NMR studies of protein folding

Juhi Juneja and Jayant B. Udgaonkar*

National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560 065, India

NMR spectroscopy is the method of choice for determining the structural details of unfolded and partially folded states of proteins. Here, the utility of NMR spectroscopy in characterizing such disordered states which populate protein folding pathways, is discussed. The relevance of the structural information obtained to protein folding mechanisms is examined critically. NMR spectroscopy can not only be applied directly for characterizing disordered states of proteins populated at equilibrium, but can also be applied indirectly, in concert with hydrogen exchange, for characterizing equilibrium as well as kinetic intermediate states of proteins. Structural and dynamic characterization by NMR spectroscopy of protein conformations in the unfolded and intermediate state ensembles are important for elucidating the early events in protein folding, and for determining how folding is channelled along specific routes to attain the unique three-dimensional native protein structure.

ONE of the most intriguing problems in biology today is to understand the mechanism by which a newly synthesized linear polypeptide chain attains its functional three-dimensional native structure. Crucial to understanding the mechanism of protein folding is to identify and delineate the roles of partially folded (disordered) conformations that the polypeptide chain samples *en route* to the native state. While the existence of such folding intermediates, and the folding pathways defined by them, have been established for several proteins[1,2], and have been used traditionally to explain how protein folding reactions occur in biologically relevant time-scales[3], their exact roles have been very difficult to define. In part, this is because there is still a paucity of detailed structural information available on protein folding pathways, especially in the sub-millisecond time domain where intermediates are initially populated. While the sub-millisecond time domain has recently become more accessible to folding studies, with the development of new methodologies, including laser-triggered methods[4–6], pressure jump methods[7,8] and microsecond-mixing methods[9–12], the methods for structure determination that can be utilized in this or slower time domains are limited and very often indirect, for example, the protein engineering method[13,14], time-resolved mass spectrometry[15,16] or time-resolved fluorescence resonance energy transfer spectroscopy[17,18]. With the recent discovery of ultrafast folding proteins[14], the time scales of experimental studies have merged with those of molecular dynamics simulations of protein (un)folding; consequently, the need for structural characterization of protein folding pathways at atomic resolution by experimental methods has been brought even more into focus.

The necessity for critically evaluating the roles of partially folded intermediate conformations during folding has become even more pressing because two developments in the past decade have, in fact, eroded the importance ascribed previously to folding intermediates. The first has been the experimental demonstration that many proteins fold in a two-state manner without any detectable accumulation of folding intermediates[19,20]. While it is possible that intermediates are indeed populated transiently, but that they cannot be detected because either they are too unstable[21] or they follow the rate-limiting step in folding[22], it is also possible that these proteins are genuine two-state folders that do not require intermediates to fold. The second has been the development of statistical models of protein folding, which describe folding intermediates not as discrete states but as ensembles of conformations[21,23–31]. The protein reaches its minimum global free energy native state from a highly heterogenous unfolded state, and does so fast via multiple routes, where individual protein molecules may traverse their own trajectories over a multi-dimensional free energy landscape to finally converge to the native state[21,23,24,31]. In this view, intermediates are not only unnecessary for folding to occur, but may actually serve as kinetic traps that impede folding[23,24]. Clearly, whether intermediates play productive roles during folding has been called into question by these two developments.

Of course, in the cell, partially or fully unstructured forms of protein are not found only on protein folding pathways. Partially unfolded forms of proteins are important for translocation across membranes[32,33]. Many proteins are intrinsically unstructured under physiological conditions[32–35]. For example, many eukaryotic proteins or protein domains involved in signal transduction, transcriptional activation, and cell cycle regulation, fold to their biologically active conformations only after interacting with their binding partners or molecular targets[32–34]. In a recent survey of 31 genomes, four eukaryotic genomes were predicted to have 27–41% of their proteins with disordered regions larger than 50 consecutive residues in length[36], and the *Drosophila melanogaster*

*For correspondence. (e-mail: jayant@ncbs.res.in)

genome was predicted to have a staggering 17% of proteins wholly disordered[36]. Apart from their significance in normal cellular processes, partial or complete unfolding of some proteins appears to be associated with their aggregation, which leads to the development of a number of diseases[37–39]. Protein aggregation may also occur inside the cell because of a kinetic competition between correct folding and aggregation or degradation of partially folded or misfolded proteins[39,40]. Thus, characterization of the structures of disordered states of proteins has a relevance that extends well beyond the study of protein folding.

The power of Nuclear Magnetic Resonance (NMR) spectroscopy in determining the structures of folded proteins in solution is evident in there now being more than 2000 solution structures deposited in the protein database, and newly developed NMR methods have extended the molecular weight limit for solution NMR spectroscopy to > 100 kDa (refs 41–43). What has, however, remained relatively under-appreciated is that NMR spectroscopy is the only tool that can be used to obtain structural information at atomic resolution about partially structured conformations of proteins, which do not form specific three-dimensional structures, and which instead fluctuate rapidly over ensembles of conformations. For the same reason, NMR spectroscopy is useful for characterizing structurally the ensemble of unfolded conformations sampled by a protein, in order to gain insights into the nature of the (un)folding free energy landscape available to it. Applications of NMR spectroscopy range from direct or indirect structural characterization of the kinetic folding process to structural and dynamic characterization of partially structured and completely unstructured states of proteins at equilibrium. Proteins can be induced to adopt partially structured forms at equilibrium under specific conditions, usually at low pH or by mutagenesis[44], and recent developments in NMR technology[45,46] have made it the ideal method for characterizing these partially structured forms, whose relevance to direct folding pathways can be established in many cases. Many studies have brought out the immense utility of NMR spectroscopy as a tool for characterizing protein folding reactions (reviewed in 45–47). In fact, the existence and structural characterization of partially folded conformations on direct folding pathways was first shown conclusively by the use of pulsed hydrogen exchange (HX) experiments coupled to NMR[48,49] and those studies were responsible to a large extent for the revival in the worldwide activity in protein folding research that has been evident over the past fifteen years.

This review aims to analyse structural data obtained by the use of NMR spectroscopy on disordered forms of proteins, to delineate the possible roles of such partially folded forms on protein folding pathways. The review assesses what has been learnt about the relative importance of hydrophobically collapsed structures and hydrophobic cores, local (secondary) structures such as turns and helices, and long range (tertiary) structures in defining the molecular features of protein folding reactions. In particular, the ability of local structures in unfolded forms to act as potential initiation sites for folding is examined critically. The first part of the review examines the use of the hydrogen exchange technique coupled with detection by NMR, to characterize in residue-specific detail not only kinetic intermediates that accumulate during protein folding, but also partially unfolded forms populated at equilibrium. The kinetic relevance of equilibrium disordered forms to folding pathways is addressed. The second part of the review describes the direct application of NMR spectroscopy to calculate structural propensities in unfolded and partially folded forms of proteins at equilibrium. The use of NMR relaxation dynamics to obtain information about backbone motions and local interactions in disordered states of proteins is also discussed. This review does not discuss the results of real-time NMR studies and dynamic NMR studies, which have been reviewed recently in this journal[50].

Indirect NMR studies: Hydrogen exchange labelling coupled with NMR

The hydrogen exchange (HX) technique exploits the constant exchange of protein backbone amide protons with solvent protons to evaluate protein structure, stability and dynamics. Amide protons in the protein which are exposed to solvent exchange faster than those which are buried or form stable hydrogen bonds. HX can occur if the H-bond is broken transiently by local fluctuations of the protein chain, or by the global unfolding reaction[51–53]. HX rates of individual NH protons can be measured either by incubating normal (protonated) protein in D_2O, or incubating ^{2}H-labelled protein in H_2O: the proton occupancy of each amide hydrogen site at each time of incubation is measured from the intensity of the corresponding NH peak in a ^{1}H NMR spectrum where deuterium is not visible. Amide protons exchange with solvent by proton transfer reactions, and are affected by pH, temperature, neighbouring side chains and the isotopes involved. All of these effects have been thoroughly calibrated[54–56]. Under experimental conditions where the rate of the structural opening event, as well as the intrinsic rate of exchange (k_{int}) of a particular NH hydrogen (measured separately in a fully solvent-exposed peptide model) is slower than the rate of structural closing, the free energy of the opening reaction can be obtained as, $\Delta G_{op} = -RT \ln k_{ex}/k_{int}$, where k_{ex} is its observed rate of exchange[51]. The ratio of k_{int} to k_{ex} for an amide hydrogen yields its protection factor, which reflects the factor by which exchange is slowed down because of protection

from exchange with solvent protons that is imparted to the amide hydrogen site by the protein structure (hydrogen bonding and/or burial in the protein core).

The HX technique is extremely sensitive because it can access partially structured kinetic intermediates that are very short-lived, as well as intermediates that exist in small quantities as high-energy excited states in equilibrium with the native state[57–59]. HX combined with NMR detection can therefore provide site-specific structural information about protein (un)folding pathways and conformational dynamics. In most cases, the kinetics of attaining protection from HX can be correlated with the folding phases measured by traditional optical methods.

Pulse-labelling by HX

The most detailed structural models of transiently populated kinetic intermediates during folding come from experiments using the technique of pulse-labelling by HX[57–61]. In a typical experiment, the unfolded protein deuterated at all backbone amide sites is diluted into D_2O refolding buffer. At different times after initiating the refolding reaction, a short exchange 'pulse' of protons at alkaline pH is given. Amide sites which are still unfolded or exposed to solvent, at the time of application of the pulse, exchange rapidly with the solvent protons, while those that have become protected by structure formation remain deuterated. The protein is then allowed to refold completely by jumping the pH down to a low value. At low pH, exchange is slow relative to folding, and so the amide sites do not undergo further HX. Thus, the structure of the protein at the time of the HX pulse is imprinted in the degree of labelling of the individual amide sites, and can be detected by a ^{1}H 2D COSY or TOCSY NMR spectrum of the folded protein. The rates at which individual amide sites get protected from exchange (or fold) can then be determined by following changes in their NH-C$^{\alpha}$H cross peak intensities in the NMR spectra, thereby providing insight into the residue-specific structural changes that occur during folding of the protein. A major limitation of the pulse-labelling technique is that only those protons which are strongly protected in the native state can be used to probe the structure of a folding intermediate; hence, the method is biased in favour of detecting native-like structure. Additionally, non-native interactions formed during folding, for example an amide proton hydrogen bonding to a non-native acceptor, are not likely to be detected, but information about non-native acceptors may be inferred from the patterns of protection of the amides[62].

The capability of the pulsed quench-flow HX technique was first demonstrated on ribonuclease A (RNase A)[48] and cytochrome *c* (cyt *c*)[49]. A kinetic intermediate with many protected amides in the native **β**-sheet structure was found to be populated early during the folding of RNase A[48]. Since the kinetics of refolding of the slow-folding species of RNase A monitored by tyrosine absorbance were considerably slower than the folding rates measured for individual NH protons in the pulsed HX experiment, it was inferred that the formation of stable secondary structure precedes the formation of tertiary interactions, providing support for the framework model of protein folding. In later experiments[63,64], the stabilities of folding intermediates of RNase A were probed by varying the pH of the exchange pulse[63] in the range 7.6 to 11, so that k_{int} varied over at least a 10^4-fold range[55]. The probe protons forming a **β**-sheet in the native protein showed unexpectedly large protection early during refolding which increased with the time of folding. This indicated that the **β**-sheet is already formed but only marginally stable in the early folding intermediate, and is stabilized later during folding. These studies also showed that the early native-like intermediate is not fully populated at the time it is fully formed, which can be attributed to the formation of a competing off-pathway intermediate or to the existence of a competing folding pathway. In similar studies with cyt *c*, several amide protons in the N- and C-terminal helices were seen to become protected within the 4 ms dead time of the quench-flow apparatus used to carry out the experiments, suggesting that these helices are the first to fold[49].

