funnel; molten globule; hydrophobic collapse.

Introduction

An important challenge in molecular biology is finding the rules that determine how a nascent polypeptide chain acquires its three-dimensional and functional structure, in other words to decipher the folding code. This second translation of the genetic message would complete the information transfer inherent in the central dogma:

DNA → RNA → Polypeptide chain $\stackrel{?}{\rightarrow}$ functional conformation

Within cells, nascent polypeptide chains are synthesized sequentially on the ribosome by a vectorial process. Although the events which lead to the formation of a polypeptide chain from the corresponding gene are well known, it remains to be understood how the monodimensional information contained in the sequence is translated into a unique or quasi-unique three-dimensional structure, within a biologically relevant time. The reversibility of the in vitro unfolding-refolding process was established from the success of Anfinsen and his group [1] in regaining up to 100% of the enzyme activity of a refolded ribonuclease which had previously been unfolded and reduced. It has been concluded that the same mechanisms are involved in refolding in vitro and in vivo. A significant corpus of information has been deduced from the increasing number of in vitro folding studies, and recent progress in instrumentation and experimental approaches has provided tremendous insights into the characteristics of the folding intermediates. The discovery of molecular chaperones has led to a reconsideration of the mechanisms of the folding process in vivo. Several questions must be addressed to understand how the information contained in the sequence is utilized in the folding process:

- Is protein folding under thermodynamic or kinetic control? How does a polypeptide chain restrict its conformational search to achieve its native structure in a short time?
- How is the folding process initiated and what is the shortest time required for folding?
- What is (arc) the pathway(s) of folding: What is the structure of the intermediates in the folding process?
- Are the main rules of protein folding deduced from the in vitro studies valid for folding in vivo?

The resolution of an increasing number of protei structures, the recent progress in instrumentation and the use of genetic engineering have stimulated both theoretical and experimental approaches, and improved our knowledge of protein folding. The spectacular developments in molecular biology, providing access to many thousands of genetic sequences, and the sequencing of the human genome, make the solution of protein folding all the more urgent. Furthermore, the refolding process is crucial for the industrial production of engineered proteins. The design of novel proteins also requires knowledge of the mechanism of protein folding. And last but not least, several pathologies are related to a wrong folding of proteins. Thus, unravelling the mechanism of protein folding represents one of the most challenging problems in biology today.

Is protein folding under thermodynamic or kinetic control?

According to Anfinsen [2], 'all the information necessary to achieve the native conformation of a protein in a given environment is contained in its amino acid sequence'. The corollary of Anfinsen's postulate is the thermodynamic control of protein folding [3]. Since the number of conformations that a polypeptide chain can adopt is astronomically large, a random search through all possible structures of the native conformation corresponding to the minimum of free energy would require

a considerable time, incompatible with the time required for folding in vitro as well as in vivo. Levinthal [4] and Wetlaufer [5] have therefore suggested that protein folding is under kinetic control. 'Levinthal's paradox' could thus be overcome, assuming the simultaneous formation of structured nuclei in several regions of a polypeptide chain through short and medium range interactions that would initiate and direct the folding process, thus restricting the number of possible conformations.

This is illustrated in terms of energy landscape by the concept of a 'folding funnel' proposed by Wolynes and coworkers [6], who describe the thermodynamic and kinetic behaviour of the transformation of an ensemble of unfolded molecules to the native state. The number of conformations remaining to be explored decreases while the exergonic process goes on (Fig. 1). As underlined by Wolynes et al. [7], 'to fold a protein navigates with remarkable ease through a complicated energy landscape'. A wide variety of folding behaviours emerge from the energy landscape depending on the energetic parameters and conditions. The authors have suggested that the folding rate is slowed by crenulations in the energy surface corresponding to local minima populated by transiently stable intermediates. Towards the bottom of the folding funnel, the number of protein conformations and chain entropy decrease. The steeper the slope, the faster the folding.

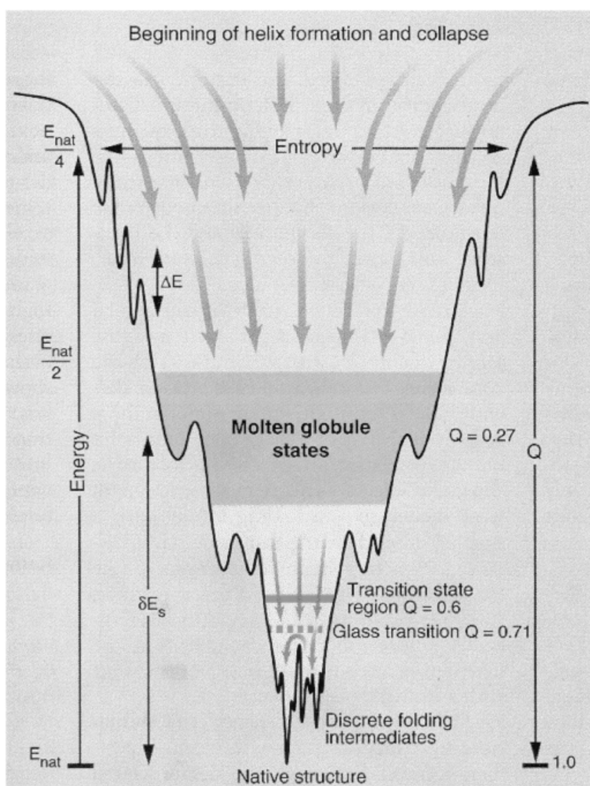

Figure 1. Scheme of the folding funnel according to Wolynes et al. [7] for a fast folding 60 residue helical protein. The width of the funnel represents entropy and depth, the energy. The fraction of native contacts Q is indicated for each collection of states. (Reprinted with permission from: Wolynes P.G. et al. (1995) Navigating the folding routes. Science **267**: 1619–1620, © 1997 American Association for the Advancement of Science).

It is now currently accepted that the three-dimensional structure of most proteins is under thermodynamic control and is reached through partially structured intermediates under kinetic control.

Pathway(s) of protein folding

Folding models. Different models arising either from theoretical considerations [8–12] or simulations [13–19] or experimental observations [20–22] have been proposed to overcome the Levinthal paradox. The classical model of nucleation-propagation, which applies to helix-coil transitions [23, 24], involves a nucleation step followed by a rapid propagation, the limiting step being the nucleation process. This model, suggested to explain the first kinetic results, accounts for the cooperativity of the folding. But the new experimental results have led this model for proteins to be discarded. The diffusion-collision model has been proposed in 1976 by Karplus and Weaver [8]. In such a model, nucleation occurs simultaneously in different parts of the polypeptide chain generating microstructures which diffuse, associate and coalesce to form substructures with a native conformation. These microstructures have a lifetime controlled by segment diffusion, so the folding of a polypeptide chain with 100 to 200 amino acids can occur in a very short time, less than a second. According to this model, the folding occurs through several diffusion-collision steps. The status of the diffusion-collision model has been re-evaluated in the light of recent experimental data [9]. A sequential and hierarchical folding by steps, in which several stretches of structure are formed and assemble at different levels following a unique route, has been proposed [20–22]. Schulz [25] has suggested a hierarchy in protein folding corresponding to the hierarchy in protein structure. In this model the first event, nucleation, is followed by the formation of secondary structures which associate to generate supersecondary structures, then domains, and eventually the active monomer; the association of domains induces the last conformational refinements which generate the functional properties. For oligomeric proteins, the association between subunits occurs in the final step of the process yielding the active protein. For most oligomeric proteins, the isolated monomer does not display any activity. A modular model of folding has been proposed by considering the three-dimensional structure of proteins. Domains have been considered as folding units by Wetlaufer [5, 26]. But it was suggested that subdomains might also refold independently, forming structural modules as folding intermediates that assemble to yield the native protein [27, 28]. The framework model [29] assumes that secondary structure is formed in an early step of folding, before tertiary struc-

ture. The hydrophobic collapse model implies that the first event of protein folding consists of a collapse which takes place before the formation of the secondary structure [13, 30]. This model originates from the work of Kauzmann [31] considering the hydrophobic effect as the driving force in protein folding and stabilization. Burial of hydrophobic residues might occur in the early steps of protein folding. Later, in the hydrophobic zipper model, Dill et al. [32] proposed that the formation of stretches of secondary structure is simultaneous with the hydrophobic collapse.