Pulsed HX experiments have shown that the folding of hen lysozyme occurs via multiple pathways[65]. Analysis of the kinetics of protection from HX of backbone amides indicated that assembly of the native helical segments occurs faster than the formation of **β**-strands during folding of the protein. Since the HX kinetics were found to be independent of the pulse pH for most of the probe protons, it could be inferred that the biphasic kinetics in these labelling experiments arose because of parallel folding routes, and were not due to the partial protection in a marginally stable folding intermediate on a sequential pathway. Refolding of lysozyme measured by far-UV CD showed that native-like secondary structure is acquired within 2 ms; yet no protection from HX was apparent, suggesting that the secondary structural elements formed are extremely labile at this stage of folding[65]. Similarly, in interleukin-1**β**, 90% of the native **β**-sheet structure forms within 25 ms of refolding monitored by CD, but pulse-labelling experiments show that the backbone amide hydrogens are protected only after ~1 s, indicating that the final packing of side chains in the **β**-sheets is necessary to make these H-bonds strong enough to provide protection[66].

Recently, the technique of pulse-labelling by HX was used for the first time to study the unfolding reaction of a protein (Figure 1 *a*). This study demonstrated that competing pathways for unfolding exist for RNase A[67]. Proton occupancies at a few amide hydrogen sites were seen to increase in a biphasic manner with time of unfolding, with substantial (35–60%) 'burst-phase' labelling (that

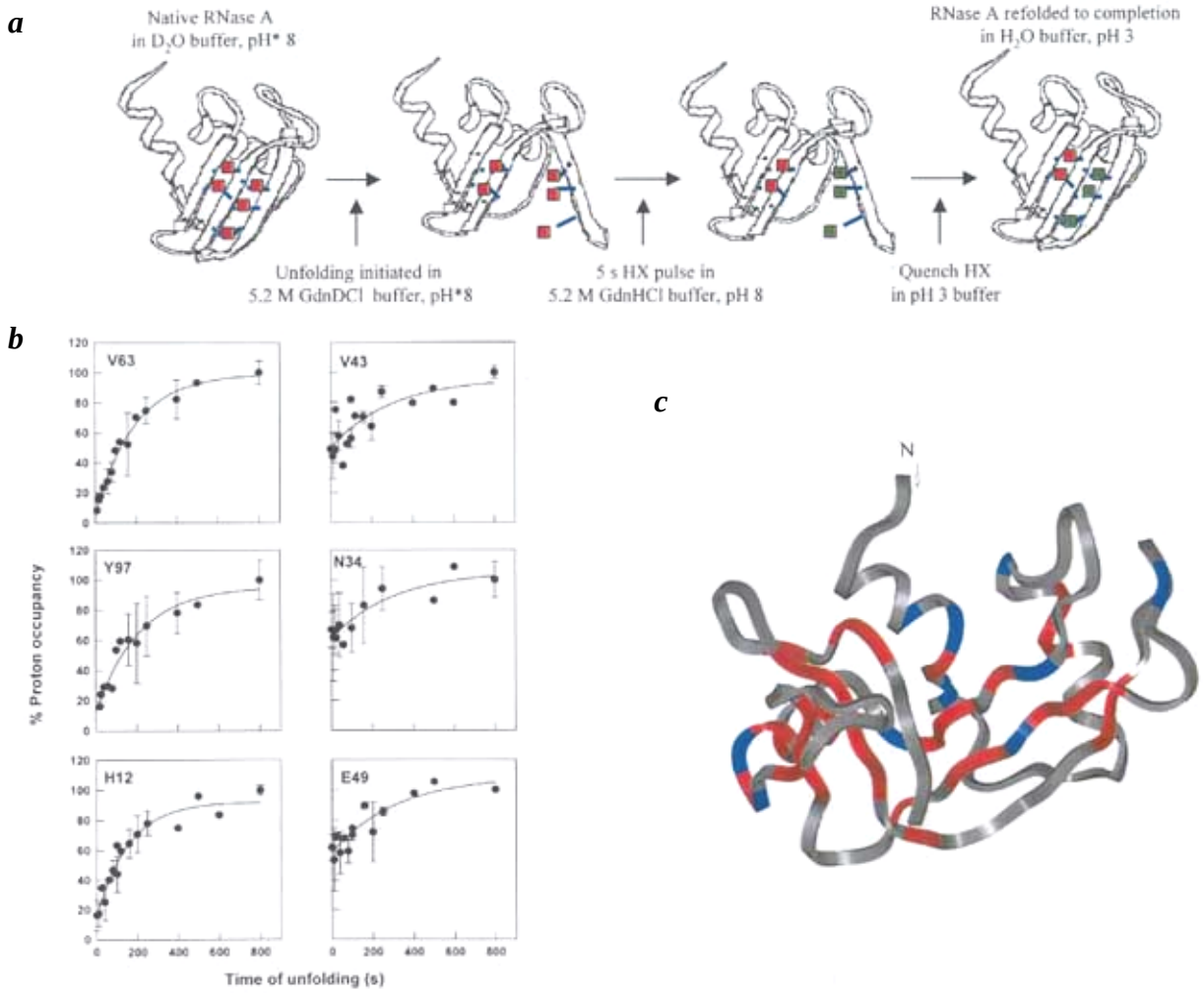

Figure 1. ***a***, Schematic representation of the experimental protocol used for characterizing the unfolding reaction of RNase A by pulsed HX experiments; ***b***, Kinetics of exposure of backbone amide deuterons of RNase A. (•), The accessibility to exchange with solvent H_2O protons vs time after initiation of unfolding in 5.2 M GdnDCl at pH* 8 (not corrected for isotope effects), 10°C is plotted for a few residues. The internally normalized intensity of each $C^{\alpha}H$-NH cross-peak in the TOCSY spectrum of the longest unfolding time-point (800 s) is taken as 100% proton occupancy of that amide proton site. All other intensities are normalized to this. The solid line in each panel is a single-exponential fit of proton occupancy vs the unfolding time. The mean rate of exposure of 43 probe protons is 0.005 s^{-1} and the standard deviation is 0.001 s^{-1}. ***c***, Location of amide proton probes used to study the unfolding pathway of RNase A. The residues marked in red belong to class I and the residues marked in blue belong to class II, and show 9 ± 4% labelling and 47 ± 9% labelling at $t = 0$ unfolding time, respectively. The RNase A figure was drawn using MOLSCRIPT from the protein NMR structure 2AAS submitted by Santoro *et al.* in the PDB[191]. The figure has been reproduced with modification from ref. 67.

occurs in the dead-time of measurement) occurring immediately upon initiation of unfolding, as shown in Figure 1 *b*, *c*. This suggested that a fraction of native molecules is transformed rapidly to a partially unfolded intermediate ensemble, as soon as the protein is exposed to strong denaturing conditions. The remaining molecules unfold via a parallel pathway on which an exchange-incompetent intermediate is formed. On this pathway of unfolding of RNase A, the secondary structure breaks down before all the tertiary structure[67]. All the probed NH sites show, however, similar observable rates of exposure to exchange with solvent protons, which are also the same as the rate of dissolution of all tertiary packing interactions in the protein, measured by near-UV CD and fluorescence. Thus, it appears that even though most of the α-helical bonds in RNase A are disrupted as indicated by the loss of far-UV CD, the amide hydrogen sites remain inaccessible to solvent until all tertiary interactions are lost and water can penetrate into the protein core[67].

Competition labelling by HX

The pH competition labelling method is useful for detecting marginally stable 'burst-phase' or early folding intermediates, and provides information about the regions of a protein that become structured in the initial steps of folding. Such intermediates in which secondary structure is not yet stabilized might be missed in a pulse-labelling HX experiment.

In a competition experiment, folding is made to occur in competition with HX. In a typical competition experiment, a denaturant-unfolded protein in H_2O is diluted into D_2O buffer to initiate refolding in a stopped-flow apparatus with a mixing dead time of ~1 ms. The pH of the refolding/exchange buffer is varied to probe the stability of individual H-bonds. At low pH, HX is slow[55], and the protein folds before none or only a few of its amide protons can exchange. As the pH is increased, HX is accelerated due to base catalysis (by OD^-), and competes with the folding reaction. This competition between folding and HX is maintained for a short time before exchange is quenched by lowering the pD, and the protein is allowed to refold completely. The amount of H-label remaining at each amide site at the time of the quench is measured in the refolded native protein from the NMR peak intensities. The pH competition labelling method is limited by the fact that folding pathways of proteins, and stabilities of the intermediate states, may vary with pH. It can be used to study only those proteins that are stable over a relatively wide range of alkaline pH. Nevertheless, it provides valuable information about structure formation in the initial few ms of refolding.

The pH competition labelling method has been used to monitor the formation of very early hydrogen-bonded and solvent-excluded structure in cyt *c*[68]. The pH profiles of some amides were found to be shifted toward basic pD compared to the pH dependence of exchange labelling in a random coil[55], indicating significant protection. The protection factors calculated for individual NH protons, showed that all α-helices in cyt *c* contain enough structure to slow exchange by 3 to 8-fold after 2 ms of refolding. Similarly, the existence of native-like secondary structure in the α-domain of lysozyme could be detected at 3.5 ms of refolding[69]. The amides that become protected at this early stage of folding are also the ones that show higher levels of protection in pulse-labelling HX experiments (discussed above)[65]. In contrast, no significantly protected structure could be detected in the first few milliseconds of refolding of ubiquitin, before the occurrence of its major cooperative folding event[69].

Do sub-millisecond folding products have specific structure?

An issue of much topical interest in protein folding is whether the rapid (submillisecond) changes in optical signals, which are seen upon denaturant dilution for many proteins, do indeed represent the formation of a very early burst-phase intermediate with specific structure. While this has been the traditional interpretation[70–72], it has been argued recently that such submillisecond changes might instead represent merely a fast readjustment of the unfolded protein ensemble at high denaturant to a more compact unfolded protein ensemble at low denaturant[73,74], and that the fast readjustment manifests itself as a non-specific collapse of all protein molecules. One experimental observation that has been used in support of this alternative interpretation of burst-phase changes is that the products of these changes for several proteins display very low protection factors (< 50) from HX when probed by competition labelling or pulse labelling experiments[69,74–76]. Whether such low protection factors do indeed indicate an absence of specific structure is questionable because they might be the result of structure that is specific yet labile[62,68,69,77]. It would be more pertinent to examine whether or not the pattern of labelling in the initially formed species is similar to that in the corresponding fully folded protein. For example, in a study of the refolding of β-lactoglobulin, which combined ultra-rapid mixing techniques with different HX labelling protocols, the pattern of HX protection showed that a kinetic folding intermediate, which is populated within 2 ms, contains marginally stable non-native α-helices near the N-terminus[62].

Do kinetic intermediates populate direct protein folding pathways?

A vexing, perennial issue in any protein folding study is to resolve whether a detected intermediate is an on-pathway productive conformation, or an off-pathway misfolded conformation that is trapped, and from which the protein must unfold before refolding to the native state[59]. The progressive formation and stabilization of structure observed during the folding of many proteins, including the examples mentioned above, provides considerable evidence in support of observed kinetic intermediates populating a direct folding pathway. One of the first direct demonstrations of a kinetic intermediate populating a direct folding pathway came from a modified pulse-chase HX experiment coupled with NMR detection, which was designed specifically to check whether the native-like kinetic intermediate, I_N (ref. 78), formed during the folding of RNase A was on- or off-pathway[79]. In this experiment, HX was allowed to occur for varying lengths of time in D_2O after initiating refolding of protonated protein in H_2O. The exchange pH was chosen such that the unfolded protein (U) rapidly exchanges with solvent deuterons while I_N does not. It was observed that the ^1H label in I_N was retained in N indicating that I_N is on-pathway[79], because if I_N were off-pathway it would have had to unfold to U before forming the native state (N), and consequently, the ^1H label would not have been retained in N.

Do equilibrium intermediates resemble kinetic intermediates?

Partly folded conformations of proteins are usually not populated at equilibrium because such conformations are

unstable relative to the native state. Many proteins can, however, be induced into partly unfolded conformations by manipulations which stabilize the partly folded conformation relative to the native state. These manipulations include for example mutagenesis[44,80,81] or more commonly, transfer of the protein to marginally stabilizing conditions such as low pH where the native state usually converts into an equilibrium molten globule form (reviewed in ref. 44). Many studies utilizing techniques such as circular dichroism (CD)[12,82,83], fluorescence[83,84], small-angle X-ray scattering[85,86], thiol–disulfide exchange[87], or real-time NMR[88], have shown that equilibrium molten globule forms are closely related to kinetic intermediates. For example, by use of a combination of CD, fluorescence and dynamic light scattering measurements, it has been shown that a molten globule-like intermediate of barstar, which is formed at high pH in the presence of high salt concentrations, is a productive intermediate on the direct folding pathway[89].

Amide exchange trapping experiments provide residue-specific structural information on equilibrium molten globule forms. In an amide exchange trapping experiment, NH protons in the partially folded intermediate populated at equilibrium are allowed to exchange with the solvent deuterons over a period of minutes to hours, before HX is quenched by returning the protein to native conditions. NMR spectra are recorded to identify protons trapped in stable secondary structures in the intermediate. Protection factors calculated for protons in molten globule-like intermediates of proteins are seen to be 10^4 to 10^5-fold smaller compared to those in the native state, reflecting the inherent flexibility of these non-native states[60]. Like the pulse-labelling by HX method, information from the amide-exchange trapping experiments is limited to regions of the protein which are strongly protected or stable in the native state.

The results of several NMR-based experiments have shown that for many proteins, partially folded states populated at equilibrium, such as molten globule forms, possess persistent native-like structures[59,60]. The kinetic relevance of partially folded, equilibrium conformations of proteins was clearly established when it was shown that the structures of such conformations, determined by amide-exchange trapping experiments[44,59,60] resemble, at the individual residue level, the structures of the corresponding kinetic intermediates determined by HX pulse-labelling experiments. Specific structural similarities between an equilibrium molten globule and an early folding intermediate have been observed in HX studies of several proteins, notably, apomyoglobin (apoMb)[82,90], RNAse H[91,92] and α-lactalbumin (α-LA)[83,93,94]. In the case of apoMb, recent HX pulse-labelling studies[95] have also been used to show that structural heterogeneity exists in the initial kinetic intermediate, suggesting that, like other proteins[65,70,77,96,97], apoMb folds via multiple pathways.

Native-state hydrogen exchange

During the past few years, native-state HX experiments have shown that even under strongly native-like conditions, the fully folded native protein continuously samples partially as well as completely unfolded conformations of the protein, which are present as high energy states[59,60] and are populated according to Boltzmann distributions. Although these high energy states would not be normally observable, they are easily detected by their accessibility to HX which is monitored by NMR. In a native-state HX experiment, the populations of these conformations are modulated by an environmental perturbation such as addition of sub-denaturing amounts of denaturant. The stabilities of individual residues are estimated from their HX rates ($\Delta G_{op} = -RT \ln k_{ex}/k_{int}$), and the sensitivity of these rates to denaturant[59].

Native-state HX experiments of oxidized cyt *c* in sub-denaturing concentrations of guanidine hydrochloride, identified the globally unfolded form and three partially unfolded forms (PUFs), all in equilibrium with the native state, on the basis of different stabilities in distinct regions of the protein[98]. The N- and C-terminal helices comprise the most stable region of the protein, and the molecule must first globally unfold for exchange with solvent at these helices to occur: ΔG_{op} for amide hydrogens belonging to these helices equals the free energy of global unfolding at all concentrations of denaturant used. Amide hydrogens belonging to the other three distinct regions show ΔG_{op} values lower than that corresponding to global unfolding. The data suggested a hierarchical folding pathway for cyt *c*, in which the most stable region folds first and the least stable region folds last. Kinetic refolding experiments of cyt *c* seem to reflect this equilibrium folding pathway. Thus, the N- and C-terminal helices are the first region to fold in pulsed HX kinetic refolding experiments[46], and misligation of the heme group in cyt *c* appears to slow the addition of the 60s helix to the folded N and C helices[99]. HX rates measured for reduced cyt *c* in relatively strong denaturing conditions, near the beginning of the equilibrium unfolding transition, also indicate several structure-opening equilibria[100]. These represent multiple partially unfolded forms, including at least one PUF, comprising the N- and C-terminal helices, which resembles the most stable PUF seen previously for the oxidized state of cyt *c*[98,101].

Native-state HX studies of RNase H demonstrate that two partially folded states co-exist at equilibrium with the native state[102]. Moreover, the first regions to fold in kinetic refolding experiments of RNase H[91] are the thermodynamically most stable regions of the native protein, as also seen previously for cyt *c*[49,98]. For several other proteins as well, the most stable region (with the most slowly exchanging amide protons) is among the earliest to acquire persistent hydrogen bonds during folding[62,103–105].

A partially unfolded state was observed to be in slow equilibrium with the predominant native state of RNase A under native conditions in the absence of denaturant. This state was also shown to populate one of the unfolding pathways of the protein[67]. An earlier native-state HX study in sub-denaturing concentrations of guanidine hydrochloride had indicated that native RNase A equilibrates with partially unfolded intermediates, but it could not be determined if these were on-pathway[106].

The correlation between the stabilities of different regions of a protein and its (un)folding pathway cannot, however, be generalized for all proteins. For example, native-state HX studies of T4 lysozyme showed that the average stability of the C-terminal subdomain of the protein is significantly higher than that of the N-terminal subdomain, but the observation of a continuum of unfolding energies for amide protons in the two domains made it impossible to identify discrete partially unfolded forms of the protein[107]. The refolding kinetics of T4 lysozyme studied by pulse-labelling by HX do not correlate with the equilibrium data, and indicate instead the formation of an early intermediate with strong protection from exchange in regions belonging to both subdomains of the protein[108].

Native-state HX was the first technique that provided information about the free energy landscape of a protein, because it allowed partially unfolded conformations in equilibrium with the native state to be detected directly. Native-state HX experiments can identify regions of the protein with global and sub-global stabilities[98,102,107]. They show, for example, that the PUFs in the mesophilic and thermophilic homologues of RNase H are similar, and that the free energies of unfolding for all regions of the thermophilic protein are higher, except for the unfolding free energy of one helix which constitutes a distinct PUF in the thermophilic protein[109].

Partially unfolded conformations are populated at equilibrium because of thermal fluctuations of the protein under native conditions, and constitute the next higher rung after the native state on the free energy ladder[23]. Native-state HX measurements are equilibrium studies that only provide information about the relative free energies of equilibrium conformations available to the protein for sampling, as well as information about fluctuations of the protein backbone[110]. They cannot be extrapolated to determine the order of events occurring during folding of the protein. Nevertheless, for some proteins, it appears that the structures of these equilibrium conformations match the structures of kinetic intermediates observed independently by HX pulse-labelling[59,60].

Direct NMR studies of disordered states of proteins at equilibrium

For those proteins for which equilibrium intermediate states have been shown to resemble transient kinetic intermediates, a direct NMR analysis of the structures of these partially folded states is particularly useful. Direct NMR studies of unfolded states of proteins provide information about the nature of the ensemble of conformations at the starting point of folding. NMR lines in unfolded state spectra are quite sharp due to the greater backbone mobility, so that sequential assignments are sometimes more easily made than for a folded protein of comparable molecular weight[45,46]. Partially folded states of proteins are more difficult to study directly by NMR, because, while their spectra typically show chemical shift degeneracies characteristic of unfolded proteins, the resonances are as broad or even broader than those of folded proteins[45,46]. Moreover, these states are rapidly interconverting between multiple conformations. Thus, their NMR spectra display small chemical shift dispersion and broadened resonance lines leading to spectral overlap which makes resonance assignment difficult[45,46]. The advent of high-field NMR spectrometers and the development of new multi-dimensional heteronuclear NMR experiments have, however, made possible the assignment and structural characterization of disordered states of proteins[45,46,111], and even large soluble protein aggregates[111,112].

Conformational preferences from NMR parameters

Conformational propensities of the polypeptide chain can be derived from chemical shifts, Nuclear Overhauser Effects (NOEs) and coupling constants. The deviations of chemical shifts from random coil values (termed secondary shifts), particularly for $^{13}C^{\alpha}$, ^{13}CO and $^{1}H^{\alpha}$, are extremely sensitive probes of secondary structural propensities of the protein backbone, as revealed by statistical studies of folded proteins[113]. Rapid conformational averaging between the α- and β-regions of partly folded and unfolded states is expected to reduce the secondary shifts below the extreme values found in folded states of proteins. The observed secondary shift can therefore be used as a measure of the relative propensity to populate the α- or β-region of dihedral space. Main chain coupling constants between $^{1}H^{N}$ and $^{1}H^{\alpha}$ also provide information about the conformational preferences of the backbone because they are sensitive to the Φ dihedral angle[114]. In disordered states, however, these alone may be less informative because conformational fluctuations average the coupling constants, making them insensitive to local structural preferences[115]. The relative intensities of the observed sequential NOEs, reflect the propensity of the backbone to preferentially populate the α- or β-region of (Φ, ψ) space, or to exist in a random coil conformation[115]. The existence of medium range NOEs reflects the presence of folded elements of secondary structure, and the secondary structure conformation can be identified from the pattern of the NOEs (ref. 115).