Another question is controversial. Is there a unique or several folding pathways? The existence of a unique pathway of correct folding with a sequential formation of intermediates until the emergence of the native structure has long been accepted by several authors [20–22, 33]. In 1985, the jigsaw puzzle model was introduced by Harrison and Durbin [34]. These authors have proposed the existence of multiple pathways, protein folding being considered as the assembly of a jigsaw puzzle with the existence of multiple routes to reach a unique solution. This model presents some similarities with that proposed by Karplus and Weaver. According to this hypothesis, the identification of intermediates represents a kinetic description rather than a structural description, each intermediate consisting of heterogeneous species.

Existence of intermediates in protein folding. The unfolding-refolding transition under equilibrium conditions has often been described as a two-state process. Indeed, the two-state approximation, which applies to very cooperative transitions, is valuable for several small proteins. The intermediates, generally unstable, are few populated under equilibrium conditions. The two-state approximation allows the determination of ΔG_0, the variation of free energy of unfolding. This value varies between 5 and 15 kcal/mol, indicating that the stability of proteins is rather weak.

The existence of intermediates has been shown from kinetic studies, even when the two-state approximation applies. For many proteins, monophasic unfolding kinetics and multiphasic refolding kinetics are generally observed. Much experimental evidence now exists to prove the occurrence of intermediates in the folding pathway of proteins. Kinetic studies allow the existence of intermediates and their localization in a phenomenological folding scheme to be determined. However, these transient forms are often very few populated and in rapid equilibrium with other species, making it difficult to obtain detailed structural information about these intermediates. In spite of these inherent difficulties, the recent development of techniques has allowed some of these intermediates to be characterized. Furthermore, for some proteins, transient folding states have accumulated either at low pH, or by genetic engineering, or in fragments unable to complete the folding process, making possible their characterization.

The early steps of protein folding

Detection and characterization of the intermediates are prerequisites to solving protein folding. Two major impediments to characterizing these species are the high cooperativity and the rapidity of the process. The intermediate species have generally a very short lifetime, making the analysis of their structural properties difficult. Nevertheless, several attempts have been made. Kinetic trapping of intermediates during the refolding of disulphide-bridged proteins has been used to detect intermediate species during the refolding of lysozyme [35] and BPTI [36–38]. An elegant method using differential chemical labelling has been developed by Ghélis to detect and characterize transient folding states in any protein [39], and applied successfully to elastase. One approach that is frequently used is the study of protein fragments [40–43]. Over the past decade, technological advances have improved our understanding of protein folding. Several methods are now available to study the early events in the folding process, particularly stopped-flow circular dichroism, and NMR using rapid hydrogen-deuterium exchange associated with a rapid mixing system in order to obtain a rapid pulse labelling of transient species [44–47]. Great progress has also been made through the use of protein engineering to stabilize intermediates or to probe particular regions of a protein during the folding process [48–52].

The formation of secondary structure in the early steps of folding has been shown for several proteins (for reviews, see refs. 21, 22). The formation of folded and flexible intermediates has been proposed by Ptitsyn and Rashin [53] from theoretical considerations. Ohgushi and Wada [54] have termed such a species observed during the folding of several proteins the molten globule. Ptitsyn et al. [55] have suggested that the molten globule is a general intermediate in the folding pathway of proteins. Since the literature on the molten globule is rather confusing, Goldberg and coworkers [56] introduced the term of specific molten globule. The molten globule is a rather compact intermediate with a high content of native secondary structure but a fluctuating tertiary structure. It contains accessible hydrophobic surfaces which bind a hydrophobic dye, anilino naphtalene sulphonate (ANS). In this intermediate, the aromatic residues can rotate in a symmetrical environment, as shown by the absence of near UV circular dichroism spectrum. The occurrence of a molten globule has been observed in the folding of several proteins, particularly α-lactalbumin [57], carbonic anhydrase [58, 59], β-lactamase [60], α-subunit [61] and β2-subunit of tryptophan synthase [56], bovine growth hormone [62, 63], and yeast phosphoglycerate kinase [51].

But one question must be addressed: are these secondary structures identical to those observed in the native protein? The molten globule of α-lactalbumin

has been well characterized. It is completely formed in the dead-time of a stopped-flow device. It is stable at low concentrations of guanidinium chloride, in acidic conditions and at neutral pH at low salt concentration in the absence of Ca^{++}. By NMR spectroscopy, it has been shown that this molten globule is a heterogeneous species in which the helical domain is structured having weak hydrophobic interactions, whereas the β-sheet is rather disordered. The side-chains of Tyr 103, Trp 104 and His 107 are packed into a hydrophobic cluster but the structure of this region is slightly different from that observed in the native protein [64–66]. Native secondary structures generated in less than 4 ms have been reported in the folding of cytochrome c [48] and barnase [67]. The folding of the hen egg white lysozyme involves transient steps corresponding to a reorganization of secondary structures [68]. Non-native secondary structures are formed during the refolding of β-lactoglobulin [69].

Ptitsyn [29] and Uversky and Ptitsyn [70, 71] have identified an intermediate which precedes the formation of the molten globule. This species has a significant content in fluctuating secondary structures; it is less compact than the molten globule and displays hydrophobic regions accessible to the solvent. Such a species has been termed the 'pre-molten globule' by Jeng and Englander [72]. It has been reported to occur during the cold denaturation of β-lactamase [70] and carbonic anhydrase [71]. Fink et al. [73, 74] have shown the occurrence of such an intermediate during the refolding of several proteins. An early species called the pre-molten globule, resulting from the rapid formation of nonspecific secondary structures, has been detected during the refolding of tryptophan synthase β-subunit [56].

There is a sufficient number of observations to assume that proteins fold through the rapid formation of transient intermediates, either molten or pre-molten globule, with a significant content of secondary structure and a small amount of fluctuating tertiary structure [74–76]. Since they appear in the dead-time of a stopped-flow device, it remains possible that their formation might be preceded by an earlier event.

It has been proposed that the first event in protein folding is a hydrophobic collapse occurring before the formation of secondary structures. Dill [30] has proposed that folding is initiated by a condensation of the polypeptide chain through hydrophobic interactions, followed by a rearrangement of a small number of condensed states. More recently, this author [32] has reported the hydrophobic zipper model in which the hydrophobic collapse and the formation of secondary structures are simultaneous. Several experimental data are consistent with a hydrophobic collapse in the early step of folding [51, 68, 77–84]. Residual microstructures persisting during the denaturation process have been characterized for several proteins. Neri et al. [85] have detected a hydrophobic cluster in a repressor. Logan et al. [86], using heteronuclear NMR, have reported the presence of residual local structures in FK 506 binding protein. A collapsed state has been observed in staphylococcal nuclease [79–81]. Matthews and coworkers have detected the formation of hydrophobic clusters during the refolding of dihydrofolate reductase [82]. Lumb and Kim [87] have observed intra-residue nuclear Overhauser effect in a 17-24 peptide from BPTI. Fluorescene emission spectroscopy has shown the occurrence of residual microstructures around W308 and W333 during the unfolding of phosphoglycerate kinase and its mutants with a unique tryptophan; these microstructures consist of hydrophobic clusters [51]. Thus, there are indications suggesting that protein folding originates from a hydrophobic collapse accompanied by the formation of native or non-native secondary structures. These microstructures may be involved as nucleation centres in the folding process.

There is an increasing amount of evidence showing that the initially extended polypeptide chain folds through a heterogeneous population of partially folded intermediates, the partially folded conformations being in a fluctuating equilibrium [88]. For example, hen egg lysozyme [68] and cytochrome c [83] refold according to parallel alternative pathways. Heterogeneous species have also been detected during the refolding of phosphoglycerate kinase [89, 90]; rapidly transient multimeric species are formed yielding the native monomeric protein in the slow step of folding. Multimeric species (dimers, trimers and tetramers) have been observed during the refolding of the N-terminal fragments of phosphoglycerate kinase under equilibrium conditions where neither the whole protein nor the isolated N-domain associate. Such species have not been observed for the C-terminal fragments. Transient multimeric species occur during the refolding of the whole protein; they are not in equilibrium, but they are rapidly formed and disappear during the slow folding step. Contrary to classical aggregates, their distribution does not depend on protein concentration and they are produced for protein concentrations as low as 0.05 μM. These associations are completely distributed at the end of the fast refolding step. To take into account all these observations, a model has been proposed (Fig. 2) which is formally similar to a reaction of copolymerization between two types of monomers. The multimeric species observed during the folding of the whole protein also occur during the folding of the N-terminal fragments, but they cannot transform in the absence of the interactions with the complementary fragment.

Pulse-labelling amide-exchange associated with NMR spectroscopy and mass spectometry have provided data to characterize heterogeneous species, particularly in lysozyme [91]. Such situations are consistent with the jigsaw puzzle model, at least in the early steps of the

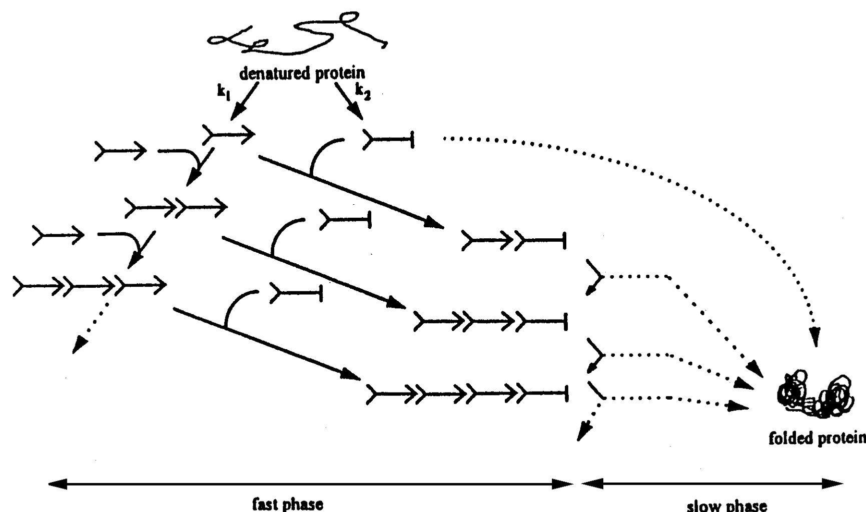

Figure 2. A model proposed to account for the formation of transient multimeric species during the refolding of phosphoglycerate kinase. The refolding of the protein produces two types of conformers; one of these can be extended either by association of the same type of monomer or by association with the other type of monomer. In this latter case, the association cannot be extended further. (Reprinted with permission from: Pecorari F. et al. (1996) Occurrence of transient multimeric species during the refolding of a monomeric protein. J. Biol. Chem. **271:** 5270–5276).

folding process. These species rapidly rearrange to generate a molten globule from which the native protein is formed in one or several sequential steps.

The kinetic methods currently used to detect the intermediates are limited to the millisecond time range. Most studies have shown the existence of extremely fast events which are complete within the dead-time of a conventional stopped-flow mixing device. The crucial problem to resolve the mechanisms of protein folding is the characterization of the very early events that initiate the folding process. Experimentally, the refolding of denatured proteins is generally initiated by rapid mixing in a regeneration buffer to dilute the denaturant. Fortunately, recent technical advances promise to improve the resolution time of the kinetic studies [88]. Submillisecond mixing techniques have been developed [92] and applied to the folding of cytochrome c [93]. Nonmixing techniques such as classical T-jump [94], nanosecond infrared laser induced T-jump [95], picosecond T-jump [96], nanosecond laser photolysis [97], optical electron transfer [98] and dynamic NMR methods [99] have been reported. From these different studies, the time of collision of sequentially distant segments of unfolded cytochrome c separated by 50 residues has been shown to be of the order of 35–40 μs, and the upper limit for the rate of protein folding has been evaluated to be around 1 μs^{-1} [100], in agreement with theoretical estimates [101]. The initial collapse of apomyoglobin into a compact state is complete in under 20 μs [95]. Using a laser T-jump method, Ballew et al. [102] have detected two very fast phases in the folding of apomyoglobin. The first phase, occurring in 250 ns, is a local collapse around Trp 14, and involves nonnative hydrophobic contacts and helix formation. In the second phase, within 5 μs tertiary contacts occur, helix A making contacts with helices H and G. Then, native apomyoglobin is generated in a final phase of 0.9 s. These methods are very promising for the investigation of the very early events which generate the folding of a polypeptide chain.

The late events in protein folding

Intermediate events in the folding time scale occur after the formation of the molten globule and before the rate limiting step of the folding process which generate the native and functional protein. In these intermediate phases, the appearance of substrate or ligand binding sites has been observed. Thus, Ca^{++} binding sites in α-lactalbumin are formed before the folded native protein [103]. Similarly, native epitopes have been found to appear in the β subunit of tryptophan synthase before completion of the native structure [104].

In the last limiting step of the folding process, the protein achieves its native conformation with the emergence of the enzyme activity. These last events correspond to the precise ordering of secondary structures [105], the correct packing of the hydrophobic core [106], subunit assembly in oligomeric proteins [29], the correct pairing of the domains in multidomain proteins [33], the reshuffling of disulphide bridges [27]. There is some evidence for the formation of incorrect disulphide bonds in the early steps of the refolding of denatured and reduced proteins. The wrong pairing must rearrange in a slow step to generate the native protein. The reshuffling of disulphide bonds can be accelerated by facilitating the thiol-disulphide interchange in suitable conditions of reagents (either traces of reducing reagents or redox mixtures) [35, 36]. Nevertheless, folding is directed by noncovalent interactions and disul-

phide bridges have only a role in stabilizing the structure. On the other hand, proline isomerization is a rate limiting step in the folding of several proteins [107]. About 60% of proteins contain at least one cisX-Pro bond in the native state. The occurrence of proline isomerization has been proposed by Brandts [107] to account for both the slow folding step and the existence of two unfolded forms in robonuclease A that have been obtained by Baldwin and coworkers [21]. Progressively, ideas about the role of proline isomerization in protein folding have evolved. Nonessential and essential proline have been reported [108]. Levitt [109] has proposed the existence of three kinds of proline, those which do not affect the folding, those which decrease the rate without preventing the folding and those which block the folding process. Thus, proline isomerization cannot be considered as specific of the late folding events.