Backbone dynamics

The movement of the NH bond axis provides information about flexibility of the backbone. ^{15}N spin relaxation measurements in a polypeptide chain can be used to probe backbone dynamics in partially folded and unfolded states of proteins. The ^{15}N spin relaxation is related to three parameters: the longitudinal relaxation rate (R_1), transverse relaxation rate (R_2), and the steady-state NOE enhancement[116]. All these parameters are sensitive to motions in the picosecond to nanosecond time scale, and the $\{^1H\}$-^{15}N heteronuclear NOE is most sensitive to higher frequency motions of the backbone[117]. The R_2 relaxation rate is more sensitive to nanosecond time scale motions, and also reflects contributions from slower microsecond to millisecond processes such as those arising from chemical exchange which may cause line broadening in the NMR spectrum[117]. Regions showing motional restrictions due to structure formation in disordered states of proteins can be identified from the analysis of the ^{15}N relaxation measurements. Analysis of the relaxation data for disordered states of proteins is usually done using the reduced spectral density method[118], and not the model-free approach used for folded proteins[119]. The latter method assumes that the molecule tumbles isotropically with a single correlation time, and is not valid for unfolded or partially folded states which have a wide variation in local correlation times between residues in the more structured regions and those in less structured and unstructured regions of the protein.

Lessons for protein folding

Unfolded proteins may not be random coils

A protein folding reaction is carried out typically starting from a protein unfolded in a high concentration of chemical denaturant such as urea or guanidine hydrochloride. Such unfolded states of proteins have been modelled traditionally as statistical random coils in which local interactions dominate conformational behaviour[120]. Support for the random coil nature of the unfolded state came initially from measurements of hydrodynamic radius[121,122], which give an ensemble average picture of global molecular properties. Low resolution spectroscopic methods such as circular dichroism or fluorescence very often fail to detect any structure in unfolded states of proteins. Nevertheless, recent simulations of unfolded polypeptides have indicated that steric clashes can prevent a protein chain from adopting any arbitrary conformation, thus constraining the unfolded state ensemble[123,124], and probably favouring conformations having native-like bond angles[125]. More importantly, there is now increasing evidence from high-resolution NMR experiments, that unfolded states of proteins are not completely random. NMR studies have shown the existence of hydrophobic clusters in the unfolded forms of some proteins[126–130], and both hydrophobic clusters and residual secondary structure in the unfolded forms of other proteins[131–135]. Identification of any residual structure in the unfolded state is clearly crucial to obtaining an understanding of the protein folding reaction, because such residual structure not only indicates local conformational preferences within the very heterogeneous ensemble of conformations that constitute the unfolded state, but can also bias the free energy landscape of folding so that folding occurs along a route determined by the structures and hence, energies of the forms that predominate in the initial unfolded state.

Unfolded states in unfolding conditions

Localized hydrophobic clusters have been observed in several proteins under high denaturant conditions where secondary structures are unstable, and appear to influence the folding of the protein by acting as nucleation sites around which structure can be formed[126–128]. In an early study, NOE measurements in high urea concentration followed by distance geometry calculations revealed a hydrophobic cluster in the N-terminal 63-residue domain of 434-repressor, involving local side-chain interactions[126]. More recently, a hydrophobic cluster has been detected in the α-subunit of tryptophan synthase, in different urea concentrations, by monitoring the change in chemical shift position of the Cε proton of a histidine residue by 1D NMR[127]. Site-directed mutagenesis as well as studies on a peptide model spanning the histidine and the mutated residues suggest that this residual structure may be stabilized by interactions with side chains distant in sequence to the histidine. In the rat intestinal fatty acid binding protein, ^1H-^{15}N HSQC spectra and measurement of HX rates in increasing urea concentration have shown that hydrophobic clusters exist in the unfolded protein[128]. These clusters appear to guide the formation of native β-strands, and finally, the stabilization of hydrogen bonds, supporting a 'hydrophobic zipper' model of folding[136]. They persist until folding is complete and can be visualized in the native protein structure.

On the other hand, NMR studies of urea-unfolded states of some other proteins, such as the immunoglobulin binding domain of streptococcal protein G[137] and of the third fibronectin type III domain from human tenascin (TNfn3)[138], indicate that these states are essentially random coil conformations with no residual native secondary structure, although regions of reduced backbone mobility in Tnfn3 correlate to regions of extreme hydrophobicity[138].

Thus, in the rapidly interconverting ensemble of conformations of the unfolded state, some structures will be extended and others compact. Interactions between side

chains such as those leading to the formation of hydrophobic clusters are likely to bias the unfolded protein ensemble towards compact states and to thereby facilitate folding[120].

Unfolded states in refolding conditions

The structural propensities of unfolded states generated under different denaturing conditions may vary with the conditions used, because disordered states of proteins are essentially ensembles of many conformations[134,139,140]. It has therefore been argued that residual structure observed under strong denaturing conditions is not likely to be representative of disordered states of proteins *in vivo*. In the case of the N-terminal SH3 domain of *Drosophila* drk protein (drkN SH3), a compact unfolded state (U_{exch}), which exists in equilibrium with the folded state in aqueous solution at neutral pH, has been characterized by studying the ^{15}N relaxation dynamics of the backbone[141,142]. Long range NOEs detected in the U_{exch} state indicate the existence of compact conformations with some native-like features[143]. Hydrophobic clustering in the compact conformations of the U_{exch} ensemble, generates multiple non-native turn-like structures, which are stabilized by long range tertiary interactions[143]. These turn-like structures are thought to limit the conformational search, and direct subsequent β-sheet formation and stabilization, and act as multiple initiation sites during the folding of this domain. This would be in agreement with lattice-model simulations of protein folding which suggest multiple folding nuclei for a nucleation-collapse mechanism[144].

The long range NOEs in the U_{exch} state of drkN SH3 disappear in 2 M guanidine hydrochloride[143], suggesting that the unfolded state ensemble is a population of extended and compact conformations in equilibrium with each other, and that the interactions which stabilize the compact conformation are disrupted in the presence of denaturants. Rapid equilibria between collapsed and extended populations of protein molecules, in which the two populations are separated by a low energy barrier, have also been seen by optically monitored kinetic experiments in the cases of the initial intermediate ensembles that form during the folding of cyt *c*[17,145] and ribonuclease A[70,77].

Chemical shift analysis, NOE effects, coupling constants measurements and the ^{15}N relaxation parameters of some other proteins such as apoplastocyanin, which is unfolded at low salt concentrations at neutral pH, show that the unfolded state is not compact, presumably due to the absence of stabilizing long-range interactions[146]. There is, however, some evidence of native-like and non-native local structuring which may serve as initiation sites for folding of the protein[146]. Thus, structural characterization of the unfolded state ensemble can provide valuable information for understanding protein folding mechanisms.

Native-like topology in the unfolded state

For many proteins, the topology of the native state is found to be a major determinant of the folding rate, and the importance of topology is related to the entropic cost of bringing two regions of the polypeptide chain together so that they can interact[147–149]. Clearly, if native-like contacts form in the unfolded state itself, there will be a smaller loss of configurational entropy during the folding process, but it is not clear if this will result in a lowering of the energy barrier to folding. In the case of at least two proteins, NMR has provided evidence for native-like spatial positioning and orientation of the polypeptide chain segments in the unfolded state[130,150,151]. By an extensive analysis of ^{15}N relaxation parameters, particularly the R_2 relaxation rates, clusters of hydrophobic structure have been shown to exist in both oxidized and reduced forms of hen lysozyme even under strongly denaturing conditions[129,130]. Six hydrophobic clusters exist in sequentially distinct regions along the polypeptide chain, and four of them involve tryptophan residues (Figure 2)[129,130]. The locations of these clusters are similar for the protein in water and in 8 M urea. Significantly, the hydrophobic clusters in both oxidized and reduced lysozyme are disrupted, in the absence or presence of urea, by a single point mutation that replaces Trp62, located at the interface of the two domains in the native protein, with glycine[130]. The changes in relaxation rates throughout the sequence in the mutant protein indicate that native-like structure in the unfolded wild-type protein is stabilized by the involvement of Trp62 in long range tertiary interactions. Trp62 is exposed in native lysozyme, but real-time NMR experiments have shown that this residue, like other tryptophan residues, is largely inaccessible to solvent during early stages of folding[152]. Other mutational[153] and HX studies[65,69] suggest that Trp62 is involved in non-native tertiary interactions in the unfolded state. Thus, non-native interactions appear to stabilize a native-like core in the unfolded state[130].

Staphylococcal nuclease can be denatured by removing a few amino acid residues from both ends of the polypeptide chain[154]. The resulting fragment, called Δ131Δ, has been characterized extensively as a model system for the unfolded state of the protein under non-denaturing conditions, and shows the presence of residual β hairpin structure in the region that forms a three-stranded antiparallel β-sheet in the native state[154,155]. This β hairpin structure appears to be stabilized by non-native local hydrophobic interactions[155]. HX studies also show that fixed inter-strand hydrogen bonds normally associated with β-sheets are lacking in this β structure[156]. An innovative method was developed to obtain long range distance

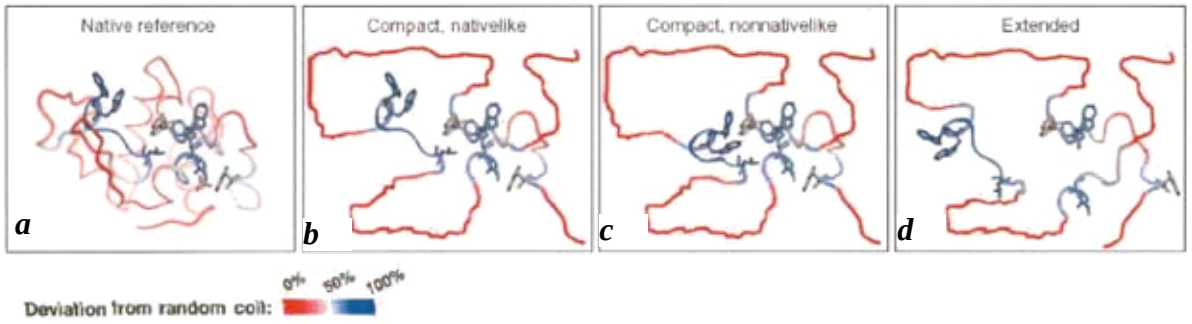

Figure 2. Schematic representation of hen lysozyme (reduced) backbone conformations in different states of the protein[166]. The colour code from red to blue depicts deviations of the R_2 relaxation rates of the backbone NHs from random coil values. Six hydrophobic clusters were shown to exist in reduced lysozyme, both in water and in 8 M urea[130]. The hydrophobic clusters appear to form a network with cooperative, long range interactions, with Trp62 playing a central role in maintaining the stability of the network. The figure has been reproduced, with permission, from ref. 130.

constraints in Δ131Δ, by measuring the enhancement of backbone amide proton R_1 and R_2 relaxation rates, induced by paramagnetic nitroxide spin labels[150,157]. These spin labels were attached at unique cysteine residues, introduced at fourteen sites along the polypeptide chain[157]. This approach appears to be a more suitable probe than NOE measurement for measuring tertiary structure in non-native states of proteins, because the dipolar interaction between a free electron and a proton has a range as large as 20 to 25 Å, where as a dipolar interaction or NOE between protons is restricted to distances below 5.5 Å. Moreover, the NOE between protons is highly sensitive to correlation times[157]. An ensemble of structures was calculated for Δ131Δ, by use of a combined distance geometry/molecular dynamics approach based on the ~700 long-range distance constraints obtained[150]. The global topology of a significant component of this unfolded ensemble was observed to be strikingly similar to that of the native state, and this component is likely to be stabilized by hydrophobic interactions[150].

Thus, it appears that the correct native-like topology of a protein can be established in unfolded states by hydrophobic interactions even in the absence of hydrogen bonds and tight packing interactions. The latter conclusion in Δ131Δ was drawn from the observation of a lack of significant protection of NH protons[156], and the observation of small order parameters for many segments of the chain determined by ^{15}N relaxation measurements[158]. A clear role in folding for residual structure in the unfolded state is also emphasized by this study. The presence of native-like structures stabilized by some non-native interactions, as seen for lysozyme and staphylococcal nuclease, may be a general feature of compact denatured states[130,155,156]. This would be consistent with a mechanism of folding in which the protein collapses to a compact state that subsequently undergoes a slower process of structural rearrangement to form the native state. A recent study has reported measurement of residual dipolar couplings in the Δ131Δ fragment oriented in strained polyacrylamide gels to impose slight orientation on the macromolecule[151]. Native-like spatial positioning and orientation of polypeptide chain segments are seen to persist even in 8 M urea, demonstrating that long-range ordering can occur much before the folding protein attains a compact conformation. The existence of long-range structure in the presence of relatively little short-range structure can occur through a network of hydrophobic clusters, of the kind observed in the denatured state of lysozyme[130], because hydrophobic interactions are not completely eliminated in 8 M urea[159]. Alternatively, the topological preference of the unfolded state may not be stabilized by long-range interactions, but instead may arise from the cumulative effects of many local interactions between each amino acid and its neighbouring side chains[123,160].

Residual secondary structure in the unfolded state

Since secondary structural elements are compact, their formation leads to a reduction in the otherwise vast conformational space that an unfolded polypeptide chain has to explore. NMR studies can provide information about conformational preferences of the protein backbone[113] in the unfolded state. Whether native-like secondary structural elements can actually form in the unfolded state depends on the conformational preference of the protein sequence, and on their stabilization by long range tertiary interactions[14]. Native-like residual secondary structure has been seen by measurements of NOE, coupling constants, and chemical shifts in the cold denatured state of barstar in moderate urea concentration[133]. Recently, a detailed NMR analysis of conformational preferences and backbone dynamics of the acid-unfolded state of apomyoglobin has shown significant population of native helical structure in some regions of the protein[131,132]. These regions of native-like residual secondary structure in the unfolded state of a protein could potentially serve as initiation sites for folding.

In the urea-denatured state of barnase, the regions of residual fluctuating secondary structure characterized by NMR[131,132], correspond fairly well to the moderately and highly structured regions of the intermediates and transition states of folding of the protein[161]. This observation led to the suggestion that protein folding is initiated around these native-like local structures that persist in the unfolded ensemble until enough stabilizing long range interactions form a nucleation site, on to which further secondary structure formation and final packing of native tertiary structure take place[135,162]. This suggestion is supported by an MD simulation of the unfolded state of barnase in which native-like residual secondary structure as well as a few fluctuating tertiary contacts have been observed[135]. On the other hand, the unfolded state of chymotrypsin inhibitor 2 (CI2) is largely unstructured as probed by NMR, with a tendency for very weak native helical structure and some weak hydrophobic clustering[163]. MD unfolding simulations of CI2 confirm these NMR observations, and it was proposed that the nucleation site for CI2 folding probably remains embryonic until a sufficient number of long range contacts are made, which are linked to collapse and secondary structure formation leading to cooperative folding[163].

Compaction and secondary structure formation during folding

Compaction of an extended polypeptide chain has been suggested as a principal driving force in the folding of proteins[164]. A fundamental question in protein folding is whether a hydrophobic collapse precedes or occurs concomitantly with the formation of secondary structure early during the folding reaction[165,166]. As is evident from the above examples, the unfolded states of proteins may be compact to varying degrees, depending on the existence of local or long range hydrophobic interactions, which may assist in consolidation of the protein core. Residual secondary structures in the unfolded state appear to act as nucleation sites, so that folding may be initiated by a nucleation–condensation/collapse mechanism[14]. If more stable secondary structural elements were to exist in the unfolded state, folding might be expected to proceed in a hierarchical manner with these units docking with each other, as the protein chain collapses and native packing interactions are established. Statistical mechanical models of protein folding also suggest that the presence of fluctuating native-like secondary structure in the unfolded state ensemble facilitates rapid folding to the native state[28].

An extensive NMR analysis of structural propensities and backbone dynamics of the acid-unfolded and partially folded states of apomyoglobin (apoMb) at equilibrium, shows the existence of cooperatively folded helical structures in these states[131,132,167]. In the acid-unfolded state, a region of local structure formation, which includes a non-native helix, as well as a region of local hydrophobic collapse, probably play an important role in the initiation of folding, and in directing the collapse of the protein chain[132]. The structural characterization and dynamics of the partially folded state of apoMb formed at equilibrium clearly suggests that stable native-like structure already exists in some regions in the intermediate, while in other regions helical structure, though unstable and not well-localized, is formed partially[132,167]. Thus, the incremental development of helical structure and progressive restriction of backbone fluctuations occurs at various stages of compaction of apoMb on an equilibrium folding pathway, as shown schematically in Figure 3. The NMR studies at equilibrium are supported by kinetic experiments of myoglobin folding[82,168,169], in which rearrangement and packing of the correct helical structure by formation of ordered tertiary structure occurs on a millisecond time scale. In the case of the folding of acid-denatured cyt *c*, no additional secondary structure forms during the initial chain collapse reaction, but thereafter, secondary structure and compactness are acquired in a concerted manner[12,170]. The step-wise formation of helical structures observed in the fast-mixing CD folding study[12], is consistent with the observation made in native-state HX experiments that the structure of cyt *c* can be divided into several cooperative folding units having different degrees of stability[98,101]. Thus, the results obtained with both apoMb and cyt *c* show that measurements at equilibrium can, indeed, be representative of kinetic folding pathway(s) of proteins.

NMR characterization and HX studies of the equilibrium molten globule or A form of bovine α-LA (BLA) show the persistence of some native helical structure, near-native compactness with a few native NOE contacts, and evidence for non-native aromatic hydrophobic clustering[171–173]. In fact, an important achievement in the detection of NOEs in partially folded states was made with the development of an NMR methodology in which NOEs were generated in the A form of BLA, and the protein was then rapidly refolded to the native state under conditions in which refolding was accomplished before the NOE intensity decayed[173]. In this way, the NOEs generated in the A form could be transferred to the well-resolved spectrum of the native state for detection, thereby overcoming the difficulty of having low intensities of cross-peaks arising from broad resonances in partially folded structures. Very interestingly, NMR-monitored equilibrium unfolding studies of the A form of BLA[174] show that the unfolding of the A form is non-cooperative: the β-domain unfolds at lower urea concentrations than does the α-domain.

Kinetic studies of the refolding of BLA, monitored by CD, fluorescence and pulse-labelling by HX, indicate that the burst-phase intermediate formed within 1 s of initiating folding, resembles the equilibrium A form of the

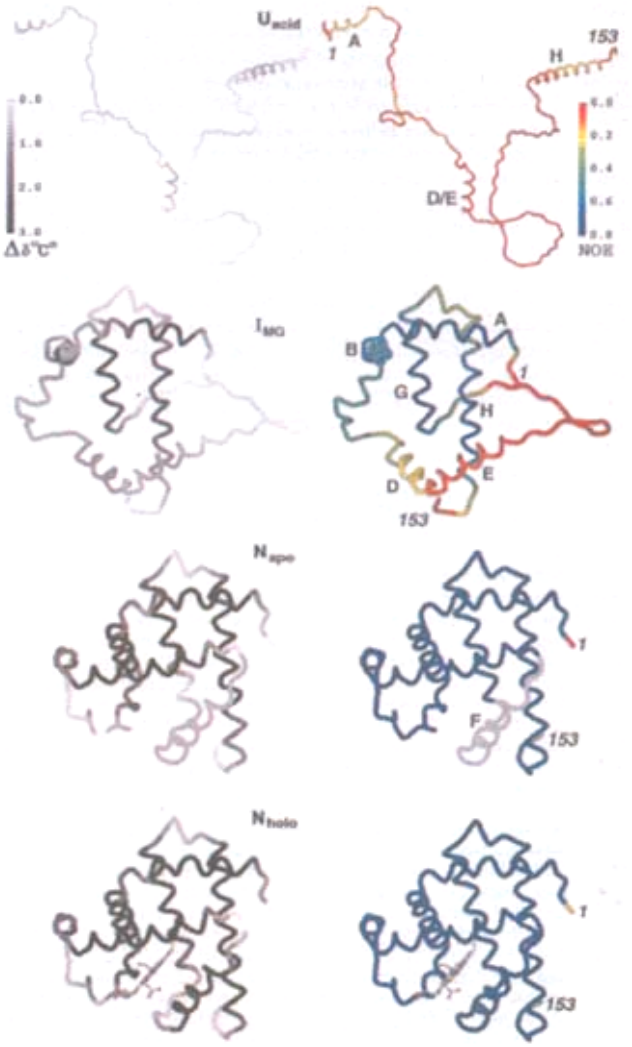

Figure 3. Schematic diagram illustrating the progressive development and stabilization of secondary structure and the increasing restriction of backbone flexibility as myoglobin folds to increasingly compact states. The left column shows the population of helices in the various states as indicated by the downfield secondary $^{13}C^{\alpha}$ chemical shift at each residue. The populations are depicted on a gray scale as indicated in the bar, with increasing grayness denoting increasing population of helix. The right column shows the $\{^1H\}$–^{15}N heteronuclear NOE at each residue, on a colour scale from dark blue (least flexible) to red (most flexible). The figure has been reproduced, with permission, from ref. 131.

protein, in terms of the amount and stability of secondary structure, degree of protection from HX and the absence of rigid side chain packing[83]. Moreover, NOESY spectra collected in real-time, under conditions where folding is slow enough to allow detection of the total steady-state NOE transfer between aromatic and aliphatic side chains as a function of time after initiating refolding, indicate the rapid formation (within the dead-time of the experiment of ~4 s) of a collapsed but partially disorganized state with near-native compactness[175]. Thus, given the close similarity between the kinetically detected molten globule form and the A form of α-LA, it seems likely that the native-like topology is generated very early in the folding process, prior to the formation of stable tertiary interactions. The slow step in folding must therefore involve a search for specific contacts between residues within a compact state that enable the protein to acquire its unique native structure rapidly and in a cooperative manner. The cooperativity of folding of α-LA was indicated by kinetic NMR experiments in which the rate at which native structure is formed was found to be identical at different sites and is effectively exponential[88,175]. The search for specific tertiary contacts in the collapsed state has to overcome energetic barriers on the folding energy landscape, arising from the close-packed nature of the structure[29]. Resonances in the NMR spectra of the partly folded species formed initially during folding are broadened because of interconversion between different conformational forms. This occurs much more slowly, on the NMR chemical shift time scale, than in highly denatured states, suggesting the existence of energy barriers[172,174].