Domains and sub-domains as folding units

Domains have been considered as folding units by Wetlaufer [5, 26]. It has also been suggested that subdomains within a protein might also refold autonomously, forming structural folded modules that assemble to generate the native protein [27]. The folding of several proteins, but not all, is consistent with the modular model. Many experimental data have shown that domains behave as independent folding units (reviewed in Ghélis and Yon [110], Wetlaufer [26], Jaenicke [33, 111]), including the N-terminal domain of γII cristallin [112], the NAD^+ binding domain of glyceraldehyde dehydrogenase [113], the phosphate binding domain of aspartate amino transferase [114], SH2 domains of p60 and p85 proteins involved in the signal transduction system [115, 116], and pancreatic elastase [117]. The engineered N- and C-terminal domains of yeast phosphoglycerate kinase have a quasi-native structure [118, 119]. The C-terminal domain still binds Mg-ATP with a stoichiometry of 1. Furthermore, these isolated domains can fold cooperatively in vitro as well as in vivo [120]. Although some local perturbations in the structure of the N-terminal domain have been detected, each domain behaves as an autonomous folding unit. Generally, isolated domains are folded faster than when integrated into the whole molecule [121].

Even though isolated structural domains are capable of folding independently, the ability of folded domains to generate an active protein has been observed only with a few proteins, such as thioredoxin [122], elastase [117], and methionyl-t-ARN synthetase [123]. Regardless of the procedure, no complementation has been observed with the domains of phosphoglycerate kinase [118]. Furthermore, circular permutations introducing a discontinuity in either one or the other domain of phosphoglycerate kinase might change the sequence of events in the folding process, but do not prevent the protein from achieving a native and functional conformation. However, it significantly decreases the stability of the enzyme (by about 4 kcal/mol), suggesting that the continuity of the domains is required for the stability but not for the correct folding of phosphoglycerate kinase [124]. The correct folding of isolated domains is not enough to yield a functional complementation of a protein. It is likely that structural patterns of proteins consisting of continuous domains have been selected during evolution on the basis of thermodynamic stability. But proteins have enough plasticity to find their functional fold, even when the continuity of the domains has been artificially disrupted. Nevertheless, continuous domains are generally considered as autonomous folding units in a protein as well as subunits in an oligomeric protein. For oligomeric proteins, subunit association occurs as a last step and induces conformational readjustments which generate the functional properties.

Do fragments exist, at a lower level, that correspond to structural units smaller than a domain and that are capable of folding autonomously? C-terminal fragments from thermolysin have been reported to fold autonomously and cooperatively when their size is equal to three helices or larger [125–127]. Oas and Kim [40] have deduced from their experiments on BPTI fragments that subdomains fold simultaneously and associate. These subdomains are proposed to be condensed states in which the close packing of the atoms in the hydrophobic core of the molecule has not yet taken place. The results obtained on the folding and complementation of fragments from barnase are also consistent with a modular model of folding [43, 128]. In contrast, this model cannot account for the folding of barley chymotrypsin inhibitor 2 [129]. Likewise, fragments from the SH2 domain of proteins p60 and p85 [115, 116], F2 fragments from staphylococcal nuclease [130–132], from tryptophan synthase, or from cytochrome c [133] smaller than a domain cannot be considered as folding units.

Seven pairs of fragments have been produced by site-directed mutagenesis followed by a chemical cleavage of phosphoglycerate kinase [89]. The cleavage positions have been chosen to correspond to limits between structural subdomains after consideration of the three-dimensional structure. Several fragments smaller than a domain have a quasi-native structure, but their unfolding transition exhibits a very low cooperativity. A discontinuity in the β-sheet perturbs the folding process. Furthermore, several pairs of fragments have been found to give functional complementation; among them, at least one of the two fragments was not significantly folded when isolated. It seems that the association of individual complementary fragments occurs in a non-native structure and subsequent conformational rearrangements through long range interactions.

The main rules of protein folding

The rules of protein folding as deduced from in vitro studies. Although understanding the transfer of information from the monodimensional to the three-dimensional structure of proteins, i.e. the folding mechanism, requires a more complete characterization of the dynamics of partially folded proteins, several rules can be deduced from the corpus of information obtained from the tremendous number of in vitro studies:

(1) The folding of a polypeptide chain is entirely determined by its sequence in a given environment (Anfinsen's postulate).

(2) For most proteins, the native structure is under thermodynamic control. It corresponds to a minimum Gibbs free energy. It is not very stable since it ranges from -5 to -15 kcal/mol.

(3) The folding pathway involves partially folded intermediates under kinetic control. According to the 'new view' of protein folding, the polypeptide chain explores the folding routes toward the native structure through intermediates consisting of heterogeneous population of partially folded species whose number decreases as the protein navigates down to the minimum of the energy landscape (Fig. 1). Some species might be trapped transiently in local minima, slowing the folding process.

The consideration of the data suggests that, initially, a hydrophobic collapse might generate a heterogeneous population with nuclei of native and non-native structures in fluctuating equilibrium, with internal rearrangements, as in a jigsaw puzzle or the diffusion-collision model. Then, correctly folded nuclei could associate to generate microdomains which diffuse and associate. Subsequently, a molten globule might be formed; even this molten globule might correspond to a population of heterogeneous species in rapid equilibrium. All these early events might occur very rapidly. From the molten globule species, the folding process might follow more restricted pathways including the formation of tertiary structure, domain pairing and the last conformational readjustments to produce the native protein in a slow folding step. For oligomeric proteins, the formation of a quasi-native but generally inactive subunit is followed by subunit association and the last conformational readjustments which generate the functional native structure. After having accepted for a long time that proteins fold according to a sequential and unique folding pathway, but considering both theoretical studies and experimental data presently available, Baldwin in 1994 [134] recognizes: 'in retrospect it seems that Harrison and Durbin may have had the right idea'. The structural characteristics of these intermediates remain a matter of discussion. The ultrafast techniques under development are very promising for the characterization of the very early events that initiate the folding process.

According to the conditions, or following definite mutations in the polypeptide chain, some of the intermediates might be trapped in minima corresponding to misfolded species which generate further aggregates. There is a kinetic competition between the correct folding and aggregation [135]. Such situations might occur in vivo as well as in vitro [136–138]. Foreign proteins overexpressed in bacterial cells frequently clumped together into insoluble inclusion bodies. Abnormal protein folding yielding typical aggregates with amyloid deposits may account for several severe pathologies, such as prion diseases including scrapie in sheep, mad cow disease, and Creutzfeld-Jakob's disease in humans; Alzheimer's disease is also characterized by the presence of amyloid fibrils in the brain tissue [139–141].

The rules of protein folding in the cellular environment and the molecular chaperones. As many previously denatured proteins, even those whose disulphide bridges have been disrupted, are able to refold spontaneously, it has long been accepted that the newly synthesized polypeptide chains fold spontaneously in vivo by the same mechanism. This view has been challenged by the discovery of molecular chaperones [142]. Therefore, this addresses the following question: are the main rules of protein folding deduced from the in vitro experiments invalidated by the discovery of molecular chaperones?

According to the postulate of Anfinsen, the information contained in the polypeptide sequence is necessary and sufficient to insure the functional conformation of a protein, which represents the structure of the lowest free energy. Neither further information nor external energy is required to achieve the correct folding. The molecular chaperones, by their transient association with nascent, stress-destabilized or translocated proteins, have a role in preventing improper folding and subsequent aggregation. But they are unable to interact with native proteins. Most data are consistent with the conclusion that they cannot refold aggregates once formed. They do not carry information capable of directing a protein to assume a structure different from the one dictated by the polypeptide amino acid sequence [143, 144]. Furthermore, they increase the yield but not the rate of folding reactions; in this respect they do not act as catalysts. Therefore, they assist the in vivo refolding without violation of Anfinsen's postulate.