NMR characterization of protein aggregation

For many proteins, the last steps in protein aggregation have been studied extensively by Fourier transform infrared (IR) spectroscopy[176,177], small-angle X-ray and neutron scattering measurements[176,178], X-ray fiber diffraction[179,180], electron microscopy[181,182] and also solid-state NMR[183]. Very little is, however, known about the initial steps of protein aggregation. The best way to characterize the initial conformational events that lead to the formation of a soluble protein aggregate, which precedes and transforms to an irreversible insoluble aggregate, would be by solution NMR methods. For instance, the dependence of R_2 relaxation rates of backbone ^{15}N nuclei has been used to study the aggregation (polymerization) process of cold shock protein A[184]. NMR structural studies of large molecular weight proteins (including soluble aggregated forms) are not straightforward because of low sensitivity and line broadening due to rapid transverse spin relaxation and extensive signal overlap in the highly complex spectra[41,185,186]. The recently developed, novel three-dimensional (3D) NMR experimental procedures of HNN and HN(C)N provide a new protocol for the sequential assignment of large proteins, including aggregated states of proteins[111]. These triple resonance experiments have been used to identify flexible regions in the 160 kDa soluble molten globule (A form) aggregate of barstar formed at low pH (Figure 4) (ref. 112). The A form of barstar exists as a symmetrical aggregate: only a single set of amide peaks is observed in the ^1H–^{15}N HSQC spectrum (Figure 4). The new assignment strategy, along with standard heteronuclear 2D and 3D NMR experiments, showed that the A form has an aggregated rigid core but with the N-terminal 20 residues

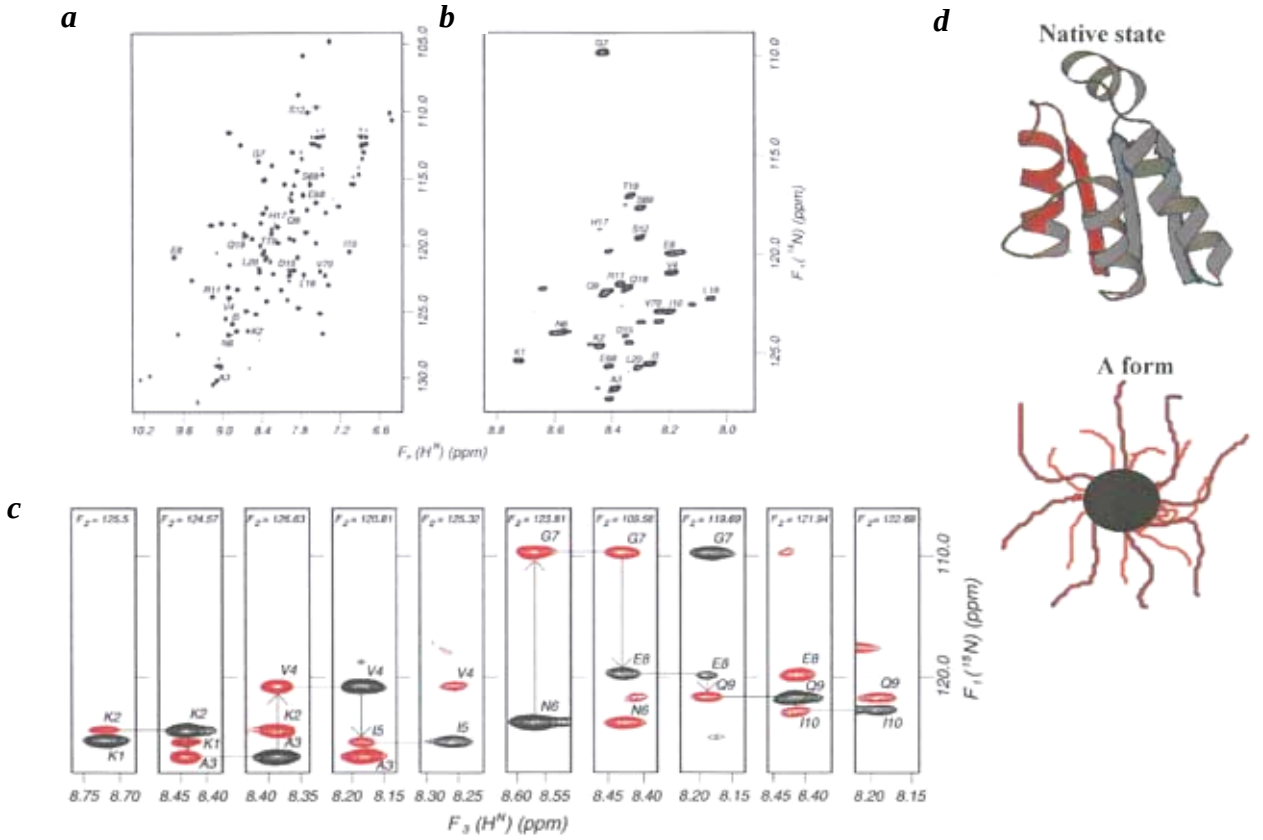

Figure 4. $^{1}H–^{15}N$ HSQC spectra of (***a***) native barstar at pH 6.8, and (***b***) A form of barstar at pH 2.7, and 25°C. In (***a***) and (***b***) only those cross peaks have been labelled for which spin systems could be unambiguously assigned at pH 2.7, and sequential connectivities obtained from the HNN and HN(C)N spectra, the HNCA and HN(CO)CA spectra, and the NOESY spectra. The low H^N chemical dispersion is evident in the HSQC spectrum of the A form of barstar, as compared to that in the native protein spectrum. ***c***, Illustrative sequential walks through $F_1–F_3$ planes in the HNN spectrum of the A form of barstar at pH 2.7. Sequential connectivities are indicated for the Lys1 to Ile10 stretch in the protein primary sequence. Strips from the spectrum at the appropriate H^N chemical shifts are shown. Black and red contours are positive and negative peaks, respectively. The distinct Gly plane in the HNN spectrum served as a starting triplet fixed point to obtain sequential connectivities. The F_2 (^{15}N) chemical shift is indicated at the top of each strip of the HNN spectrum. ***d***, Schematic representation of the aggregated A form of barstar formed at low pH, showing a rigid aggregated core and with the N-terminal 20 residues (coloured red in the native barstar structure) of each of the sixteen monomeric subunits hanging free in solution. The figure has been reproduced with modification from ref. 112.

of each of the monomeric subunits hanging free in solution, depicted schematically in Figure 4*d*. Secondary chemical shifts, $^{3}J_{HN,H\alpha}$ coupling constants, lowered H^N temperature coefficients, and the nature of sequential and a few medium range NOEs indicate that the free N-terminals are in an essentially random coil conformation, but with some local ordering of structure[112]. ^{15}N relaxation dynamics measurements and calculation of reduced spectral densities indicate large-amplitude high-frequency motions of a highly flexible protein backbone in the N-terminal segments[112]. This study demonstrated the utility of NMR to carry out structural and dynamic characterization of large soluble protein aggregates.

HX[187] and other NMR studies[188] have been useful for characterizing the partially unfolded intermediate forms of proteins which aggregate irreversibly to form the insoluble amyloid fibrils implicated in several neurodegenerative diseases[37–39]. The amyloid state itself has eluded structural characterization at the individual residue level, until recently[189]. On the basis of a comparison of the HX pattern of amide protons in the soluble native state and the amyloid state of β_2-microglobulin (β_2m), a structural model of the amyloid fibrils was obtained. It was shown that the β-sheet structure of native β_2m is rearranged during formation of the amyloid cross-β-sheet structure, such that some loop residues are converted into strands[189]. In this study, the problem of obtaining high-resolution structural data of large insoluble protein aggregates was overcome by a novel HX methodology which exploited the ability of dimethylsulfoxide to solubilize amyloid fibrils to the monomeric form, while suppressing, simultaneously, the intrinsic exchange rate of the NH protons.

Concluding remarks

It is clear that equilibrium NMR studies of disordered states of proteins can provide invaluable information at

the level of individual residues about the progressive accumulation of secondary structure and restriction of backbone dynamics, as the polypeptide chain collapses during folding to form more compact states. It has long been recognized that the burial of non-polar side chains, that is hydrophobic surfaces, is the main source of free energy change that drives the folding of proteins[190]. It appears from recent NMR studies that hydrophobic interactions are important in unfolded and partially folded states of proteins, either in the formation of local clusters around hydrophobic residues, or for constraining the polypeptide chain to a volume and topology like that of the native fold. It appears unlikely that short range van der Waals and electrostatic interactions, or hydrogen bonding, are involved in the long range order seen in the unfolded states of some proteins.

NMR studies have shown that the denatured states of several proteins are compact to varying degrees, and may contain native-like secondary structural elements[14]. Variations in the stabilities of these secondary structural elements for different proteins may lead to different folding mechanisms, either nucleation–condensation or hierarchical folding[14]. HX studies coupled with NMR play an important role in determining stabilities of different regions in a protein. For some proteins, the most stable region of the protein determined from native-state HX experiments, also appears to be the first to fold[49,60,62,91,103–105]. Thus, these studies, along with NMR analysis of residual structure in the unfolded state, can provide information about nucleation sites for initiation of folding. The role of a hydrophobic collapse in initiating folding, the role of secondary structure formation in guiding the folding process, and the question of whether non-native interactions present in unfolded and partially folded states of proteins necessarily retard folding, are issues which are currently being addressed.