The in vitro experiments indicate that chaperonins are unable to bind a totally unfolded polypeptide chain. It seems rather well established that they transiently associate with an early intermediate on the folding path. It has been suggested that this intermediate might be a molten globule or rather a pre-molten globule. However, so far this has not been clearly demonstrated. Whatever it may be, the folding intermediate has large hydrophobic regions exposed and it is likely that it associates with the chaperone through hydrophobic interactions. The hydrolysis of ATP is required in some

cases, but not all, for the release of the protein. The energy of ATP hydrolysis is used to provoke a conformational change of the chaperone allowing the release of the protein. This is well illustrated by the work of Hartl and coworkers [145] who have shown that GroES binding and release can drive GroEL-mediated protein folding in the absence of ATP hydrolysis. The authors have concluded that: 'The role of ATP hydrolysis is mainly to induce conformational changes in GroEL that results in GroES cycling at a physiologically relevant rate.' The structural data now available, joined with the functional studies have given rise to significant progress in the understanding of the mechanisms used by the chaperone machinery to assist protein folding [143, 144].

The in vitro experiments on protein folding indicate that misfolding and subsequent aggregation result from a kinetic competition between the correct folding and an off-pathway. When the formation of the correct structure is kinetically favoured in cells, molecular chaperones are not required. This is true even for oligomeric proteins. In this regard, the folding of p22 Arc repressor dimer studied by Milla and Sauer [146] offers an interesting example. Each protomer consists of a small polypeptide chain of 53 amino acids, and in the three-dimensional structure it is interwound with the other. The refolding process is described as a two-state reaction with a bimolecular rate-limiting step ($k + 10^7\ M^{-1}\ s^{-1}$), a rate comparable to the fastest protein-protein association step. Any chaperone-mediated Arc folding reaction would be slower than the unassisted folding reaction. Furthermore, in cells the overexpression of Arc does not induce the expression of heat-shock proteins such as GroEL and DnaK. In contrast, when the rate of misfolding is greater than the rate of correct folding, it is likely that the role of the molecular chaperone is to prevent the off-pathway. The basic role of the molecular chaperone is assumed to prevent the wrong interactions, maintaining the folding intermediate in the correct folding pathway, or allowing the protein to be in a conformation easily transported across membranes. The probability of incorrect interactions will be greater if the rate of biosynthesis is much slower than the rate of folding. In fact, only a small fraction of proteins requires the presence of a chaperone to fold correctly. It is likely that, in vivo, only a few proteins need the assistance of molecular chaperones to find their native structure.

Other accessory molecules are able to play the role of helpers in the folding of proteins in vivo. Protein disulphide isomerase, an abundant component of the lumen of the endoplasmic reticulum, catalyzes the formation of disulphide bonds in secretory proteins. Another enzyme, peptidyl-prolyl cis-trans isomerase, catalyzes the cis-trans isomerization of X-Pro peptide bonds. These enzymes only accelerate the process, which can also be achieved in their absence in vitro under adequate conditions (for reviews see Schmid [147] and Lorimer [148]). In other respects some proteins undergo chemical modifications such as glycosylations. It has been shown that glycosylation does not modify the folding pathway of a protein in vivo or in vitro; it only increases its stability. It has been demonstrated for several proteins [149, 150]. The proposed mechanism for this stabilization includes the role of carbohydrate moiety to solubilize the denatured state and reinforce hydrophobic patches. This stabilization appears to result mainly from entropic effects [150].

From all recent studies, it is clear that 'the principles of protein refolding established in vitro also govern the de novo folding of proteins in vivo' [143].

Acknowledgments. The author is very grateful to Drs David Perahia and Charis Ghélis for stimulating discussions.

1 Anfinsen C. B., Haber E., Sela M. and White F. H. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl Acad. Sci. USA **47:** 1309–1314

2 Anfinsen C. B. (1973) Principles that govern the folding of protein chains. Science **181:** 223–230

3 Anfinsen C. B. and Scheraga H. A. (1975) Experimental and theoretical aspects of protein folding. Adv. Prot. Chem. **29:** 205–300

4 Levinthal C. (1968) Are there pathways for protein folding? J. Chem. Phys. **65:** 44–45

5 Wetlaufer D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl Acad. Sci. USA **70:** 697–701

6 Bryngelson J. D., Onuchic J. N., Socci N. D. and Wolynes P. G. (1995) Funnels, pathways and the energy landscape of protein folding: a synthesis. Protein Struct. Funct. Genet. **21:** 167–195

7 Wolynes P. G., Onuchic J. N. and Thirumalai D. (1995) Navigating the folding routes. Science **267:** 1619–1620

8 Karplus M. and Weaver D. L. (1976) Protein folding dynamics. Nature **260:** 404–406

9 Karplus M. and Weaver D. L. (1994) Protein folding dynamics: the diffusion-collision model and experimental data. Protein Sci. **3:** 650–668

10 Karplus M. and Sali A. (1995) Theoretical studies of protein folding and unfolding. Curr. Opin. Struct. Biol. **5:** 58–73

11 Dill K. A., Bromberg S., Yue K., Fiebig K. M., Yee D. P., Thomas P. D. et al. (1995) Principles of protein folding: a perspective from simple extract models. Protein Sci. **4:** 561–602

12 Bryngelson J. D. and Wolynes P. G. (1987) Spin-glass and the statistical mechanics of protein folding. Proc. Natl Acad. Sci. USA **84:** 7524–7528

13 Levitt M. and Warshel A. (1975) Computer simulation of protein folding. Nature **253:** 693–698

14 Taketomi H., Kano F. and Gô N. (1988) The effect of amino acid substitutions on protein folding and unfolding transition studied by computer simulations. Biopolymers **27:** 527–560

15 Skolnick J. and Kolinski A. (1990) Simulations of the folding of a globular protein. Science **250:** 1121–1125

16 Godzik A., Skolnick J. and Kolinski A. (1992) Simulations of the folding pathway of triose phosphate isomerase-type α/β barrel proteins. Proc. Natl Acad. Sci. USA **89:** 2629–2633

17 Sali A., Shakhnovich E. and Karplus M. (1994) How does a protein fold? Nature **369:** 248–251

18 Kolinski A. and Skolnick J. (1994) Monte Carlo simulations of protein folding, I – Lattice model and interaction scheme. Proteins Struct. Funct. Genet. **18:** 338–352

19 Daggett V. and Levitt M. (1993) A model of the molten globule state from molecular dynamics simulations. J. Molec. Biol. **232:** 600–619
20 Baldwin R. L. (1975) Intermediates in protein folding reactions and the mechanism of protein folding. Ann. Rev. Biochem. **44:** 453–475
21 Kim P. S. and Baldwin R. L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Ann. Rev. Biochem. **51:** 459–489
22 Kim P. and Baldwin R. L. (1990) Intermediates in the folding reactions of small proteins. Ann. Rev. Biochem. **59:** 631–660
23 Zimm B. H. and Bragg J. K. (1959) Theory of the phase transition between helix and random coil polypeptide chains. J. Chem. Phys. **31:** 526–535
24 Lifson S. and Roig A. (1961) On the theory of helix-coil transition in polypeptides. J. Chem. Phys. **34:** 1963–1974
25 Schulz G. E. (1977) Structural rules for globular proteins. Angew. Chem. (Int. Ed. Engl.) **16:** 23–32
26 Wetlaufer D. B. (1981) Folding of protein fragments. Adv. Prot. Chem. **34:** 61–92
27 Levitt M. and Chothia C. (1976) Structural patterns in globular proteins. Nature **261:** 552–558
28 Chothia C. (1984) Principles that determine the structure of proteins. Ann. Rev. Biochem. **53:** 537–572
29 Ptitsyn O. B. (1991) How does protein synthesis give rise to the 3D-structure? FEBS Lett. **285:** 176–181
30 Dill K. A. (1985) Theory for the folding and stability of globular proteins. Biochemistry **24:** 1501–1509
31 Kauzmann W. (1959) Some factors in the interpretation of protein denaturation. Adv. Prot. Chem. **14:** 1–64
32 Dill K. A., Fiebig K. M. and Chan H. S. (1993) Cooperativity of protein folding kinetics. Proc. Natl Acad. Sci. USA **90:** 1942–1946
33 Jaenicke R. (1987) Folding and association of proteins. Prog. Biophys. Molec. Biol. **49:** 117–237
34 Harrison S. C. and Durbin R. (1985) Is there a single pathway for the folding of a polypeptide chain? Proc. Natl Acad. Sci. USA **82:** 4028–4030
35 Wetlaufer D. B. and Ristow S. (1973) Acquisition of the three-dimensional structure of proteins. Ann. Rev. Biochem. **42:** 135–158
36 Crieghton T. E. (1974) The single-disulfide intermediates in the refolding of reduced pancreatic trypsin inhibitor. J. Molec. Biol. **87:** 563–577; 579–602; 603–624
37 Crieghton T. E. (1974) Experimental studies of protein folding and unfolding. Prog. Biophys. Molec. Biol. **33:** 231–297
38 Weissman J. A. and Kim P. S. (1992) Reexamination of the folding of BPTI: predominance of native intermediates. Science **253:** 1386–1393
39 Ghélis C. (1980) Transient conformational states in proteins followed by differential labeling. Biophys. J. **32:** 503–514
40 Oas T. G. and Kim P. S. (1988) A peptide model of a protein intermediate. Nature **336:** 42–48
41 Sancho J. and Ferscht A. R. (1992) Dissection of an enzyme by protein engineering. J. Molec. Biol. **224:** 741–747
42 Pecorari F., Minard P., Desmadril M. and Yon J. M. (1993) Structure and functional complementation of engineered fragments from yeast phosphoglycerate kinase. Prot. Engng. **6:** 313–325
43 Kippen A. D., Sancho J. and Fersht A. R. (1994) Folding of barnase in parts. Biochemistry **33:** 3778–3786
44 Roder H., Elöve G. A. and Englander S. W. (1988) Structural characterization of folding in cytochrome c by H-exchange labelling and proton NMR. Nature **335:** 700–704
45 Baldwin R. L. and Roder H. (1991) Characterization of protein folding intermediates. Curr. Opin. Struct. Biol. **1:** 218–220
46 Baldwin R. L. (1993) Pulse H/D exchange studies of folding intermediates. Curr. Opin. Struct. Biol. **3:** 84–91
47 Dobson C. M. (1991) Characterization of protein folding intermediates. Curr. Opin. Struct. Biol. **1:** 22–27
48 Matouschek A. and Fersht A. R. (1991) Protein engineering in analysis of protein folding and stability. Meth. Enzymol. **202:** 82–112
49 Fersht A. R., Matouschek A. and Serrano L. (1992) The folding of an enzyme. 1 – Theory of protein engineering analysis of stability and pathway of protein folding. J. Molec. Biol. **224:** 771–782
50 Ballery N., Desmadril M., Minard P. and Yon J. M. (1993) Characterization of an intermediate in the folding pathway of phosphoglycerate kinase: chemical reactivity of genetically introduced cysteinyl residues during the folding process. Biochemistry **32:** 708–714
51 Garcia P., Desmadril M., Minard P. and Yon J. M. (1995) Evidence for residual structures in an unfolded form of yeast phosphoglycerate kinase. Biochemistry **34:** 397–404
52 Fersht A. R. and Serrano L. (1993) Principles of protein stability derived from protein engineering experiments. Curr. Opin. Struct. Biol. **3:** 75–83
53 Ptitsyn O. B. and Rashin A. A. (1973) Stagewise mechanism of protein folding. Dokl. Akad. Nauk SSSR **213:** 473–475
54 Ohgushi M. and Wada A. (1983) Molten globule state: a compact form of protein with mobile side-chains. FEBS Lett. **164:** 21–24
55 Ptitsyn O. B., Pain R. H., Semisotnov G. V., Zerovnik E. and Razgulyaev O. I. (1990) Evidence for a molten globule state as a general intermediate in protein folding. FEBS Lett. **26:** 20–24
56 Chaffotte A. F., Cadieux C., Guillou Y. and Goldberg M. E. (1992) A possible folding initial intermediate: the C-terminal proteolytic domain of tryptophan synthase β chain folds in less than 4 milliseconds into a condensed state with non-native-like secondary structure. Biochemistry **31:** 4303–4308
57 Kuwajima K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular protein structure. Proteins Struct. Genet. **6:** 87–103
58 Wong K. P. and Hamlin L. M. (1974) Acid denaturation of bovine carbonic anydrase B. Biochemistry **13:** 2678–2683
59 Jagannadham M. V. and Balasubramanian D. (1985) The molten globule intermediate form in the folding pathway of human carbonic anhydrase. FEBS Lett. **188:** 326–330
60 Goto Y. and Fink A. L. (1989) Conformational states of β-lactamase: molten globule state at acidic and alkaline pH with high salt. Biochemistry **28:** 945–952
61 Beasty A. M. and Matthews C. R. (1985) Characterization of an early intermediate in the folding of the α subunit of tryptophan synthase by hydrogen exchange measurement. Biochemistry **24:** 3547–3553
62 Brems D. N., Plaisted S. M., Havel H. A., Kaufmann E. W., Stodola J. D., Eaton L. C. et al. (1985) Equilibrium denaturation of pituitary and recombinant-derived bovine growth hormone. Biochemistry **24:** 7662–7668
63 Brems D. N. and Havel H. A. (1989) Folding of bovine growth hormone is consistent with the molten globule hypothesis. Proteins Struct. Funct. Genet. **5:** 93–94
64 Kuwajima K. (1996) The molten state of α-lactalbumin. FASEB J. **10:** 102–109
65 Alexandrescu A. T., Evans P. A., Pitkeathly M., Baum J. and Dobson C. M. (1993) Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: a two-dimensional NMR study. Biochemistry **32:** 1707–1718
66 Balbach J., Forge V., van Nuland N. A. J., Winder S. L., Hore P. J. and Dobson C. M. (1995) Following protein folding in real time using NMR-spectroscopy. Nature Struct. Biol. **2:** 866–870
67 Bycroft M., Matouschek A., Kellis J. T., Serrano L. and Fersht A. R. (1990) Detection and characterization of a folding intermediate in barnase by NMR. Nature **346:** 488–490
68 Radford S. E., Dobson C. M. and Evans P. A. (1992) The folding of hen lysozyme involves partially structure intermediates and multiple pathways. Nature **358:** 302–307
69 Shiraki K., Nishikawa K. and Goto Y. (1995) Trifluoroethanol induced stabilization of the α-helical structure of β-lactoglobulin: implication for non-hierarchical protein folding. J. Mol. Biol. **245:** 180–194
70 Uversky V. N. and Ptitsyn O. B. (1994) Partly folded state, a new equilibrium state of protein molecules: four-state guani-