1. Matthews, C. R., *Annu. Rev. Biochem.*, 1993, **62**, 653–683.
2. Roder, H. and Colón, W., *Curr. Opin. Struct. Biol.*, 1997, **7**, 15–28.
3. Levinthal, C., *J. Chim. Phys.*, 1968, **65**, 44–45.
4. Jones, C. M. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 11860–11864.
5. Philips, C. M., Mizutani, Y. and Hochstrasser, R. M., *ibid*, 1995, **92**, 7292–7296.
6. Nölting, B., Golbik, R. and Fersht, A. R., *ibid*, 1995, **92**, 10668–10672.
7. Panick, G., Malessa, R., Winter, R., Rapp, G., Frye, K. J. and Royer, C. A., *J. Mol. Biol.*, 1998, **275**, 389–402.
8. Jacob, M., Holtermann, G., Perl, D., Reinstein, J., Schindler, T., Geeves, M. A. and Schmid, F. X., *Biochemistry*, 1999, **38**, 2882–2891.
9. Callender, R. H., Dyer, R. B., Gilmanshin, R. and Woodruff, W. H., *Annu. Rev. Phys. Chem.*, 1998, **49**, 173–202.
10. Takahashi, S., Yeh, S. R., Das, T. K., Chan, C. K., Gottfried, D. S. and Rousseau, D. L., *Nat. Struct. Biol.*, 1997, **4**, 44–50.
11. Shastry, M. C., Luck, S. D. and Roder, H., *Biophys. J.*, 1998, **74**, 2714–2721.
12. Akiyama, S., Takahashi, S., Ishimori, K. and Morishima, I., *Nat. Struct. Biol.*, 2000, **7**, 514–520.
13. Fersht, A. R., in *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein folding*, W. H. Freeman and Co, New York, 1999.
14. Fersht, A. R. and Daggett, V., *Cell*, 2002, **108**, 573–582.
15. Konermann, L., Collings, B. A. and Douglas, D. J., *Biochemistry*, 1997, **36**, 5554–5559.
16. Simmons, D. A. and Konermann, L., *ibid*, 2002, **41**, 1906–1914.
17. Lakshmikanth, G. S., Sridevi, K., Krishnamoorthy, G. and Udgaonkar, J. B., *Nat. Struct. Biol.*, 2001, **8**, 799–804.
18. Lyubovitsky, J. G., Gray, H. B. and Winkler, J. R., *J. Am. Chem. Soc.*, 2002, **124**, 5481–5485.
19. Jackson, S. E., *Folding Des.*, 1998, **3**, R81–R91.
20. Bhuyan, A. K. and Udgaonkar, J. B., *J. Mol. Biol.*, 2001, **312**, 1135–1160.
21. Dinner, A. R., Šali, A., Smith, L. J., Dobson, C. M. and Karplus, M., *Trends Biochem. Sci.*, 2000, **25**, 331–339.
22. Chu, R. A. and Bai, Y., *J. Mol. Biol.*, 2002, **315**, 759–770.
23. Dill, K. A. and Chan, H. S., *Nat. Struct. Biol.*, 1997, **4**, 10–19.
24. Chan, H. S. and Dill, K. A., *Proteins: Struct., Funct., Genet.*, 1998, **30**, 2–33.
25. Baldwin, R. L., *Nature*, 1994, **369**, 183–184.
26. Karplus, M., *Folding Des.*, 1997, **2**, S69–S75.
27. Miranker, A. D. and Dobson, C. M., *Curr. Opin. Struct. Biol.*, 1996, **6**, 31–42.
28. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. and Wolynes, P. G., *Proteins: Struct., Funct., Genet.*, 1995, **21**, 167–195.
29. Dobson, C. M., Šali, A. and Karplus, M., *Angew. Chem. Int. Ed. Engl.*, 1998, **37**, 868–893.
30. Chan, H. S. and Dill, K. A., *Annu. Rev. Biophys. Biophys. Chem.*, 1991, **20**, 447–490.
31. Dill, K. A., *Protein Sci.*, 1999, **8**, 1166–1180.
32. Wright, P. E. and Dyson, H. J., *J. Mol. Biol.*, 1999, **293**, 321–331.
33. Dyson, H. J. and Wright, P. E., *Curr. Opin. Struct. Biol.*, 2002, **12**, 54–60.
34. Plaxco, K. W. and Groß, M., *Nature*, 1997, **386**, 657–659.
35. Uversky, V. N., *Protein Sci.*, 2002, **11**, 739–756.
36. Dunker, A. K. *et al.*, *J. Mol. Graph. Model.*, 2001, **19**, 26–59.
37. Fink, A. L., *Folding Des.*, 1998, **3**, R9–R23.
38. Kelly, J. W., *Curr. Opin. Struct. Biol.*, 1998, **8**, 101–106.
39. Bucciantini, M. *et al.*, *Nature*, 2002, **416**, 507–511.
40. Radford, S. E. and Dobson, C. M., *Cell*, 1999, **97**, 291–298.
41. Riek, R., Pervushin, K. and Wüthrich, K., *Trends Biochem. Sci.*, 2000, **25**, 462–468.
42. Fiaux, J., Bertelsen, E. B., Horwich, A. L. and Wüthrich, K., *Nature*, 2002, **418**, 207–211.
43. Rüdiger, S., Freund, S. M. V., Veprintsev, D. B. and Fersht, A. R., *Proc. Natl. Acad. Sci. USA*, 2002, **88**, 11085–11090.
44. Arai, M. and Kuwajima, K., *Adv. Protein Chem.*, 2000, **53**, 209–282.
45. Dyson, H. J. and Wright, P. E., *Nat. Struct. Biol.* (*NMR Supplement*), 1998, **5**, 499–503.
46. Dyson, H. J. and Wright, P. E., *Methods Enzymol.*, 2001, **339**, 258–270.
47. Dobson, C. M. and Hore, P. J., *Nat. Struct. Biol.* (*NMR Supplement*), 1998, **5**, 504–507.
48. Udgaonkar, J. B. and Baldwin, R. L., *Nature*, 1988, **335**, 694–699.
49. Roder, K., Elöve, G. A. and Englander, S. W., *ibid*, 1988, **335**, 700–704.
50. Bhuyan, A. K. and Udgaonkar, J. B., *Curr. Sci.*, 1999, **77**, 942–950.
51. Hvidt, A. and Nielsen, S. O., *Adv. Protein Chem.*, 1966, **21**, 287–386.

52. Woodward, C. K. and Hilton, B. D., *Biophys. J.*, 1980, **32**, 561–575.
53. Woodward, C., Simon, I. and Tüchsen, E., *Mol. Cell Biochem.*, 1982, **48**, 135–160.
54. Molday, R. S., Englander, S. W. and Kallen, R. G., *Biochemistry*, 1972, **11**, 150–158.
55. Bai, Y., Milne, J. S., Mayne, L. and Englander, S. W., *Proteins: Struct., Funct., Genet.*, 1993, **17**, 75–86.
56. Connelly, G. P., Bai, Y., Jeng, M.-F. and Englander, S. W., *ibid*, 1993, **17**, 87–92.
57. Englander, S. W. and Mayne, L., *Annu. Rev. Biophys. Biomol. Struct.*, 1992, **21**, 243–265.
58. Baldwin, R. L., *Curr. Opin. Struct. Biol.*, 1993, **3**, 84–91.
59. Englander, S. W., *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 213–238.
60. Chamberlain, A. K. and Marqusee, S., *Adv. Protein Chem.*, 2000, **53**, 283–328.
61. Dyson, J. H. and Wright, P. E., *Annu. Rev. Phys. Chem.*, 1996, **47**, 369–395.
62. Kuwata, K., Shastry, M. C. R., Cheng, H., Hoshino, M., Batt, C. A., Goto, Y. and Roder, H., *Nat. Struct. Biol.*, 2001, **8**, 151–155.
63. Udgaonkar, J. B. and Baldwin, R. L., *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 8197–8201.
64. Udgaonkar, J. B. and Baldwin, R. L., *Biochemistry*, 1995, **34**, 4088–4096.
65. Radford, S. E., Dobson, C. M. and Evans, P. A., *Nature*, 1992, **358**, 302–307.
66. Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H. and Clore, G. M., *Science*, 1993, **260**, 1110–1113.
67. Juneja, J. and Udgaonkar, J. B., *Biochemistry*, 2002, **41**, 2641–2654.
68. Sauder, J. M. and Roder, H., *Folding Des.*, 1998, **3**, 293–301.
69. Gladwin, S. T. and Evans, P. A., *ibid*, 1996, **1**, 407–417.
70. Houry, W. A., Rothwarf, D. M. and Scheraga, H. A., *Biochemistry*, 1996, **35**, 10125–10133.
71. Kuwajima, K., Yamaya, H. and Sugai, S., *J. Mol. Biol.*, 1996, **264**, 806–822.
72. Shastry, M. C. and Roder, H., *Nat. Struct. Biol.*, 1998, **5**, 385–392.
73. Sosnick, T. R., Shtilerman, M. D., Mayne, L. and Englander, S. W., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 8545–8550.
74. Qi, P. X., Sosnick, T. R. and Englander, S. W., *Nat. Struct. Biol.*, 1998, **5**, 882–884.
75. Takei, J., Chu, R.-A. and Bai, Y., *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 10796–10801.
76. Bieri, O. and Kiefhaber, T., *J. Mol. Biol.*, 2001, **310**, 919–935.
77. Houry, W. A. and Scheraga, H. A., *Biochemistry*, 1996, **35**, 11734–11746.
78. Cook, K. H., Schmid, F. X. and Baldwin, R. L., *Proc. Natl. Acad. Sci. USA*, 1979, **76**, 6157–6161.
79. Laurents, D. V., Bruix, M., Jamin, M. and Baldwin, R. L., *J. Mol. Biol.*, 1998, **283**, 669–678.
80. Sanz, J. M. and Fersht, A. R., *Biochemistry*, 1993, **32**, 13584–13592.
81. Nath, U. and Udgaonkar, J. B., *ibid*, 1995, **34**, 1702–1713.
82. Jennings, P. A. and Wright, P. E., *Science*, 1993, **262**, 892–896.
83. Arai, M. and Kuwajima, K., *Folding Des.*, 1996, **1**, 275–287.
84. Jamin, M. and Baldwin, R. L., *J. Mol. Biol.*, 1998, **276**, 491–504.
85. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y. and Kuwajima, K., *J. Mol. Biol.*, 1998, **275**, 149–162.
86. Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. and Fujisawa, T., *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 1329–1334.
87. Ha, J. H. and Loh, S. N., *Nat. Struct. Biol.*, 1998, **5**, 730–737.
88. Balbach, J., Forge, V., van Nuland, N. A. J., Winder, S. L., Hore, P. J. and Dobson, C. M., *ibid*, 1995, **2**, 865–870.
89. Rami, B. R. and Udgaonkar, J. B., *Biochemistry*, 2002, **41**, 1710–1716.
90. Hughson, F. M., Wright, P. E. and Baldwin, R. L., *Science*, 1990, **249**, 1544–1548.
91. Raschke, T. M. and Marqusee, S., *Nat. Struct. Biol.*, 1997, **4**, 298–304.
92. Dabora, J. M., Pelton, J. G. and Marqusee, S., *Biochemistry*, 1996, **35**, 11951–11958.
93. Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A. and Baum, J., *ibid*, 1993, **32**, 5681–5691.
94. Schulman, B. A., Redfield, C., Peng, Z-Y., Dobson, C. M. and Kim, P. S., *J. Mol. Biol.*, 1995, **253**, 651–657.
95. Nishimura, C., Dyson, H. J. and Wright, P. E., *ibid*, 2002, **322**, 483–489.
96. Shastry, M. C. and Udgaonkar, J. B., *ibid*, 1995, **247**, 1013–1027.
97. Kiefhaber, T., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 9029–9033.
98. Bai, Y., Sosnick, T. R., Mayne, L. and Englander, S. W., *Science*, 1995, **269**, 192–197.
99. Sosnick, T. R., Mayne, L., Hiller, R. and Englander, S. W., *Nat. Struct. Biol.*, 1994, **1**, 149–156.
100. Bhuyan, A. K. and Udgaonkar, J. B., *Proteins: Struct., Funct., Genet.*, 1998, **32**, 241–247.
101. Xu, Y., Mayne, L. and Englander, S. W., *Nat. Struct. Biol.*, 1998, **5**, 774–778.
102. Chamberlain, A. K., Handel, T. M. and Marqusee, S., *ibid*, 1996, **3**, 782–787.
103. Walkenhorst, W. F., Edwards, J. A., Markley, J. L. and Roder, H., *Protein Sci.*, 2002, **11**, 82–91.
104. Woodward, C., *Trends Biochem. Sci.*, 1993, **18**, 359–360.
105. Bai, Y., Milne, J. S., Mayne, L. and Englander, S. W., *Proteins*, 1994, **20**, 4–14.
106. Mayo, S. L. and Baldwin, R. L., *Science*, 1993, **262**, 873–876.
107. Llinas, M., Gillespie, B., Dahlquist, F. W. and Marqusee, S., *Nat. Struct. Biol.*, 1999, **6**, 1072–1078.
108. Lu, J. and Dahlquist, F. W., *Biochemistry*, 1992, **31**, 4749–4756.
109. Hollien, J. and Marqusee, S., *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 13674–13678.
110. Clarke, J. and Fersht, A. R., *Folding Des.*, 1996, **1**, 243–254.
111. Panchal, S. C., Bhavesh, N. S. and Hosur, R. V., *J. Biomol. NMR*, 2001, **20**, 135–147.
112. Juneja, J., Bhavesh, N. S., Udgaonkar, J. B. and Hosur, R. V., *Biochemistry*, 2002, **41**, 9885–9899.
113. Wishart, D. S. and Sykes, B. D., *Methods Enzymol.*, 1994, **239**, 363–392.
114. Karplus, M., *J. Chem. Phys.*, 1959, **30**, 11–15.
115. Dyson, H. J. and Wright, P. E., *Annu. Rev. Biophys. Biophys. Chem.*, 1991, **20**, 519–538.
116. Abragam, A., in *Principles of Nuclear Magnetism*, Clarendon Press, Oxford, 1961.
117. Kay, L. E., *Nat. Struct. Biol.* (*NMR Supplement*), 1998, **5**, 513–517.
118. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A. and Kay, L. E., *J. Biomol. NMR*, 1995, **6**, 153–162.
119. Lipari, G. and Szabo, A., *J. Am. Chem. Soc.*, 1982, **104**, 4546–4559.
120. Smith, L. J., Fiebig, K., Schwalbe, H. and Dobson, C. M., *Folding Des.*, 1996, **1**, 95–106.
121. Damaschun, G., Damaschun, H., Gast, K., Gernat, C. and Zirwer, D., *Biochim. Biophy. Acta*, 1991, **1078**, 289–295.
122. Calmettes, P., Roux, B., Durand, D., Desmadril, M. and Smith, J. C., *J. Mol. Biol.*, 1993, **231**, 840–848.
123. Pappu, R. V., Srinivasan, R. and Rose, G. D., *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 12565–12570.
124. Van Gunsteren, W. F., *Angew. Chem. Int. Ed. Engl.*, 2001, **40**, 351–355.