dinium chloride-induced unfolding of β-lactamase at low temperature. Biochemistry **33:** 2782–2791
71 Uversky V. N. and Ptitsyn O. B. (1996) Further evidence on the equilibrium ''pre-molten globule state'': four state guanidinium chloride induced unfolding of carbonic anhydrase B at low temperature. J. Molec. Biol. **255:** 215–228
72 Jeng M. F. and Englander S. W. (1991) Stable submolecular folding units in a non compact form of cytochrome c. J. Molec. Biol. **221:** 215–228
73 Fink A. L., Calciano L. J., Goto Y., Kurotsu T. and Palleros D. R. (1994) Classification of acid denaturation of proteins: intermediates and unfolded states. Biochemistry **33:** 12504–12511
74 Fink A. L. (1995) Compact intermediate states in protein folding. Ann. Rev. Biophys. Biomol. Struct. **24:** 495–522
75 Dobson C. M. (1991) Characterization of protein folding intermediates. Curr. Opin. Struct. Biol. **1:** 22–27
76 Ptitsyn O. B. (1995) Molten globule and protein folding. Adv. Prot. Chem. **47:** 83–229
77 Amir D. and Haas E. (1988) Reduced bovine pancreatic trypsin inhibitor has a compact structure. Biochemistry **27:** 8889–8893
78 Amir D., Kraus S. and Haas E. (1992) Detection of local structures in reduced unfolded bovine pancreatic trypsin inhibitor. Proteins Struct. Funct. Genet. **13:** 162–173
79 Shortle D., Meeker A. K. and Freire E. (1988) Stability mutants of staphylococcal nuclease. Large compensating enthalpy-entropy changes for reversible denaturation reaction. Biochemistry **27:** 4761–4768
80 Shortle D. and Meeker A. K. (1989) Residual structure in large fragments of staphylococcal nuclease: effect of amino acid substitutions. Biochemistry **28:** 936–944
81 Shortle D. and Abeygunawardana C. (1993) NMR analysis of the residual structure in the denatured state of an unusual mutant of staphylococcal nuclease. Curr. Biol. Struct. **1:** 121–134
82 Garvey E. P., Swank J. and Matthews C. R. (1989) A hydrophobic cluster forms early in the folding of dyhydrofolate reductase. Proteins Struct. Funct. Genet. **6:** 259–266
83 Elöve G. A. and Roder H. (1991) Structure and stability of cytochrome c folding intermediates. In: Protein refolding, pp. 50–63, Georgiou (ed.), ACS Symposium series 470
84 Kosen P. A., Creighton T. E. and Blout E. R. (1981) Circular dichroism spectroscopy of bovine pancreatic trypsin inhibitor and five altered conformation states. Relationship of conformation and the refolding pathway of the trypsin inhibitor. Biochemistry **20:** 5744–5754
85 Neri D., Billeter M., Wider G. and Wuthrich K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434 repressor. Science **257:** 1559–1563
86 Logan T. M., Theriault Y. and Fesik S. W. (1994) Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. J. Molec. Biol. **236:** 637–648
87 Lumb K. J. and Kim P. S. (1994) Formation of a hydrophobic cluster in denatured bovine pancreatic trypsin inhibitor. J. Molec. Biol. **236:** 412–420
88 Plaxco K. W. and Dobson C. M. (1996) Time-relaxed biophysical methods in the study of protein folding. Curr. Opin. Struct. Biol. **6:** 630–636
89 Pecorari F., Guilbert C., Minard P., Desmadril M. and Yon J. M. (1996) Folding and functional complementation of engineered fragments from yeast phosphoglycerate kinase. Biochemistry **35:** 3465–3476
90 Pecorari F., Minard P., Desmadril M. and Yon J. M. (1996) Occurrence of transient multimeric species during the refolding of a monomeric protein. J. Biol. Chem. **271:** 5270–5276
91 Miranker A., Robinson C. V., Radford S. E. and Dobson C. M. (1996) Investigation of protein folding by mass spectrometry. FASEB J. **10:** 93–101
92 Regenfuss P., Clegg R. M., Fulwyler M. J. and Gray H. B. (1985) Mixing liquids in microseconds. Rev. Sci. Instrum. **56:** 283–290
93 Takahashi S., Yeh S. R., Das T. K., Chan C. K., Gottfried D. S. and Rousseau D. L. (1997) Folding of cytochrome c initiated by submillisecond mixing. Nature Struct. Biol. **4:** 44–50
94 Nötling B., Golbik R. and Fersht A. R. (1995) Submillisecond events in protein folding. Proc. Natl Acad. Sci. USA **92:** 10668–10672
95 Ballew R. M., Sabelko J. and Gruebele M. (1996) Direct observation of fast protein folding. Proc. Natl Acad. Sci. USA **93:** 5759–5764
96 Phillips C. M., Mizutani Y. and Hochstrasser R. M. (1995) Ultrafast thermally induced unfolding of RNase A. Proc. Natl Acad. Sci. USA **92:** 7292–7296
97 Jones C. M., Henry E. R., Hu U., Chan C. K., Luck S. D., Bhuyan A. et al. (1993) Fast events in protein folding initiated by nanosecond laser photolysis. Proc. Natl Acad. Sci. USA **90:** 11860–11864
98 Pascher T., Chesik J. P., Winkler J. R. and Gray H. B. (1996) Protein folding triggered by electron transfer. Science **271:** 1558–1560
99 Burton R. E., Huang G. S., Daugherty M. A., Fullbright P. W. and Oas T. G. (1996) Microsecond protein folding through a compact transition state. J. Molec. Biol. **263:** 311–322
100 Hagen S. J., Hofrichter J., Szabo A. and Eaton W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. Proc. Natl Acad. Sci. USA **93:** 11615–11617
101 McCammon J. A. (1996) A speed limit of protein folding. Proc. Natl Acad. Sci. USA **93:** 11426–11427
102 Ballew R. M., Sabelko J. and Gruebele M. (1996) Observation of distinct nanosecond and microsecond protein folding events. Nature Struct. Biol. **3:** 923–926
103 Kuwajima K., Mitani M. and Sugai S. (1989) Characterization of the critical state in protein folding. Effects of guanidine hydrochloride and specific calcium binding on the folding kinetics of α-lactalbumine. J. Molec. Biol. **206:** 547–561
104 Blond-Elguindi S. and Goldberg M. E. (1990) Kinetic characterization of early immunoreactive intermediates during the refolding of guanidine-unfolded Escherichia coli tryptophan synthase β_2 subunits. Biochemistry **29:** 2409–2417
105 Matouschek A., Kellis J. T., Serrano L., Bycroft M. and Fersht A. R. (1990) Transient folding intermediates characterized by proteins engineering. Nature **346:** 440–445
106 Lecomte J. T. J. and Matthews C. R. (1993) Unraveling the mechanism of protein folding: new tricks for an old problem. Prot. Engng. **6:** 1–10
107 Brandts J. F., Halvorson H. R. and Brennan M. (1975) Consideration of the possibility that the slow step in protein folding is due to cis-trans isomerism of proline residues. Biochemistry **14:** 4953–4963
108 Schmid F. X. and Baldwin R. L. (1978) Acid catalysis of the formation of the slow folding species of RNase A: evidence that the reaction is proline isomerization. Proc. Natl Acad. Sci. USA **75:** 4764–4768
109 Levitt M. (1981) Effect of proline residues in protein folding. J. Molec. Biol. **145:** 251–263
110 Ghélis C. and Yon J. M. (1982) Protein folding. Acad. Press, New York
111 Jaenicke R. (1991) Protein folding: local structures, domains, subunits and assemblies. Biochemistry **30:** 3147–3161
112 Sharma A. K., Minke-Gogl V., Gohl P., Siebendritt R., Jaenicke R. and Rudolph R. (1990) Limited proteolysis of γII-cristallin from calf eye lens. Eur. J. Biochem. **194:** 603–609
113 Jecht M., Tomschy A., Kirschner K. and Jaenicke R. (1994) Autonomous folding of the excised coenzyme-binding domain of D-glyceraldehyde 3-phosphate dehydrogenase from Thermogena maritima. Protein Sci. **3:** 411–418
114 Herold M., Leister B., Hage A., Luger K. and Kirschner K. (1991) Autonomous folding and coenzyme-binding of the excised pyridoxal 5′-phosphate domain of asparate aminotransferase from Escherichia coli. Biochemistry **30:** 3612–3620

115 Williams K. P. and Shoelson S. E. (1993) Cooperative self assembly of SH2 domain fragments restores phosphopeptide binding. Biochemistry **32:** 11279–11284
116 Shoelson S. E., Sivaraja M., Williams K. P., Hu P., Schlessinger J. and Weiss M. A. (1993) Specific phosphopeptide binding regulates a conformational change in the PI 3-kinase SH2 domain associated with enzyme activation. EMBO J. **12:** 795–802
117 Ghélis C., Tempête-Gaillourdet M. and Yon J. M. (1978) The folding of pancreatic elastase: independent domain refolding and inter-domain interaction. Biochem. Biophys. Res. Commun. **84:** 31–36
118 Minard P., Hall L., Betton J. M., Missiakas D. and Yon J. M (1989) Efficient expression and characterization of isolated structural domains of yeast phosphoglycerate kinase generated by site-directed mutagenesis. Prot. Engng. **3:** 55–60
119 Fairbrother W. J., Minard P., Hall L., Betton J. M., Missiakas D., Yon J. M. et al. (1989) Nuclear magnetic resonance studies of isolated structural domains of yeast phosphoglycerate kinase. Prot. Engng. **3:** 5–11
120 Missiakas D., Betton J. M., Minard P. and Yon J. M. (1990) Unfolding-refolding of the domains in yeast phosphoglycerate kinase: comparison with the isolated engineered domains. Biochemistry **29:** 8683–8689
121 Missiakas D., Betton J. M., Chaffotte A., Minard P. and Yon J. M. (1992) Kinetic studies of the refolding of yeast phosphoglycerate kinase: comparison with the isolated engineered domains. Protein Sci. **1:** 1485–1493
122 Slaby I. and Holmgren A. (1979) Structure and enzymatic function of thioredoxin refolded by complementation of two tryptic peptide fragments. Biochemistry **18:** 5584–5599
123 Shiba K. and Schimmel P. (1992) Functional assembly of a randomly cleaved protein. Proc. Natl Acad. Sci. USA **89:** 1880–1884
124 Ritco-Vonsovici M., Minard P., Desmadril M. and Yon J. M. (1995) Is the continuity of the domains required for the correct folding of a two-domain protein? Biochemistry **34:** 16543–16551
125 Vita C., Fontana A., Seeman J. R. and Chaiken I. M. (1979) Conformational and immunochemical analysis of the cyanogen bromide fragments of thermolysin. Biochemistry **18:** 3023–3031
126 Vita C. and Fontana A. (1982) Domain characteristics of the carboxy-terminal fragment 206-316 of thermolysin: unfolding thermodynamics. Biochemistry **21:** 5196–5202
127 Vita C., Fontana A. and Jaenicke R. (1989) Folding of thermolysin fragments: hydrodynamic properties of isolated domains and subdomains. Eur. J. Biochem. **183:** 513–518
128 Sancho J., Neira J. L. and Fersht A. R. (1992) An N-terminal fragment of barnase has residual structure to that in a refolding intermediate. J. Molec. Biol. **224:** 749–758
129 Prat Gay G. D., Ruiz Sanz J., Davis B. and Fersht A. R. (1994) The structure of the transition state of two fragments of the barley chymotrypsin inhibitor to generate a native like protein: implication for protein folding. Proc. Natl Acad. Sci. USA **91:** 10943–10946
130 Taniuchi H. and Anfinsen C. (1969) An experimental approach to the study of the folding of the staphylococcal nuclease. J. Biol. Chem. **244:** 3864–3975
131 Taniuchi H., Parker D. S. and Bonhert J. L. (1977) Study of equilibration of the system involving two alternative, enzymatically active complementing structures simultaneously formed from two overlapping fragments of staphylococcal nuclease. J. Biol. Chem. **252:** 125–140
132 Taniuchi H., Gary R. P. and Juillerat M. A. (1986) Complementation in folding and fragment exchange. Meth. Enzymol. **131:** 185–217
133 Wu L. C., Laub P. B., Elöve G. A., Carey J. and Roder H. (1993) A non covalent peptide complex as a model for an early folding intermediate of cytochrome c. Biochemistry **32:** 10271–10276
134 Baldwin R. L. (1994) Matching speed and stability. Nature **369:** 183–184
135 Yon J. M. (1996) Protein aggregation. Encyclopedia of molecular biology and molecular medicine, vol. V, pp. 73–93, Meyers R. A. (ed.), VCH, Weinheim
136 Mitraki A. and King J. (1989) Protein folding intermediates and inclusion body formation. Biotechnology **7:** 690–697
137 Yon J. M. and Betton J. M. (1991) Protein folding in vitro and in the cellular environment. Biol. Cell **71:** 17–23
138 Mitraki A., Fane B., Haase-Pettingel C., Sturtevant J. and King J. (1991) Global suppression of protein folding defects and inclusion body formation. Science **253:** 54–58
139 Prusiner S. (1996) Prions. Encyclopedia of molecular biology and molecular medicine, vol. V, pp. 57–73, Meyers R. A. (ed.), VCH, Weinheim
140 Yon J. M. (1996) Le prion. Regard sur la Biochimie. SFBBH
141 Kelly J. W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior. Curr. Opin. Struct. Biol. **6:** 11–17
142 Ellis R. J. (1991) Molecular chaperones. Ann. Rev. Biochem. **60:** 321–347
143 Hartl F. U. (1996) Molecular chaperones in cellular protein folding. Nature **381:** 571–580
144 Ellis R. J. and Hartl F. U. (1996) Protein folding in the cell: competing models and chaperonin function. FASEB J. **10:** 20–26
145 Hayer-Hartl M. K., Weber F. and Hartl F. U. (1996) Mechanism of chaperonin action: GroES binding and release can drive GroEL-mediated protein folding in the absence of ATP hydrolysis. EMBO J. **15:** 6111–6121
146 Milla M. E. and Sauer R. T. (1994) P22Arc repressor folding kinetics of a single domain dimeric protein. Biochemistry **33:** 1125–1133
147 Schmid F. X. (1990) Catalysis and assistance in protein folding. Curr. Opin. Struct. Biol. **1:** 36–41
148 Lorimer G. H. (1992) Role of accessory proteins in protein folding. Curr. Opin. Struct. Biol. **2:** 26–34
149 Wang C., Eufemi M., Turano C. and Giartosie A. (1996) Influence of the carbohydrate moiety on the stability of glycoproteins. Biochemistry **35:** 7299–7307
150 DeKoster G. T. and Robertson A. D. (1997) Thermodynamics of unfolding for Kazal-type serine protease inhibitors: entropic stabilization of ovomucoid first domain by glycosylation. Biochemistry **36:** 2323–2331