125. Shortle, D., *Protein Sci.*, 2002, **18**, 11–26.
126. Neri, D., Billeter, M., Wider, G. and Wüthrich, K., *Science*, 1992, **257**, 1559–1563.
127. Saab-Rincón, G., Gualfetti, P. J. and Matthews, C. R., *Biochemistry*, 1996, **35**, 1988–1994.
128. Hodsdon, M. E. and Frieden, C., *ibid*, 2001, **40**, 732–742.
129. Schwalbe, H. *et al.*, *ibid*, 1997, **36**, 8977–8991.
130. Klein-Seetharaman, J. *et al.*, *Science*, 2002, **295**, 1719–1722.
131. Eliezer, D., Yao, J., Dyson, H. J. and Wright, P. E., *Nat. Struct. Biol.*, 1998, **5**, 148–155.
132. Yao, J., Chung, J., Eliezer, D., Wright, P. E. and Dyson, H. J., *Biochemistry*, 2001, **40**, 3561–3571.
133. Wong, K. B., Freund, S. M. and Fersht, A. R., *J. Mol. Biol.*, 1996, **259**, 805–818.
134. Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M. and Fersht, A. R., *ibid*, 1995, **254**, 305–321.
135. Wong, K., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R. and Daggett, V., *J. Mol. Biol.*, 2000, **296**, 1257–1282.
136. Dill, K. A., Fiebig, K. M. and Chan, H. S., *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 1942–1946.
137. Frank, M. K., Clore, G. M. and Gronenborn, A. M., *Protein Sci.*, 1995, **4**, 2605–2615.
138. Meekhof, A. E. and Freund, S. M. V., *J. Mol. Biol.*, 1999, **286**, 579–592.
139. Shortle, D., *Curr. Opin. Struct. Biol.*, 1993, **3**, 66–74.
140. Logan, T. M., Thériault, Y. and Fesik, S. W., *J. Mol. Biol.*, 1994, **236**, 637–648.
141. Farrow, N. A., Zhang, O., Forman-Kay, J. D. and Kay, L. E., *Biochemistry*, 1995, **34**, 868–878.
142. Farrow, N. A., Zhang, O., Forman-Kay, J. D. and Kay, L. E., *Biochemistry*, 1997, **36**, 2390–2402.
143. Mok, Y.-K., Kay, C. M., Kay, L. E. and Forman-Kay, J. D., *J. Mol. Biol.*, 1999, **289**, 619–638.
144. Klimov, D. K. and Thirumalai, D., *J. Mol. Biol.*, 1998, **282**, 471–492.
145. Hagen, S. J. and Eaton, W. A., *ibid*, 2000, **297**, 781–789.
146. Bai, Y., Chung, J., Dyson, H. J. and Wright, P. E., *Protein Sci.*, 2001, **10**, 1056–1066.
147. Plaxco, K. W., Simons, K. T. and Baker, D., *J. Mol. Biol.*, 1998, **277**, 985–994.
148. Clementi, C., Nymeyer, H. and Onuchic, J. N., *ibid*, 2000, **298**, 937–953.
149. Miller, E. J., Fischer, K. F. and Marqusee, S., *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 10359–10363.
150. Gillespie, J. R. and Shortle, D., *J. Mol. Biol.*, 1997, **268**, 170–184.
151. Shortle, D. and Ackerman, M. S., *Science*, 2001, **293**, 487–489.
152. Hore, P. J., Winder, S. L., Roberts, C. H. and Dobson, C. M., *J. Am. Chem. Soc.*, 1997, **119**, 5049–5050.
153. Rothwarf, D. M. and Scheraga, H. A., *Biochemistry*, 1996, **35**, 13797–13807.
154. Alexandrescu, A. T., Abeygunawardana, C. and Shortle, D., *ibid*, 1994, **33**, 1063–1072.
155. Wang, Y. and Shortle, D., *ibid*, 1995, **34**, 15895–15905.
156. Wang, Y. and Shortle, D., *Protein Sci.*, 1996, **5**, 1898–1906.
157. Gillespie, J. R. and Shortle, D., *J. Mol. Biol.*, 1997, **268**, 158–169.
158. Alexandrescu, A. T. and Shortle, D., *ibid*, 1994, **242**, 527–546.
159. Dill, K. A. and Shortle, D., *Annu. Rev. Biochem.*, 1991, **60**, 795–825.
160. Srinivasan, R. and Rose, G. D., *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 14258–14263.
161. Serrano, L., Matouschek, A. and Fersht, A. R., *J. Mol. Biol.*, 1992, **224**, 805–818.
162. Fersht, A. R., *Curr. Opin. Struct. Biol.*, 1997, **7**, 3–9.
163. Kazmirski, S. L., Wong, K.-B., Freund, S. M. V., Tan, Y.-J., Fersht, A. R. and Daggett, V., *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 4349–4354.
164. Chan, H. S. and Dill, K. A., *ibid*, 1990, **87**, 6388–6392.
165. Agashe, V. R., Shastry, M. C. R. and Udgaonkar, J. B., *Nature*, 1995, **377**, 754–757.
166. Baldwin, R. L., *Science*, 2002, **295**, 1657–1658.
167. Eliezer, D., Chung, J., Dyson, H. J. and Wright, P. E., *Biochemistry*, 2000, **39**, 2894–2901.
168. Nishii, I., Kataoka, M., Tokunaga, F. and Goto, Y., *ibid*, 1994, **33**, 4903–4909.
169. Ballew, R. M., Sabelko, J. and Gruebele, M., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 5759–5764.
170. Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. and Fujisawa, T., *ibid*, 2002, **99**, 1329–1334.
171. Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. and Dobson, C. M., *Biochemistry*, 1993, **32**, 1707–1718.
172. Baum, J., Dobson, C. M., Evans, P. A. and Hanley, C., *ibid*, 1989, **28**, 7–13.
173. Balbach, J., Forge, V., Lau, W. S., Jones, J. A., van Nuland, N. A. J. and Dobson, C. M., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 7182–7185.
174. Schulman, B. A., Kim, P. S., Dobson, C. M. and Redfield, C., *Nat. Struct. Biol.*, 1997, **4**, 630–634.
175. Forge, V., Wijesinha, R. T., Balbach, J., Brew, K., Robinson, C. V., Redfield, C. and Dobson, C. M., *J. Mol. Biol.*, 1999, **288**, 673–688.
176. Khurana, R., Gillespie, J. R., Talapatra, A., Minert, L. J., Zanetti, C. I., Millett, I. and Fink, A. L., *Biochemistry*, 2001, **40**, 3525–3535.
177. Zurdo, J., Guijarro, J. I., Jimenez, J. L., Saibil, H. R. and Dobson, C. M., *J. Mol. Biol.*, 2001, **311**, 325–340.
178. Yong, W., Lomakin, A., Kirkitadze, M. D., Teplow, D. B., Chen, S. H. and Benedek, G. B., *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 150–154.
179. Inouye, H., Fraser, P. E. and Kirschner, D. A., *Biophys. J.*, 1993, **64**, 502–519.
180. Blake, C. and Serpell, L., *Structure*, 1996, **4**, 989–998.
181. Serpell, L. C. *et al.*, *J. Mol. Biol.*, 1995, **254**, 113–118.
182. Zurdo, J., Guijarro, J. I., Jiménez, J. L., Saibil, H. R. and Dobson, C. M., *ibid*, 2001, **311**, 325–340.
183. Lansbury, P. T. Jr. *et al.*, *Nat. Struct. Biol.*, 1995, **2**, 990–998.
184. Alexandrescu, A. T. and Rathgeb-Szabo, K., *J. Mol. Biol.*, 1999, **291**, 1191–1206.
185. Kay, L. E. and Gardner, K. H., *Curr. Opin. Struct. Biol.*, 1997, **7**, 722–731.
186. Clore, G. M. and Gronenborn, A. M., *Curr. Opin. Chem. Biol.*, 1998, **2**, 564–570.
187. Liu, K., Cho, H. S., Lashuel, H. A., Kelly, J. W. and Wemmer, D. E., *Nat. Struct. Biol.*, 2000, **7**, 754–757.
188. McParland, V. J., Kalverda, A. P., Homans, S. W. and Radford, S. E., *ibid*, 2002, **9**, 326–331.
189. Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H. and Goto, Y., *ibid*, 2002, **9**, 332–336.
190. Kauzman, W., *Adv. Protein Chem.*, 1959, **14**, 1–63.
191. Santoro, J., Gonzalez, C., Bruix, M., Neira, J. L., Nieto, J. L., Hernanz, J. and Rico, M., *J. Mol. Biol. I*, 1993, **229**, 722–734.

ACKNOWLEDGEMENTS. This work was funded by the Tata Institute of Fundamental Research and the Wellcome Trust. J.B.U. is the recipient of the Swarnajayanti Fellowship from the Government of India.

Received 18 September 2002; revised accepted 21 November 2002